Methodology of classification, forecast and prediction of healthcare providers accredited in high quality in Colombia

Tomás Fontalvo-Herrera*

Faculty of Economics Sciences,
University of Cartagena,
Cartagena, Colombia
Email: tfontalvoh@unicartagena.edu.co
*Corresponding author

Enrique Delahoz-Dominguez

Industrial Engineering Department,
Engineering Faculty,
Universidad Tecnológica de Bolívar,
Cartagena, Colombia
Email: edelahoz@utb.edu.co

Orianna Fontalvo

Engineering Faculty,
Universidad del Norte,
Barranquilla, Colombia
Email: oriannaf@uninorte.edu.co

Abstract: This research presents a methodology for classification, forecasting and prediction of healthcare providers accredited in Colombia. For this purpose, a quantitative, descriptive and predictive analysis was carried out of 27 institutions accredited in Colombia by 2016. Consequently, the machine learning techniques cluster analysis and artificial neural networks were used to define business profiles of the institutions under study. The method classifying, forecasting and predicting the membership of a healthcare provider to a business profile, previously created based on the high-quality patterns of accreditation. The input variables were *assets*, *account receivable*, *inventory*, *property* and *equipment* and the output variables *health service sales* and *net profit*. The cluster analysis defined two main groups. 1) accredited institutions in the process of financial consolidation; 2) accredited institutions financially sound. The process of forecasting and prediction through the creation of an artificial neural network yielded a 95% CI (088, 0.9975) precision in the classification, and 100% and 80% for sensitivity and specificity values respectively. The results evidence the capacity of the proposed methodology to recognise the characteristics and association patterns of HCP accredited in high quality.

Keywords: cluster-analysis; neural networks; quality; business profiles; healthcare; Colombia.

Reference to this paper should be made as follows: Fontalvo-Herrera, T., Delahoz-Dominguez, E. and Fontalvo, O. (2021) 'Methodology of classification, forecast and prediction of healthcare providers accredited in high quality in Colombia', *Int. J. Productivity and Quality Management*, Vol. 33, No. 1, pp.1–20.

Biographical notes: Tomás Fontalvo-Herrera is a Chemical Engineer from the Universidad del Atlántico. He holds a Master's in Business Administration from the National University of Colombia. In 2012, he obtained his PhD from Atlantic International University. He is the author of more than 100 scientific publications. He currently works as a Professor and researcher at the University of Cartagena. His areas of interest are data mining in the business sector and multivariate data analysis.

Enrique Delahoz-Dominguez is an Industrial Engineer with a Masters in Operations Research from Universitat de Barcelona. He is an author of more than 15 papers in Scopus and WOS journals. Currently, he works at the Industrial Engineering Department, Technological University of Bolivar. He does research in machine learning, educational data mining and information systems.

Orianna Fontalvo holds a Bachelor's in Industrial Engineering. Currently, she is researching in efficiency analysis and big data for business development.

1 Introduction

In the social sphere, the positive impact on a society is proven by having strong and solid health institutions, as can be seen in different studies (Almost et al., 2018; Askim et al., 2015; Tsofa et al., 2017). On the other hand, the process of accreditation in high quality to which healthcare institutions voluntarily submit themselves serves as a certification of efficiency and safety in healthcare. However, quite often these institutions do not have objective decision-making tools to define processes and activities that allow evaluation of achievements and results associated with accreditation. Thus, the administrative and financial processes and the quality of the healthcare service are strongly associated, in such a way that it is necessary to study how these processes interact to define grouping structures that allow a joint and global assessment. In addition, currently the healthcare provider institutions follow guidelines for quality assurance, even without having an explicit will to obtain the quality label or implementing high quality standards, doing this only for the motivation of continuous improvement and organisational development.

Currently, it is possible to develop sectorial studies in healthcare due to the progress of the certification and regulatory organisations to collect, manage and publish quality and reliable financial and operational information. This allows the implementation of artificial intelligence techniques such as machine learning, for the development of new methodologies that add value to the organisational development of healthcare institutions. Consequently, the focus of this research consists in the characterisation of groups of healthcare institutions accredited in high quality, developing an inferential analysis of the grouping variables that allow establishing patterns, characteristics and determining relationships between the financial results and HCP accreditation. In this same line of work other authors have carried out studies that analyse clusters and patterns to subsequently make forecasts associated with these groups. In this sense, studies that analyse the behaviour associated with performance patterns (Nasseri and Kiaei, 2019) stand out. Likewise, the researchers (Chamboko and Bravo, 2018) worked on a forecast model for decision making in different organisations. Other research (Sreenivasan and Sundaram, 2018) also developed a forecast model in order to analyse the level of service. In this same sense, other studies (Kamble and Wankhade, 2017) analyse the factors associated with a sector that allows to predict significant productivity variables. In the logistics sector for example, Dweiri et al. (2015) develop a AHP based model to make predictions about inventory levels to satisfy customers.

The previous background, leads us to ask ourselves how to identify characteristics and patterns associated with the implementation of high standards of high quality in accredited institutions in Colombia, with which you can have quantitative criteria, to forecast the membership of an HCP to this type of excellent HCP? And in technical terms, ask ourselves: How to classify, forecast and predict the profiles and standards of excellence of accredited HCP in Colombia? With this, we can analyse the impact that accreditation in high quality in health has on the HCP that implement it. This leads to the following questions: How to define profiles or clusters of HCP accredited in Colombia? What are the significant items to characterise the profiles of HCP accredited in Colombia? What artificial neural network (ANN) structure is required to forecast companies with patterns associated with high-quality accreditation? How to forecast and predict the membership of a high-quality HCP to previously defined business profiles, supported by an ANN?

2 Literature review

2.1 Health and quality at the international level

Health and quality have been studied by many researchers. At the international level, different authors in their studies mention that the reason for the accreditation of most health entities has been the improvement in quality and organisational development. However, these accreditation processes were not subject to international standardisation but carried out by accreditation organisations that developed their accreditation standards without having legislation for accreditation in each country (Shaw et al., 2013). Consequently, Dammaj et al. (2016) develop a study to measure the quality of service for the health sector in Jordan, relating knowledge sharing and quality.

Likewise, quality is a very important issue in the performance of these entities. This is how Alolayyan et al. (2013) in their study claim that total quality management has a significant impact on hospitals. In effect, it is usual to apply methods to measure or evaluate the quality of services. Different authors used the gap approach, a methodology, to assess the quality of service in private hospitals (Khraisat et al., 2017). In a different approach, Jahantigh (2019) implement a Fuzzy logic approach to set a frontier between expectations and perceptions of patients. Based on a SERVQUAL approach, Raina et al. (2018) estimate the gaps between perception and expectations, determining that the hospital studied is far from meeting the expectations of patients.

On the other hand, when talking about quality, the costs that this entails are taken into account. Consequently, Zahar et al. (2016) in their study on quality costs in medical care, applied the quality cost model (COQ) to estimate the activities related to COQ in a clinical laboratory resulting in the majority of costs invested in 'good quality' costs. In the same way the study of quality costs can determine the quality status of an organisation. In Singhtaun and Hattayanon (2017), they claim that quality cost analysis can indicate inefficient quality cost elements and help managers understand the seriousness of their quality problems.

