



Artificial neural networks and risk stratification in emergency departments

Greta Falavigna¹ · Giorgio Costantino² · Raffaello Furlan³ · James V. Quinn⁴ · Andrea Ungar⁵ · Roberto Ippoliti⁶

Received: 25 June 2018 / Accepted: 16 October 2018 / Published online: 23 October 2018
© Società Italiana di Medicina Interna 2018

Abstract

Emergency departments are characterized by the need for quick diagnosis under pressure. To select the most appropriate treatment, a series of rules to support decision-making has been offered by scientific societies. The effectiveness of these rules affects the appropriateness of treatment and the hospitalization of patients. Analyzing a sample of 1844 patients and focusing on the decision to hospitalize a patient after a syncope event to prevent severe short-term outcomes, this work proposes a new algorithm based on neural networks. Artificial neural networks are a non-parametric technique with the well-known ability to generalize behaviors, and they can thus predict severe short-term outcomes with pre-selected levels of sensitivity and specificity. This innovative technique can outperform the traditional models, since it does not require a specific functional form, i.e., the data are not supposed to be distributed following a specific design. Based on our results, the innovative model can predict hospitalization with a sensitivity of 100% and a specificity of 79%, significantly increasing the appropriateness of medical treatment and, as a result, hospital efficiency. According to Garson's Indexes, the most significant variables are exertion, the absence of symptoms, and the patient's gender. On the contrary, cardio-vascular history, hypertension, and age have the lowest impact on the determination of the subject's health status. The main application of this new technology is the adoption of smart solutions (e.g., a mobile app) to customize the stratification of patients admitted to emergency departments (ED)s after a syncope event. Indeed, the adoption of these smart solutions gives the opportunity to customize risk stratification according to the specific clinical case (i.e., the patient's health status) and the physician's decision-making process (i.e., the desired levels of sensitivity and specificity). Moreover, a decision-making process based on these smart solutions might ensure a more effective use of available resources, improving the management of syncope patients and reducing the cost of inappropriate treatment and hospitalization.

Keywords Decision processes · Artificial neural networks (ANNs) · Risk stratification · Syncope · Emergency departments (ERs)

Introduction

The emergency department (ED) is a key component of hospital activities, supplying health treatment and services to patients. The physicians' activities involve making a rapid provisional diagnosis, and choosing the most appropriate treatment. To help physicians to carry out a rational evaluation in a stressful environment, a series of support decision-making rules have been offered by scientific societies. This decision-making process might have a significant economic impact [1–3], and a trade-off between efficiency and effectiveness might arise [4]. On one hand, physicians may be more interested in stratifying patients based on the type of classification chosen (i.e., expected levels of true positives and false negatives). In other words, sensitivity should drive

✉ Greta Falavigna
greta.falavigna@ircres.cnr.it

¹ CNR-IRCRES, Research Institute on Sustainable Economic Growth, Moncalieri, Italy

² Clinical Medicine Department, Fondazione IRCCS, Ca' Granda, Ospedale Maggiore Policlinico, Milan, Italy

³ Department of Biomedical Sciences, Humanitas University, Humanitas Research Hospital, Rozzano, Italy

⁴ Division of Emergency Medicine, Stanford University, Stanford, CA, USA

⁵ Syncope Unit, Geriatric Medicine and Cardiology, Careggi University Hospital, Florence, Italy

⁶ Department of Management, University of Turin, Turin, Italy

Table 1 Variables and functioning of the main stratification rules adopted in EDs [19]

Rule	Variables	Functioning
SFSR	Congestive heart failure, history of: Hematocrit < 30% ECG, abnormal Shortness of breath Systolic blood pressure at triage < 90 mmHg	Patient at high risk if a single variable is present
OESIL	History of cardio-vascular disease Abnormal ECG Age > 65 years Absence of prodromal symptoms	Single factor counts as 1 Sum ≤ 1 (low risk) Sum > 1 (high risk)

the setting of every stratification rule, properly supporting the decision-making process in EDs [5]. On the other hand, the hospital management may be more interested in misclassification (i.e., expected levels of true negatives and false positives), which can affect the appropriateness of treatment and the hospitalization of patients, thus conditioning hospital efficiency [6, 7]. Moreover, the appropriateness of treatment has a large impact upon the quality of hospitals as perceived and assessed by patients and policy makers [8–10]. This is even more relevant if we consider the current period of global austerity and widespread spending cuts in the healthcare sector [11], which might lead hospital managers to adopt innovative solutions to face budget constraints [12–14].