Consequently, quality is a fundamental issue in health entities, and this brings many advantages to these entities. Besides, nowadays quality and data analysis also have a causal relationship, different authors in their study claim that hospitals that used clinical quality metrics more effectively had a higher performance from the hospital management staff in setting goals and operational (Khraisat et al., 2017).

2.2 Health and quality in Latin America

As for Latin America, research conducted in the city of Maracaibo shows the need to implement high quality models and standards in health institutions, in order to improve the operations and normalisation processes within health entities and for this to be reflected in the provision of quality healthcare services, generating satisfaction for the end-users (González et al., 2013).

In Cuba, the importance of assuming and implementing high quality standards was concluded, because in the end this is reflected in the internal and external results of the hospitals or clinics where they are implemented (Forrellat Barrios, 2014) improving strategic management, care processes and control processes and audits for the continuous improvement of the services. Similarly, in another study in Mexico, there is evidence of the need to evaluate the performance of quality in terms of indicators that allow measuring the opportunity and relevance associated with the provision of medical services (Hernández et al., 2013).

2.3 Quality in the health sector in Colombia

Quality in health services in Colombia is regulated by the Mandatory Health Quality Assurance System (SOGCS), which consists of a set of norms, requirements, mechanisms and processes to generate, maintain, and improve the quality of health services in the country and is made up of the single system of qualification (SUH), audit program for the improvement of quality (PAMEC), unique system of accreditation (SUA) and the information system for health quality (*Decreto Único Reglamentario 780 de 2016*, 2016).

Therefore, the SUH is a set of rules established to regulate compliance with basic requirements in terms of technological and scientific capacity, equity and financial adequacy and technical-administrative capacity. This is intended to ensure users from possible risks in the provision of services and is required to be an authorised (*Decreto Único Reglamentario 780 de 2016*, 2016). On the other hand, the PAMEC is part of the mandatory health quality system in Colombia and constitutes an audit program for the improvement of quality, focusing its efforts on identifying and minimising the potential adverse risks in the provision of the medical service in health institutions in Colombia.

The mandatory guarantee system of health quality in Colombia also includes high-quality accreditation which is an optional process where health entities assume or not the commitment to implement high-quality standards and must demonstrate whether they are public or private HCP that can generate results and satisfaction to users. Therefore, compliance with this standard of high quality in health requires not only good administrative management but is linked to the availability of resources and infrastructure to be able to respond to system users.

Lastly, the quality information system guarantees the presentation of the indicators associated with the control of potential adverse risks in health entities. Therefore, it constitutes a competitive reference information system for users, in such a way that they can analyse and select the institutions with the best performance associated with the monitoring, control and improvement of adverse risks in the provision of assistance services.

In Colombia, the users who perceive the quality of healthcare are those who belong mostly to prepaid medicine, followed by the EPS (health provider entity) of the contributory scheme and lastly the EPS of the subsidised regime. There are also faults and complaints in the provision of services and permanent tension between payers and providers (Guerrero et al., 2011). In effect, the accreditation system is a call made to health establishments, and which aims to promote in the organisations the necessary changes or adjustments, to provide clients with guarantees in all processes and to bring quality to a plane of total optimisation (Fontalvo Herrera et al., 2016). Referring to Colombia, accreditation in health entities has had a positive impact on profitability indicators, specifically in the gross and operational margins. These variables and indicators are considered later for the determination of the financial items selected in this work.

2.4 Clustering in the health sector

The cluster analysis as a machine learning technique allows grouping for future prediction tasks. That is, clustering is responsible for dividing and classifying data that constitute an input to obtain a future result through an ANN.

Regarding the concept of De la Garza et al. (2013), the cluster analysis aims to form groups or segments as similar as possible within each group, but the most different compared to other groups. In this sense, clustering has been used in various activities. For example, Vaver (2014) used clustering to group geographic entities into a set of groups and identify whether each geographic entity is an ambiguously classified entity or a definitively classified entity. In addition to this, it established a measurement criterion for a group of groups determining an attribute of the ambiguously classified and definitively classified entities, compared according to their measurements with other conglomerates.

In the same way, Cong et al. (2015) in their article, service discovery acceleration with hierarchical clustering, they used clustering to group units of study through a type of conglomerate called hierarchical grouping, resulting in an improvement in the complexity of time with an acceptable loss of precision in the grouping of services and the discovery of services. On the other hand, clustering focused on the health sector also has wide applications. In Gøeg et al. (2015), they applied clustering in clinical models from local electronic health records based on semantic similarity. In this sense, the clinical models registered in the entity with different terminologies were grouped estimating intrinsic similarities between the different models, to be able to compare them and offer an overview of the multiple models. In the same way, the grouping or conglomerate was also used to group the clinical profile of the users (Mettler, 2013).

2.5 Neural network for forecasting and prediction in the health sector

The ANN according to studies (Cruz and Herrera, 2011), consists of a system of nonlinear mappings, which simulates the behaviour of human thought, by means of mathematical structures and networks, similar to those generated in the human being in its process thinking. Consequently, ANN has been widely used in the health sector, especially as an instrument for health professionals, but it has also been used for patient satisfaction. Carlucci et al. (2013) in their study, they propose the use of the ANN as a knowledge discovery technique to identify service quality factors that are important for outpatients.

Likewise, ANN as a predictive tool has been used to analyse the behaviour of customers of new companies. For example, in Ansari and Riasi (2016), they evaluated the factors that affect the loyalty of the clients. It was also found that the use of neural networks is a better approach to analyse the loyalty of the client, the satisfaction and the perceived value, besides of a beneficial technique for new companies.

Currently, it is important to measure the efficiency of hospitals due to the amount of resources invested in public health systems. In Tosun (2012), an enveloping data analysis is applied and a model based on the ANN to measure and evaluates hospital efficiency scores. Thus, a hospital is classified as efficient or inefficient. The results show that a well-trained ANN performs a good classification and even offers better solutions than a discriminant analysis DA. In addition, ANN shows the advantage of using less CPU time and computer resources than data envelopment analysis, especially in large data sets.

Finally, the application of ANN and clustering has made great contributions to the competitiveness of different organisations. For their part, Fontalvo-Herrera et al. (2018) show the capacity of cluster analysis to discriminate and classify accredited institutions in high quality profiles. In turn, Fontalvo et al. (2018) show an application in other contexts, which allowed grouping the entities into four competitive profiles that associate their characteristics, in this case ANNs showed an 85.7% capacity to discriminate and classify companies according to their competitive profile.

All the questions posed make it possible to define as a general objective in this study to propose a methodology for classifying, forecasting and predicting high quality accredited HCP in Colombia. The above mentioned allows to identify patterns of behaviour within the universe of accredited institutions and based on these patterns identify the characteristics that emerge from the previously defined profiles. To subsequently have the possibility of being able to predict whether an accredited IPS can be a) large IPS or b) IPS in the process of financial consolidation. The above can be verified once you have an IPS that is not part of this group, but that provides all the variables object of this investigation. Therefore, it is possible to offer relevant information to users and health service providers. The following specific objectives were considered:

1 identify the variables and financial items relevant for the classification, forecasting and preaching of the Accredited HCP in Colombia

2 define the profiles or clusters of the accredited HCP in Colombia

3 design an ANN structure to forecast and predict high quality HCP.

2.6 Performance metrics

The success of the classification process occurs by minimising the difference between the predicted value and the actual value. This relationship is described by the positive true (VP), true negative (VN), false positive (FP), and false negative (FN) metrics. The metrics used to evaluate performance will be the accuracy, positive predictive value (PPV), negative predictive value (PPN), sensitivity (S) and specificity (E) and the area under the ROC curve (AUC). The area under the curve represents the rate of TP and FP at various discrimination thresholds. A model with a perfect classification will have an AUC = 1. On the other hand, a totally random model would yield an AUC value = 0.5.

$$Accuracy = \frac{TP + TN}{n} \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

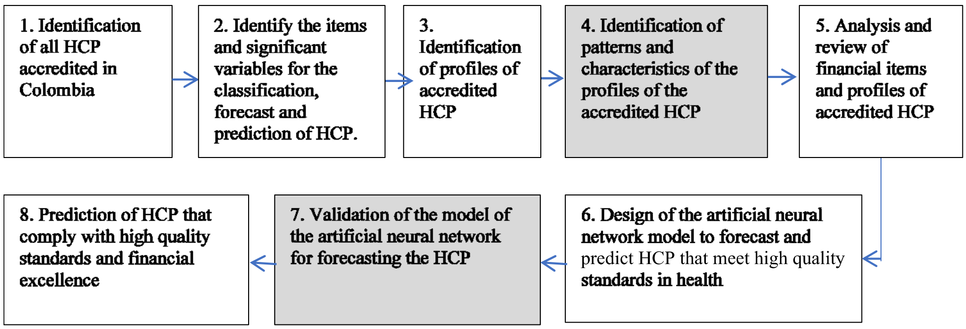
3 Methodology

As an epistemological conception of this research, the logical positivism paradigm was used, with which the scientific verification and the logical analysis for the development of the characterisation, forecast and prediction in this applied research were sought. The type of research had a quantitative, classificatory and predicative approach. As a population, all 27 Institutions that provide HCP services accredited in high quality, which presented their financial statements in 2016, were taken. This is to say all public hospitals and private clinics that in 2016 chose to implement high quality standards and accredit themselves. For this investigation the inductive and deductive method were used.

For the development of this study, variables and significant financial items were initially selected, associated with input variables such as: assets, accounts receivable, inventory, property, plant and equipment and the output variables associated with the financial items related to income, operational, sale of medical services, and gross profit. Figure 1 shows the methodology designed, its detailed and systematic stages, associated with the articulation of the different techniques of machine learning, which allows us to achieve the objectives of this research.

As can be seen from Figure 1, this research was carried out with a two-stage approach, the process of grouping and classification of accredited entities. As a first stage, it was based on a cluster approach, which allowed grouping the accredited institutions of the health sector into two large clusters. The first that was defined as institutions in the process of financial consolidation and a second cluster where the large health institutions were located. The above was established through the dendogram and the silhouette test and the two-dimensional graphic representation. In stage two, the design of the structure of the neural network was defined, which allowed establishing a structure that facilitated predicting previously established clusters. In order to measure the accuracy of the forecasting process, the confusion matrix and performance metrics were used.

Figure 1 Methodology for classification and forecasting of HCP (see online version for colours)



As primary sources, the information generated by the health superintendence and the Ministry of Health of Colombia, related to the HCP object of this investigation, was used for the year 2016.

As analytical techniques, the concepts of machine learning were used, related to the hierarchical cluster model used to classify the profiles of the HCP that implemented accreditation standards and identify the patterns of quality excellence. As well as the ANN, proposed for the prediction process of the HCP to one of the previously defined profiles. The type of ANN implemented consists of an MLP, which stands for multilayer perceptron, developed to extend the capabilities of the single perceptron which is basically a lineal discriminator on its own. The MLP allows to study problems that are not linearly separated, and is characterised by its three type of layers: input layer, hidden layers and output layer, and its feed forward connections; MLP are generally fully connected which is the case for the implementation on this research.

For the calculations and analysis of the information object of this research work, the software R was used as a tool for the processing of the information and data of the accredited HCP. Based on that, the profiles of the accredited HCP were calculated, as well as the design of the structure of the ANN, which makes it possible to forecast and predict other non-accredited HCP.

3.1 Cluster analysis

Cluster analysis is a multivariate analysis technique, whose purpose is to create groups trying to make the observations belonging to one group very close to each other and away from the observations located in another group. There are four categories of clustering algorithms:

1 partitioning
2 based on density
3 based on networks
4 hierarchical.

Partition algorithms such as k-means and partition around medoids iteratively refine a set of k clusters and usually do not scale well for large data sets. Density-based algorithms are capable of generating arbitrarily sized clusters and dealing with extreme values. Network-based algorithms reduce the clustering space to cells within a network, allowing efficient grouping of large data sets. Hierarchical algorithms can be agglomeration or divisive, in the agglomerated form two clusters are repeatedly merged, while in the divisive form a cluster is repeatedly divided into two. The Euclidean distance will be used as a similarity measure. So, the distance between and objects *i* and *j* is given by the equation (4)

$$d_{ij} = \sqrt{\sum_{p=1}^{k} \left(x_{ip} - x_{jp} \right)^2} \tag{4}$$

For the development of the first stage of the proposed methodology, a non-hierarchical cluster analysis is performed, through a partition algorithm around the medodides (PAM) (Kaufman and Rousseeuw, 1990), which defines a medoid as the representative element of a cluster from which the average dissimilarity towards each of the other members of the cluster is minimal. Consequently, a silhouette test was developed as it is developed in Menardi (2011), to evaluate the quality of the belonging of the observations to its group, this test delivers for each observation a weighting that ranges between values of –1 and 1, –1 being the assessment for observations that would be better represented in another cluster; The value 0 for observations that is in the border between 2 clusters; and 1 for observations well coupled to the current cluster.

3.2 *Artificial neural networks (ANN)*

The ANN is a statistical and computational technique, based on replicating the human learning process using data. The learning process is modelled by a nonlinear optimisation model (Chojaczyk et al., 2015). In another concept, Scardapane and Wang (2017) defines ANNs as networks composed of nonlinear processing nodes. In an applied approach, Ronao and Cho (2016) defines ANNs as structures capable of replicating human learning functions to identify and relate complex patterns to later develop a forecast on new observations. In the mathematical model, the sum of the different values modified by the synaptic weights, which determine whether the neuron is activated or not, is calculated using the equation (5).

$$Net_j = \sum\nolimits_{i=1}^{N} x_i * W_{ji} + \theta_j \tag{5}$$

Where *Net* represents the net input, w_i the synaptic weight of the neuron *j* over the input *x*2. The neuron is activated by an activation function, which propagates the y_i output of the neuron to another. Neuron or network output as shown in equation (6).