In this work, the authors focus on a specific type of decision-making problem, i.e., whether to hospitalize a patient after a syncope event, to prevent subsequent severe short-term outcomes [15–17]. Several rules have been proposed to support physicians' decision-making in this context, such as the *San Francisco Syncope Rule* (SFSR) [18] and that of *Osservatorio Epidemiologico sulla Sincope del Lazio* (OESIL) [19]. The OESIL risk score is the simple arithmetic sum of the number of independent end-point predictors found for every single patient. This score is timesaving, because it can be easily calculated during triaging in EDs. The SFSR is structured as a kind of decision tree, with a series of questions about the clinical condition of the patient, and, at the end of the process, the physician is able to stratify the risk of short-term severe outcomes. Table 1 offers an overview of these rules, highlighting the adopted variables and their functioning.

Using multivariate logistic regression models, several studies have sought to validate these scores, resulting in different sensitivity and specificity values according to the selected sample of observations. Considering the SFSR, two studies have been conducted on the same sample (i.e.,

684 patients), but assessing different variables as predictors of serious events after syncope. In detail, the former obtains a sensitivity of 96% (confidence interval 92–100%) and a specificity of 62% (CI 58–66%) [20], while the latter achieves an apparent improvement in performance, i.e., a sensitivity of 86% (CI 71–94%) and a specificity of 70% (CI 66–74%) [21]. According to the results collected in another study on the same sample of patients, the OESIL score shows a sensitivity of 88% (CI 70–98%) and a specificity of 60% (CI 55–64%), while the SFSR has a sensitivity of 81% (CI 61–93%) and a specificity of 63% (CI 58–67%) in predicting adverse outcomes within 10 days [22].

The current literature proposes an alternative approach to the classic multivariate logistic regression models: artificial neural networks (ANNs). This innovative technique can outperform the traditional models, since it does not require a specific functional form, i.e., the data are not supposed to be distributed following a specific design. Indeed, multivariate logistic regression models require specific functional forms (i.e., parametric models), which might significantly affect the final results [4, 15]. In other words, neural networks can outperform the classic models, since they can adapt to the initial data sample, showing greater generalization ability.

Against this background, the main goal of our study is to propose a new algorithm based on artificial neural networks (ANNs), which, compared to the traditional logistic regression models, is able to stratify the risk of severe outcomes after a syncope spell with greater accuracy [20–22]. More precisely, we have elaborated an innovative algorithm through which the physician can select the expected levels of sensitivity and specificity (e.g., sensitivity of 100% and specificity of 79%). Based on this, the model then estimates the weights of the different variables, making it possible to predict severe short-term outcomes after a syncope event with the expected level of accuracy.

Data and methodology

Data

The data used in the analysis are taken from three different studies, providing information about 1844 patients for our case study on the stratification of severe short-term outcomes [20, 23, 24]. These outcomes include death, the need for major therapeutic procedures (i.e., cardiopulmonary resuscitation, pacemaker or implantable cardioverter-defibrillator insertion, intensive-care unit admittance, and acute antidysrhythmic therapy), and early (within 10 days) readmission to hospital due to similar symptoms. In particular, according to our data sources, major therapeutic procedures refer to those procedures undertaken after the ED patient is hospitalized or discharged. Table 2 shows key information

Table 2 Descriptive statistics of the selected sample of patients according to the data sources

Source	No. of patients	Female (%)	Male (%)	Total (%)	Severe short-term outcomes (%)
First study [20]	684	21.85	15.24	37.09	37.16
Second study [23]	695	21.10	16.59	37.69	20.18
Third study [24]	465	11.50	13.72	25.22	42.66
Total	1844	54.45	45.55	100.00	100.00

Table 3 Descriptive statistics of the available variables

Variables	Female		Male	
	< 65 (%)	> 65 (%)	< 65 (%)	> 65 (%)
Syncope during exertion	4.03	4.41	4.07	5.63
Trauma following syncope	15.25	20.11	17.89	21.00
Abnormal electrocardiography	10.38	43.10	21.14	51.73
History of cerebro-vascular disease	6.36	36.21	13.82	48.70
History of cardio-vascular disease	5.93	33.91	14.09	43.94
Absence of symptoms	31.57	30.84	27.10	35.06
Hypertension	11.02	55.94	21.95	49.57
History of syncope	43.64	37.16	37.13	36.15
Severe short-term outcomes	3.18	13.41	7.59	20.35