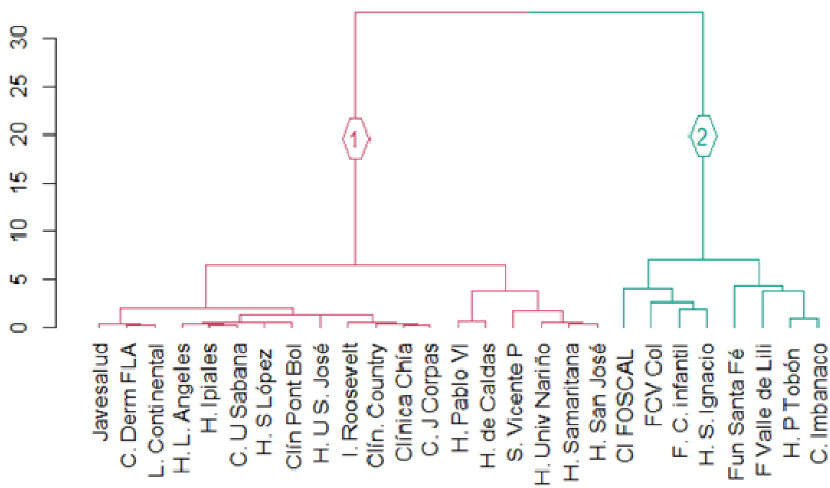
$$y_i = Fact\left(Net_j \right) \tag{6}$$

4 Results

4.1 Clustering results

The dendogram resulting from this analysis can be seen in Figure 2, where the formation of two clearly defined accredited HCP groups is evident. To develop clustering processes it is fundamental to have a wide knowledge of the problem studied, so based on the experience of the researchers in topics of organisational classification and unsupervised learning it was decided to analyse the conformation of clusters from size 2 to 5, which allows a greater detail in the generation of knowledge of the problem and support for the forecasting and prediction phase of the proposed methodology. The maximum silhouette is 0.65, corresponding to select two groups the k value of clusters (see Figure 3).

Figure 2 Cluster dendogram for Hierarchical cluster of HCP (see online version for colours)



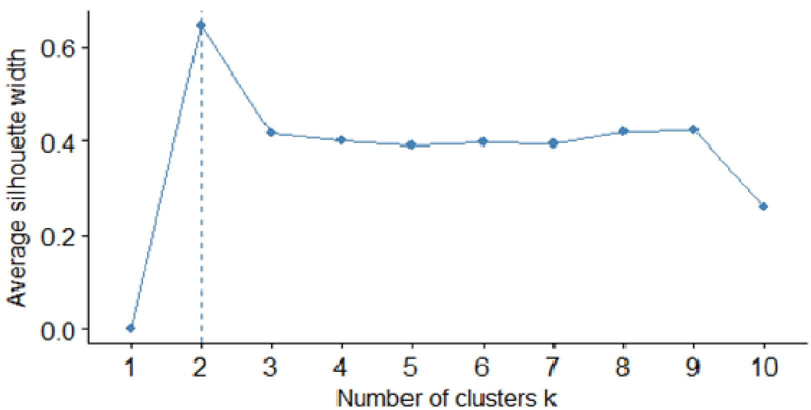
Using two groups as input parameter to the PAM analysis, we proceeded to analyse the representative elements of each group (see Table 1), in addition to generating a two-dimensional visualisation using a principal component analysis (PCA) (see Figure 4). In this way, two accredited HCP profiles were defined.

Table 1 Representative medoid elements by cluster

	Cluster 1	*Cluster 2*
Variable	H. Susana López	H. Pablo Tobón Uribe
Current assets (CA)	87, 392, 973, 000	586, 000, 000, 000
Account receivable (AC)	34, 867, 392, 000	141, 000, 000, 000
Inventory	1, 050, 327, 000	10, 340, 049, 000
Property and equipment (PE)	19, 279, 587, 000	345, 000, 000, 000
Operational income (OI)	65, 936, 446, 000	342, 000, 000, 000
Health service sales (HS)	65, 780, 155, 000	339, 000, 000, 000
Net profit (NP)	25, 670, 096, 000	60, 232, 788, 000

Cluster 2 has seven HCP, these are: Cardioinfantil Foundation, Valle del Lili Foundation, Pablo Tobón Uribe Hospital, Imbanaco Clinic, Cardiovascular Foundation of Colombia, Santa Fé Foundation and Foscal Clinic. The analysis of these HCP shows that all are private, while 58% are specialised in cardiology. In the geographical distribution it is found that these HCP are located in cities with a population of more than one million inhabitants. The patterns of belonging to this group are the HCP with assets over 300,000 Colombian pesos. In Figure 5 we can observe the total separation of the groups when comparing the variable assets with the rest of the six predictor variables of the PAM clustering model. The representative element of this group is the hospital Pablo Tobón Uribe, private institution of the department of Antioquia which is among the first four in the relationship to a greater value of the items assets, inventory and properties plants and equipment, is inferred in this group companies with a strong financial support, which is why this cluster will be labelled 'large accredited HCP'.

Figure 3 Silhouette analysis for number of clusters (k) from 1 to 10 (see online version for colours)



Cluster 1 has 20 HCP, these are: FLA Dermatological Center, University Hospital of Ipiales, Hospital Pablo VI of Bosa, University Hospital of Caldas, Clinic of the Pontificia Bolivariana University, Clinic of the University of La Sabana, San José University Hospital , Los Angeles Hospital, Chía Clinic, San José Hospital, Juan N Corpas Clinic, San Ignacio Hospital, Javesalud, Continental Clinical Laboratory , San Vicente de Paul Hospital and the Roosevelt Institute. In this cluster there is a balanced relation of the origin of the HCP, 50% public and 50% private, characterised as being general health institutions, providers of highly complex services, except for a clinical laboratory and a dermatological centre. In this group HCP are located in large cities and also medium and small cities of less than 500,000 inhabitants.

The factors of membership to this cluster are associated with the variable *operational income*. Here are the HCP with an operational income of less than 171,000 Colombian pesos. The representative element of this cluster is the hospital Susana Lopez Valencia, public entity of the department of Cauca. Therefore, in this group are the HCP with a value of low assets and the lowest values of utility among accredited institutions. This cluster will be labelled 'accredited institutions in the process of financial consolidation' for the rest of the investigation.

From Figure 5 it can be observed that in cluster 2, when the output variables associated with *operational income* and sale of health services, related to the assets of this IPS are reviewed, all IPSs in cluster 2 are well above all IPS associated with cluster 1.

Figure 4 Two-dimensional representation of the PAM analysis (see online version for colours)

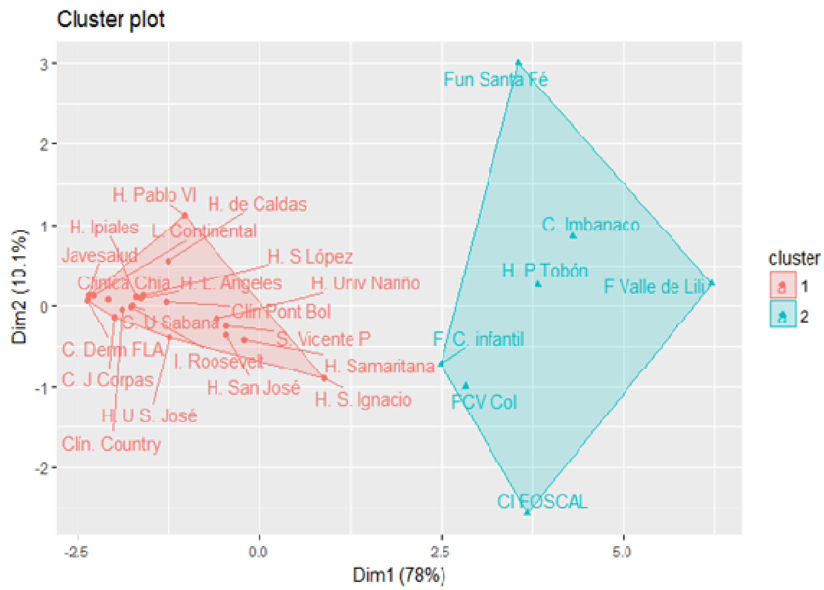
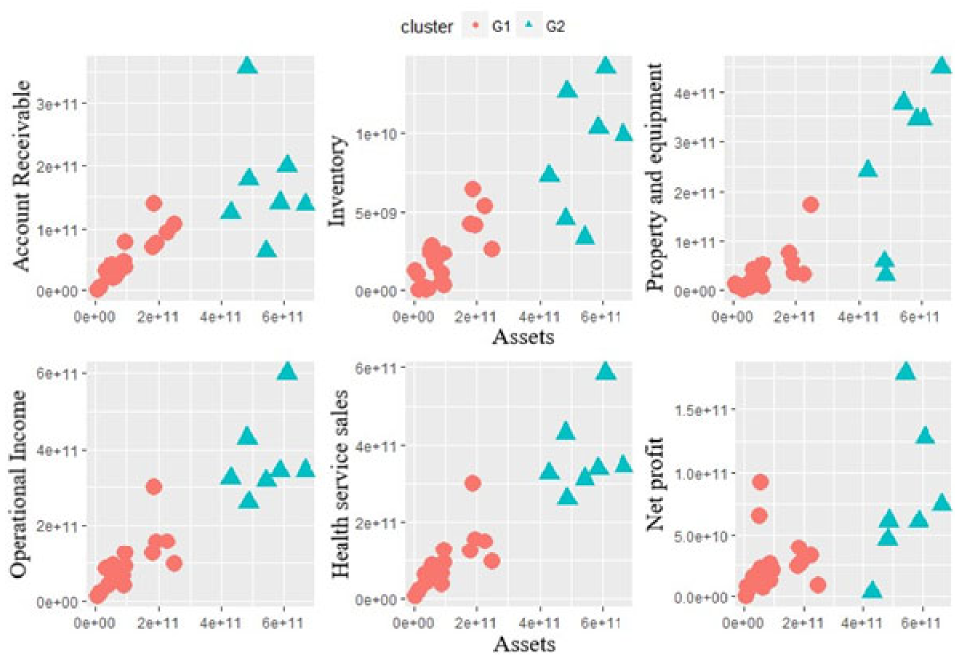


Figure 5 Comparison of predicted variables vs. assets (see online version for colours)



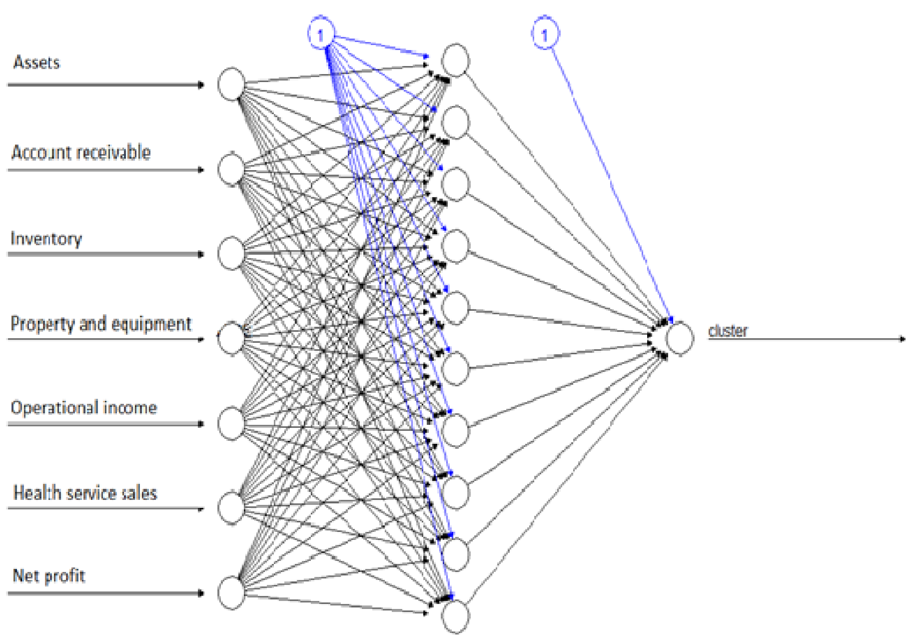
4.2 Artificial neural network (ANN) stage results

This section develops the procedure, development and results obtained from the implementation of an ANN for the classification of healthcare providers. In this way, in the first instance, the ANN construction process is described, and then the ANN validation process is executed through the cross-validation technique. At the end, the results are presented according to the precision, sensitivity, specificity and an analysis in the context of the health service provision of the proposed methodology.

Neural networks require quantitative variables for their application, therefore, in the context of this research where all the predictor variables are financial items expressed in continuous numerical values, the proposed technique is ideal. Considering that the available data and the observations of the analysed HCP are 27. The validity of the application of ANN in this research is supported by studies carried out by other authors (Ricks and Ventura, 2003) in which samples of size 16 to 40 are used to develop ANN, obtaining higher precision results for these houses to 98%. In turn Setiono (2001) developed classification problems using an ANN for data sets with less than 40 observations, in the same sense Peng et al. (2013) classify sick patients with ADHD using from ten to 110 observations, concluding that the ANNs are able to adapt to the size of the sample.

In function of the objective of classification of the present investigation, the creation of an ANN was designed as the following structure as observed in Figure 6:

Figure 6 Topology of the ANN (see online version for colours)



- seven input nodes, representing one for each of the seven active variable predictors, accounts receivable, inventory, plant and equipment properties, operating income, sale of health services and gross profit, these variables are standardised to avoid factors of bias and distortion by the magnitude of the variable.

- two exit nodes, representing the two categories found in the unsupervised phase of the proposed methodology, these are
    1. accredited HCP in the process of financial consolidation
    2. accredited HCP financially consolidated
- ten intermediate learning nodes
- one node of bias

For the internal development of ANN, a logarithmic sigmoidal transfer function was chosen, which has the advantage of being differentiable and restricting the output values to the range of 0 and 1. In this way, the transfer function is integrated to the purpose of the binary classification object of the investigation. It is important to note one of the problems commonly found in classification processes using supervised learning methodologies, over-training the model, which is the ability to generate high precision values for training data and poor performance against new data. In order to avoid this problem, ten perceptrons were used, which yielded good values according to the adjustment and precision of the training and evaluation phase, as will be taught below. The set of equations associated with each perceptron is represented in the following equations:

$$\phi_{-}(A, CC, INV, PPE, IO, VSS, UB) = \sum_{i=1}^{10} \beta_{i1} \,?(w_{Ai} A + \ldots + w_{UBi} UB + w_{0i}) \quad (7)$$

$$\phi_{+}(A, CC, INV, PPE, IO, VSS, UB) = \sum_{i=1}^{10} \beta_{i2} \,?(w_{Ai} A + \ldots + w_{UBi} UB + w_{0i}) \quad (8)$$

In the presented design of the ANN, each function represents for each HCP, its degrees of representativeness with each of the clusters established in the grouping phase, in relation to the values that the seven predictor variables yield. It should be noted that belonging to a group is not absolute, given that the ANN does not yield integer values of 0 and 1, but rather a continuous value associated with the representability mentioned above.