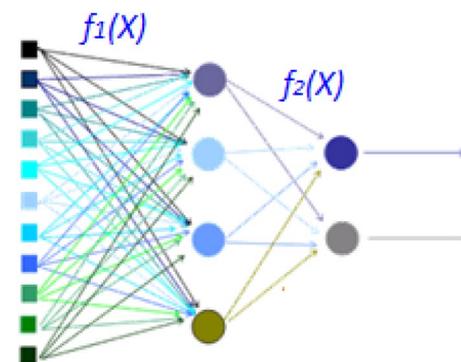
about the above studies and some preliminary descriptive statistics of the sample of patients.

The information included in the whole database refers to physical and biological characteristics of the patients (e.g., gender, age, syncope during exertion, trauma following syncope, abnormal electrocardiography, history of cardio-vascular disease, history of cerebro-vascular disease, absence of prodrome before syncope, and previous history of syncope). Nevertheless, not all the variables were available for all the patients, and, after the data sets were merged, total observations decreased to 1825. To reduce potential weakness of representation, a bootstrap procedure was applied to the initial sample [25, 26]. Table 3 presents detailed descriptive statistics of the selected sample of patients, showing the available variables according to gender and age. As stated in the current literature [23], these variables are statistically significant risk factors for severe short-term outcomes (within 10 days).

In this work, we focus on the above factors, comparing different stratification rules adopted by physicians in EDs with the innovative methodology, i.e., ANNs.

Artificial neural networks (ANNs) as new stratification rule

Artificial neural networks (ANNs) are complex models organized in layers (multilayer) formed by neurons (also called perceptrons) interconnected via synapses (weights). Due to their well-known ability to generalize behaviors [27],

**Fig. 1** Feed-forward MultiLayer Perceptron framework [45]

ANNs have been successfully applied to many fields, such as urinary tract infections and celiac disease [28, 29], acute myocardial infarction and hemodynamic changes [30, 31], as well as in EDs to stratify patients [4, 15, 32–34].

As highlighted in Fig. 1, the first layer is called “input layer”, and it is composed of a number of neurons (or nodes) equal to that of the variables analyzed (in our specific case, as many neurons as the available pieces of information on the patients). The last layer is the “output layer”, from which the result of model is derived. The number of nodes in this layer depends on the type of answer expected. Typically, there is only one neuron, because the result is expressed in dichotomous form. Between the input layer and the output layer, there are hidden layers, which can be more than one.

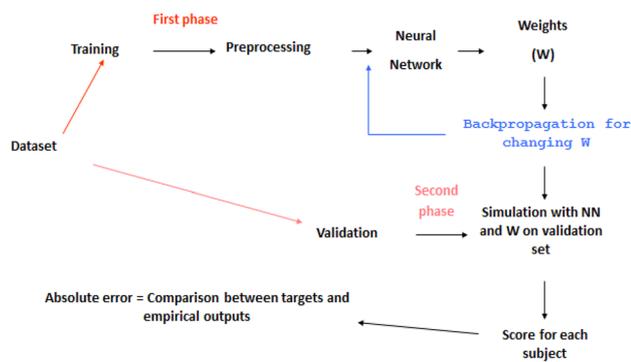


Fig. 2 Supervised learning of MLP with back-propagation algorithm [45]