For the process of estimation of the parameters of the ANN, the cross-validation technique was used, creating ten new data sets randomly and making an evaluation and test process for these using the performance metric as the precision, in the which minimises the sum of the squares of errors in the real output (–1, 0) for the first category and (0.1) for the second.

Consequently, each of the ten new data sets created was divided into two, a training sample and an evaluation one, corresponding to 65% and 35% respectively, generating a set of solutions that allow obtaining a robust solution, analysed not only because of its punctual value but because of the distribution associated with a dispersion factor. The final structure of the specific weights of the ANN for the first layer can be seen in Table 1, and those of the second layer in Table 2.

Based on the result, the criterion of belonging to the categories is defined by assigning an HCP to the category ‘financially consolidated accredited institutions’; if the function
ϕ_{-} (*A*, *CC*, *INV*, *PPE*, *IO*, *VSS*, *UB*) > ϕ^{+} (*A*, *CC*, *INV*, *PPE*, *IO*, *VSS*, *UB*) and in the category ‘accredited institutions in the process of financial consolidation’ in the opposite case.

Table 2 Weights structure for the first layer of the ANN

Input	*N1*	*N2*	*N3*	*N4*	*N5*	*N6*	*N7*	*N8*	*N9*	*N10*
Bias	0.4	–0.3	0.66	–0.28	–1.63	0.76	0.64	–1.44	1.66	1.05
CA	–2.73	–1.94	1.04	–1.95	0.37	–0.45	–0.77	–1.66	–0.61	–0.81
AC	–0.2	–0.96	1.04	–0.89	–0.52	–0.38	–0.22	0.33	–0.54	–0.21
I	1.22	–0.07	0.82	0.59	–0.5	0.58	–0.36	–0.58	1.03	0.7
PE	1.26	1.22	–0.5	1.02	0.71	–0.25	1.29	0.25	–0.75	0.43
OI	–0.97	–0.44	0.66	0.17	–1.75	0.07	–0.08	0.06	–0.48	0.48
HS	–0.67	–2.8	0.81	0.16	0.83	–0.43	–0.12	–0.63	0.45	0.05
NP	–0.88	–0.27	–0.34	–0.42	1.01	0.57	1.17	1.24	2.17	–1.54

Table 3 Weights structure for the second layer of the ANN

	N1	*N2*	*N3*	*N4*	*N5*	*N6*	*N7*	*N8*	*N9*	*N10*
Cluster	0.22	–0.57	–0.43	0.84	–0.59	–0.37	1.2	1.95	0.61	0.86

Based on the evaluation of the performance of the established ANN, the resulting confusion matrix, the sensitivity and specificity indices and the operating curve (ROC) were analysed. In Table 4 the results of the confusion matrix for the training phase and the evaluation phase in which have been used as input vectors to the corresponding values 2016 are shown. In the rows of the confusion matrix the real membership of the HCP is represented and in the columns the prediction of classification of the model. For the training phase there is a precision value of 95% with a confidence interval of 95% confidence of (088, 0.9975). Next, the classification results for the year 2016 show an accuracy of 93%. The results show a good performance of the ANN for the training and evaluation phases, which is why problems of overfitting the model are ruled out. Therefore, the effectiveness of the methodology proposed in the present investigation is shown to predict and predict the membership of HCP to one of the two defined profiles.

Table 4 Confusion matrix for the training and test process

		Actual			
		Training		*Test*	
	Cluster	*C1*	*C2*	*C1*	*C2*
Prediction	C1	20	0	8	1
	C2	2	5	0	1

5 Discussion

Research by other authors Fontalvo et al. (2018) show the relevance of analysing business groups and forecasting them with ANNs. Therefore, with this knowledge framework, you can predict the membership of new companies or some of the previously identified groups. Granadillo et al. (2019) identify clusters and model them with discriminant analysis; however, the accuracy to predict does not achieve better results than when using neural networks. The above shows the relevance of this research and the use of the proposed techniques. In this same line of work, the authors Hasani et al. (2018) analyse a group of companies and, based on these, define the factors associated with their organisational performance. Other researchers, Wang et al. (2017) have also used neural networks as tools for the process of predicting significant variables in other contexts, which demonstrates the relevance of this tool for forecasting processes.

However, the performance of a classification model is not only given by its accuracy, but by its ability to discriminate between categories. For the proposed developed model, the ANN, sensitivity and specificity values are 100% and 80%. In this sense, the model is able to identify all health entities of the category 'large accredited HCP' and presents a case of misclassification when classifying the category 'small accredited HCP'. The overall effectiveness of the model can be analysed using the ROC curve which generated a 94.4% performance value (see Figure 7).

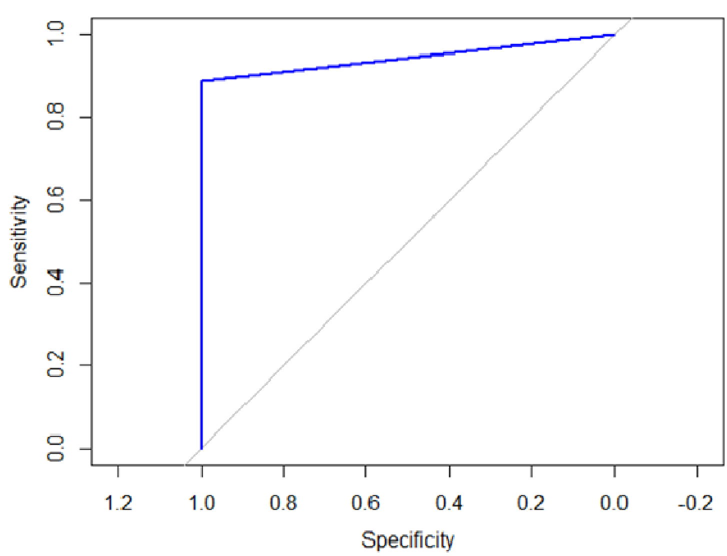
The values obtained in the performance indexes thrown by the model (see Table 5), validate the effectiveness of the model, which generates the possibility of identifying the profiles of HCP accredited through the financial items in different years of performance. From the results obtained, we can affirm that HCP accredited in Colombia are made up of two business profiles associated with the financial consolidation of accredited HCP and a second profile of HCP accredited with financial consolidation.

From Table 5 we can state that the structure of the ANN, designed is able to predict and predict the belonging of another HCP not belonging to the HCP studied with 94.4% reliability as can be seen in the ROC curve.

Table 5 Performance metric of the ANN

	Results
Performance indicator	1
Sensitivity	100%
Specifity	80%
Positive predicted value	100%
Negative predictive value	80%
Prevalence	90%

Figure 7 Roc curve for the classification based on ANN (see online version for colours)



6 Conclusions

As a conclusion to this research, the new international scientific knowledge was generated from the articulation of a series of concepts associated with the analysis and implementation of high-quality standards, automatic learning, conglomerate analysis and ANN for the HCP. With the above, articulated knowledge was generated through the design of an automatic learning methodology for the classification, forecasting and prediction of healthcare providers HCP of Colombia, which allows analysing the dynamics of the sector based on financial variables, generating knowledge of the impact, to be able to identify patterns and characteristics of excellence and based on this being able to forecast and predict if another institution has such characteristics. And in this way, we can analyse if another HCP complies with characteristics or patterns associated with the accreditation of high quality in its financial performance.