The literature suggests that a single hidden layer can approximate any functional form [31, 35]. The number of neurons in the hidden layers must be found empirically [36, 37], although some authors have tried to define specific rules. For instance, some researchers suggest using the formula $(2i + 1)$ where $i = 1, \dots, I$ represents the number of variables considered [38–40]. A better performing and less time-consuming criterion has been validated, and it uses the proportion $0.75i$ [41, 42]. The links between the layers are the “synapses”, mathematically called weights, which collect information about the relationships between the input variables and the expected outputs. These relationships are formalized through functions that are, in the majority of cases, non-linear (e.g., logsigmoidal, tansigmoidal, hardlim, and so on). The activation functions used in this model are linear from the input layer to the hidden one, and tansigmoidal from the hidden layer to the output one.¹ The relationships between the layers are collected in weight matrixes, and their analysis makes it possible to evaluate the contribution of each piece of information to the definition of the expected outputs [39, 43]. The MultiLayer Perceptron (MLP) network is represented in Fig. 1, and its links are feed-forward, because the connections come from the input layer to the hidden one and from the hidden layer to the output one. Backward or recursive relationships are not considered in this framework. The feed-forward MultiLayer Perceptron works with a supervised learning technique through a back-propagation algorithm. Note that there are some network frameworks that use an unsupervised procedure, i.e., Self-Organizing Map or Kohonen networks [44].

Figure 2 shows how supervised learning works with the back-propagation algorithm. The initial sample is divided into two sub-samples: the training sample and the validation sample. In the first phase, only elements from the training

sample are introduced into the model, and, through the back-propagation algorithm, the network attempts to minimize the mean-square output error over the entire training set. The ANN computes weights matrixes until a predefined error threshold is reached. In this step, the model is fed information about the patients, but also about the “target”, i.e., their health status (severe short-/long-term outcomes). This stage is very important, because the ANN learns from the data, and collects, within weight matrixes, information about the relationships between the variables. It clearly emerges that dividing the initial sample into training and validation is critical: the training set must represent all possible types of patients with their specific characteristics. In our case, two proportions are set: 4/5 in the training sample and 1/5 in the validation one (first case); 9/10 in the training sample; 1/10 in the validation one (second case). Once the weights and ANN framework are defined (i.e., type of activation functions between layers; number of hidden layers and their nodes; other technical parameters, such as the search function for the optimal gradient, etc.), these parameters are applied to the validation sample, which is introduced into the ANN without any information on severe short-/long-term outcomes. The ANN applies the framework to the new data to evaluate results and the ability of the model to provide correct classification.

In our model, information about the patients is introduced into the input layer with the aim of obtaining an outcome for each patient, indicating whether severe short- or long-term outcomes are likely (1 if they are; 0 if not).

Sensitivity–specificity search algorithm

According to the current literature [5], sensitivity is defined as true positives/(true positives + false negatives), whereas specificity as true negatives/(true negatives + false positives). The positive predictive value is defined as true positives/(true positives + false positives), whereas the negative predictive value is defined as true negatives/(true negatives + false negatives). When initiating the algorithm, physicians can set the desired values of sensitivity and specificity. In this manner, the complex model trains itself, and obtains parameters based on the expected performance, while, up to now, model performance has been measured considering the instances of correct classification [6]. This approach meets the physicians’ expectations, since, by means of the algorithm presented here, the model parameters are estimated starting from preset sensitivity and specificity values. In detail, the authors have chosen to set sensitivity at 90% and specificity at 85%, with a validation sample of 1/10, and sensitivity at 100% and specificity at 80%, with a validation sample of 1/5.

With the purpose of finding the best results in terms of errors, the first version of the algorithm [45] has been

¹ The tansigmoidal function has a logsigmoidal form with a codomain range from -1 to $+1$.

changed to attain the optimal threshold that minimizes errors, and allows sensitivity and specificity values to converge toward those chosen by the physicians. Specifically, a threshold vector, ranging from 0 to 1 with a step of 0.001, has been defined, and the model results (i.e., true positive, false negative, false positive, true negative, sensitivity, and specificity) have been computed for each element of this threshold. Then, the empirical values of sensitivity and specificity have been compared to the expected ones. The following sensitivity vector ($\text{diff_sens}_{i,1}$, where i represents the value of the tested threshold) shows an example of the difference in absolute value between the expected value and the empirical one. The same vector has been calculated for specificity. This indicates a sort of “error” of the result compared to the expected value, so the algorithm computes the minimum value of this vector and memorizes its position, i.e., the corresponding threshold:

$$\text{diff_sens}_{i,1} = \begin{bmatrix} \text{expected_sens} - \text{empirical_sens}_1 \\ \text{expected_sens} - \text{empirical_sens}_2 \\ \text{expected_sens} - \text{empirical_sens}_3 \\ \vdots \\ \text{expected_sens} - \text{empirical_sens}_i \end{bmatrix}. \quad (1)$$