Likewise, this methodology is relevant for the different actors of the health sector, because it allows HCP to have a methodology and operational positioning tool for making objective decisions, which allows establishing business similarities between HCP, and in turn it serves as a roadmap or reference for a non-accredited HCP to map its current financial operational management to decide to undertake the task of obtaining a high quality accreditation. Similarly, the regulatory agencies of the health sector were designed a methodology and tool to determine public policies of surveillance, support and monitoring based on objective information to accredited HCP, in order to associate certain financial behaviour with the implementation of good practices established in a high quality accreditation. Also, users of the health sector can benefit from the current methodology by having criteria and information to choose their health service provider, depending on the financial strength of a company and the positive perception that generates a high-quality seal.

From the results of this investigation it can also be concluded that in Colombia there are two profiles or HCP cluster accredited in high quality, one cluster represented by the Pablo Tobón Uribe hospital and another cluster represented by the Susnan López Hospital.

It was also possible to demonstrate with empirical evidence that the implementation of high quality standards in the accredited HCP studied is associated with the administrative management and allocation of resources related to the disposition of assets, which guarantee the provision of services for end users and compliance with the high quality standards that enable the achievement of accreditation.

In the same way, in this research a robust ANN structure was established, which allows forecasting and predicting the belonging of an HCP to one of the two previously defined profiles. Likewise, this structure of ANN allows forecasting and predicting if an institution different from the object of study to this research, complies with patterns and characteristics necessary to be part of this group of institutions of excellence and quality. This allows society and users in general to analyse and select the institutions with the best quality performance and financial operational management that are part of the service provision network and meet the conditions of excellence.

As a unique and practical contribution, this research provides the national and international community with scientific knowledge, materialised in a quantitative methodology that allows classifying, forecasting and predicting excellent health institutions that meet high-quality standards and good financial performance. The above through the articulation of knowledge associated with multivariate calculation, specifically cluster analysis, neural networks and high-quality accreditation in health service institutions, which ultimately allowed for a practical method. This allows the selection of the IPS by all interest groups or interested parties, which can meet their needs as patients or users of a network providing health services, in the international context where such methodology is applied.

As a limitation of this investigation, it is possible to point out the failure to have other variables associated with the accredited IPS, which allow the appearance of other profiles or the two profiles found are modified or increased, in that case, it would change the results. As future research, researchers and the scientific community are invited to apply this two-stage method to classify, forecast and predict other business sectors. Or consider the use of other variables that allow grouping the IPS taking into account other parameters that emerge when using other variables.

References

Almost, J.M., Van Den Kerkhof, E.G., Strahlendorf, P., Caicco Tett, L., Noonan, J., Hayes, T., Van Hulle, H., Adam, R., Holden, J., Kent-Hillis, T., McDonald, M., Paré, G.C., Lachhar, K. and Silva e Silva, V. (2018) 'A study of leading indicators for occupational health and safety management systems in healthcare', *BMC Health Serv. Res.*, Vol. 18, p.296, DOI: https://doi.org/10.1186/s12913-018-3103-0.

Alolayyan, M.N., Ali, K.A.M. and Idris, F. (2013) 'Total quality management and operational flexibility impact on hospitals performance: a structural modelling approach', *Int. J. Product. Qual. Manag.*, Vol. 11, No. 2, pp.212–227.

Ansari, A. and Riasi, A. (2016) 'Modelling and evaluating customer loyalty using neural networks: evidence from startup insurance companies', *Future Bus. J.*, Vol. 2, No. 1, pp.15–30.

Askim, J., Christensen, T. and Lægreid, P. (2015) 'Accountability and performance management: the Norwegian hospital, welfare, and immigration administration', *Int. J. Public Adm.*, Vol. 38, pp.971–982, DOI: https://doi.org/10.1080/01900692.2015.1069840.

Carlucci, D., Renna, P. and Schiuma, G. (2013) 'Evaluating service quality dimensions as antecedents to outpatient satisfaction using back propagation neural network', *Health Care Manag. Sci.*, Vol. 16, No. 1, pp.37–44.

Chamboko, R. and Bravo, J.M. (2018) 'Modelling and forecasting recurrent recovery events on consumer loans', *Int. J. Appl. Decis. Sci.*, Vol. 12, No. 3, pp.271–287, DOI: 10.1504/IJADS.2019.100440.

Chojaczyk, A.A., Teixeira, A.P., Neves, L.C., Cardoso, J.B. and Guedes Soares, C. (2015) 'Review and application of artificial neural networks models in reliability analysis of steel structures', *Struct. Saf.*, Vol. 52, pp.78–89, DOI: https://doi.org/10.1016/j.strusafe.2014.09.002.

Cong, Z., Fernandez, A., Billhardt, H. and Lujak, M. (2015) 'Service discovery acceleration with hierarchical clustering', *Inf. Syst. Front.*, Vol. 17, No. 4, pp.799–808.

Cruz, P.P. and Herrera, A. (2011) *Inteligencia Artificial con Aplicaciones a la Ingeniería*, Vol. 1, Marcombo, Mexico DF, Mexico.

Dammaj, A., Alawneh, A., Hammad, A.A. and Sweis, R.J. (2016) 'Investigating the relationship between knowledge sharing and service quality in private hospitals in Jordan', *Int. J. Product. Qual. Manag.*, Vol. 17, pp.437, DOI: https://doi.org/10.1504/IJPQM.2016.075248.

De la Garza, J., Morales, B. and González, B. (2013) *Análisis Estadístico Mutivariante, Un Enfoque Teórico y Práctico*, pp.150–178, McGraw Hill, México DF, México.

Decreto Único Reglamentario 780 de 2016 [WWW Document] (2016) [online] https://www.minsalud.gov.co/Normativa/Paginas/decreto-unico-minsalud-780-de-2016.aspx (accessed 24 October 18).

Dweiri, F., Khan, S.A. and Jain, V. (2015) 'Production planning forecasting method selection in a supply chain: a case study', *Int. J. Appl. Manag. Sci.*, Vol. 7, pp.38, DOI: https://doi.org/10.1504/IJAMS.2015.068056.

Fontalvo Herrera, T.J., Mendoza Mendoza, A.A., Cadavid, V. and Delimiro, A. (2016) 'Evaluación del comportamiento de los indicadores de productividad y rentabilidad en las empresas prestadores de salud del Régimen Contributivo en Colombia', *Rev. Salud Uninorte*, Vol. 32, No. 3, pp.419–428.

Fontalvo, T., De La Hoz, E. and De La Hoz, E. (2018) 'Data envelopment analysis method and neural networks in the evaluation and prediction of the technical efficiency of small exporting companies [Método análisis envolvente de datos y redes neuronales en la evaluación y predicción de la eficiencia técnica de pequeñas empresas exportadoras]', *Inf. Tecnol.*, Vol. 29, pp.267–276 [online] https://doi.org/10.4067/S0718-07642018000600267.

Fontalvo-Herrera, T.J., Delahoz, E.J. and Mendoza-Mendoza, A.A. (2018) 'Application of data mining for the classification of university programs of industrial engineering accredited in high quality in Colombia [Aplicación de minería de datos para la clasificación de programas universitarios de ingeniería industrial acreditados en alta calidad en Colombia]', *Inf. Tecnol.*, Vol. 29, pp.89–96 [online] https://doi.org/10.4067/S0718-07642018000300089.

Forrellat Barrios, M. (2014) 'Calidad en los servicios de salud: un reto ineludible', *Rev. Cuba. Hematol. Inmunol. Hemoter.*, Vol. 30, No. 2, pp.179–183.

Gøeg, K.R., Cornet, R. and Andersen, S.K. (2015) 'Clustering clinical models from local electronic health records based on semantic similarity', *J. Biomed. Inform.*, Vol. 54, No. 1, pp.294–304, DOI: 10.1016/j.jbi.2014.12.015.

González, V.V., Valecillos, J. and Hernández, C. (2013) 'Calidad en la prestación de servicios de salud: parámetros de medición', *Revista de ciencias sociales, Facultad de Ciencias Sociales*, Vol. 19, No. 4, pp.663–671.

Granadillo, E.D.L.H., Gomez, J.M. and Herrera, T.J.F. (2019) 'Methodology with multivariate calculation to define and evaluate financial productivity profiles of the chemical sector in Colombia', *Int. J. Product. Qual. Manag.*, Vol. 27, pp.144–160, DOI: https://doi.org/10.1504/IJPQM.2019.100141.

Guerrero, R., Gallego, A.I., Becerril-Montekio, V. and Vásquez, J. (2011) 'Sistema de salud de Colombia', *Salud Pública México*, Vol. 53, No. 2, pp.s144–s155, DOI: 10.1590/S0036-36342011000800010.

Hasani, H., Jalali, S.M.J., Rezaei, D. and Maleki, M. (2018) 'A data mining framework for classification of organisational performance based on rough set theory', *Asian J Manag. Sci. Appl.*, Vol. 3, p.156, DOI: https://doi.org/10.1504/AJMSA.2018.091020.

Hernández, M., Hernández, A. and Bringas, N. (2013) 'El contexto actual de la calidad en salud y sus indicadores', *Rev Mex Med Fis Rehab*, Vol. 25, No. 1, pp.26–33.

Jahantigh, F.F. (2019) 'Evaluation of healthcare service quality management in an Iranian hospital by using fuzzy logic', *Int. J. Product. Qual. Manag.*, Vol. 26, p.160, DOI: https://doi.org/10.1504/IJPQM.2019.097764.

Kamble, R. and Wankhade, L. (2017) 'Perspectives on productivity: identifying attributes influencing productivity in various industrial sectors', *Int. J. Product. Qual. Manag.*, Vol. 22, p.536, DOI: https://doi.org/10.1504/IJPQM.2017.087868.

Kaufman, L. and Rousseeuw, P.J. (1990) 'Partitioning around medoids (program pam)', *Find. Groups Data Introd. Clust. Anal.*, No. 1, pp.68–125.

Khraisat, A., Sweis, R.J., Saleh, R., Suifan, T., Hiyassat, M. and Sarea, A. (2017) 'The assessment of service quality in private hospitals in Amman area using the gap approach', *Int. J. Product. Qual. Manag.*, Vol. 22, No. 3, pp.281–308.

Menardi, G. (2011) 'Density-based Silhouette diagnostics for clustering methods', *Stat. Comput.*, Vol. 21, pp.295–308, DOI: https://doi.org/10.1007/s11222-010-9169-0.

Mettler, T. (2013) 'Explorative clustering of clinical user profiles: A first step towards user-centered health information systems'.

Nasseri, S.H. and Kiaei, H. (2019) 'Ranking of efficient units on the basis of distance from virtual ideal and anti-ideal units', *Int. J. Appl. Decis. Sci.*, Vol. 12, p.361, DOI: https://doi.org/10.1504/IJADS.2019.102640.

Peng, X., Lin, P., Zhang, T. and Wang, J. (2013) 'Extreme learning machine-based classification of ADHD using brain structural MRI data', *PloS one*, Vol. 8, No. 11, p.e79476.

Raina, S.H., Bhat, R.L. and Dar, K.H. (2018) 'Service quality in private hospitals of Jammu and Kashmir – an empirical assessment from District Srinagar', *Int. J. Healthc. Technol. Manag.*, Vol. 17, p.197, DOI: https://doi.org/10.1504/IJHTM.2018.098390.

Ricks, B. and Ventura, D. (2003) 'Training a quantum neural network', *Proceedings of the 16th International Conference on Neural Information Processing Systems, NIPS'03*, Whistler, British Columbia, Canada, MIT Press, pp.1019–1026.

Ronao, C.A. and Cho, S-B. (2016) 'Human activity recognition with smartphone sensors using deep learning neural networks', *Expert Syst. Appl.*, Vol. 59, pp.235–244, DOI: https://doi.org/10.1016/j.eswa.2016.04.032.

Scardapane, S. and Wang, D. (2017) 'Randomness in neural networks: an overview', *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, Vol. 7, p.e1200, DOI: https://doi.org/10.1002/widm.1200.

Setiono, R. (2001) 'Feedforward neural network construction using cross validation', *Neural Computation*, Vol. 13, No. 12, pp.2865–2877.

Shaw, C.D., Braithwaite, J., Moldovan, M., Nicklin, W., Grgic, I., Fortune, T. and Whittaker, S. (2013) 'Profiling health-care accreditation organizations: an international survey', *Int. J. Qual. Health Care*, Vol. 25, pp.222–231.

Singhtaun, C. and Hattayanon, R. (2017) 'An application of quality cost analysis as a tool for quality management', *Int. J. Product. Qual. Manag.*, Vol. 22, pp.205–222.

Sreenivasan, S. and Sundaram, M. (2018) 'A probabilistic model for predicting service level adherence of application support projects', *Int. J. Product. Qual. Manag.* Vol. 25, pp.305–330, DOI: https://doi.org/10.1504/IJPQM.2018.095648.

Tosun, Ö. (2012) 'Using data envelopment analysis-neural network model to evaluate hospital efficiency', *Int. J. Product. Qual. Manag.*, Vol. 9, pp.245–257.

Tsofa, B., Molyneux, S., Gilson, L. and Goodman, C. (2017) 'How does decentralisation affect health sector planning and financial management? A case study of early effects of devolution in Kilifi County', *Kenya. Int. J. Equity Health*, Vol. 16, p.151, DOI: https://doi.org/10.1186/s12939-017-0649-0.

Vaver, J. (2014) 'Evaluating techniques for clustering geographic entities'.

Wang, S., Wang, L., Gao, S. and Bai, Z. (2017) 'Stock price prediction based on chaotic hybrid particle swarm optimisation-RBF neural network', *Int. J. Appl. Decis. Sci.*, Vol. 10, p.89, DOI: https://doi.org/10.1504/IJADS.2017.084307.

Zahar, M., Barkany, A.E. and Biyaali, A.E. (2016) 'Cost of quality in healthcare: a case study in a clinical laboratory', *Int. J. Product. Qual. Manag.*, Vol. 17, pp.536–548.