It is clear that the optimum threshold for sensitivity is often not the same as that for specificity, so the fundamental question is: how can the optimum threshold be chosen? As a first step, the algorithm computes the mean value for each observation (i.e., patient) of the variables diff_sens and diff_spec . In formal terms, we have:

$$\text{diff_sens}_{i,1} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_i \end{bmatrix} \quad \text{diff_spec}_{i,1} = \begin{bmatrix} dd_1 \\ dd_2 \\ \vdots \\ dd_i \end{bmatrix} \quad (2)$$

$$\text{diff_sens}_{i,1} = \begin{bmatrix} (d_1 + dd_1)/2 \\ (d_2 + dd_2)/2 \\ \vdots \\ (d_i + dd_i)/2 \end{bmatrix}.$$

The lower level of the vector $\text{diff_mean}_{i,1}$ identifies the best joint values of sensitivity and specificity, and their linked threshold.

Once the optimum value of the threshold is established, the neural network model runs as explained in the above section, and following the optimization algorithm previously proposed [45], considering a single threshold value and computing neural weight matrixes. Concerning computational issues, a bootstrap procedure is used in this study to improve the robustness of estimates. Moreover, since, in the training phase, ANN weights are randomly

set every time, an algorithm allows us to run the ANN for a defined number of times (i.e., replications).

The algorithm presented here runs in the training phase and yields network parameters ensuring levels of sensitivity and specificity defined at the beginning of the computation.

Thus, when a physician introduces new patient data into the model, the answer obtained about a potential future negative event is specifically based on set levels of sensitivity and specificity.

Applicability of the innovative technology to EDs

The main application of this new technology is the adoption of smart solutions (e.g., a mobile app) to customize the stratification of patients admitted to EDs after a syncope event. On one hand, clinical information about the patients is collected during triage and then processed by the algorithm based on ANNs to support the decision-making process regarding hospitalization and specialist investigations. On the other hand, physicians can use a dedicated mobile app to stratify the patients based on their clinical records, collected during triage, and preset the levels of sensitivity and specificity. The adoption of these smart solutions gives the opportunity to customize risk stratification according to the specific clinical case (i.e., the patient’s health status) and the physician’s decision-making process (i.e., the desired levels of sensitivity and specificity). Obviously, by implementing new algorithms in different contexts (e.g., to stratify chest pain linked to suspected cardiac problems), this app can be extended to the stratification of other risks to comprehensively support clinical decision-making in EDs. Finally, direct collection of data about incoming patients during triage makes it possible to gather new evidence to refine the algorithm, so that updated versions of our innovative technology will become ever more effective at every access. In other words, we can provide an effective learning decision-making system for EDs.

The implementation of this solution would require a framework based on the *Internet of Things* (IoT) for the exchange of information among healthcare actors in hospitals (i.e., triage personnel and physicians). A web-based dashboard and a decision-making support system based on ANNs would display the results of the customized risk stratification. From a technical point of view, the proposed algorithm is built and tested using *Matlab code*, which can then be converted into the *C programming language* and integrated into *iOS* and *Android apps*. Alternatively, we might develop the algorithm in *R-stats*, writing the app in *Java* or *Python*.

Table 4 Results on validation sample comparing ANN, OESIL, and SFSR

	ANN		OESIL		SFSR	
	1/5	1/10	1/5	1/10	1/5	1/10
Tot. subjects	1825	1825	1825	1825	492	492
No. training sample	1460	1642	1460	1642	240	348
No. validation sample	365	183	365	183	252	144
Sensitivity	93%	100%	76%	81%	83%	88%
CI (95%)	80–98%	83–100%	60–88%	58–95%	61–95%	62–98%
Specificity	67%	79%	67%	77%	69%	76%
CI (95%)	61–72%	72–85%	62–72%	70–83%	63–75%	67–83%
Total errors (TEs)	109	34	116	41	74	33
(%) TEs	29.86%	18.58%	31.78%	22.40%	29.37%	22.92%
AUC	0.798	0.895	0.714	0.791	0.76	0.816
Likelihood ratio (LR –)	0.109	0	0.362	0.247	0.25	0.165
Likelihood ratio (LR +)	2.81	4.76	2.31	3.54	2.7	3.61

Results

Table 4 presents comparative results in terms of sensitivity, specificity, number of total errors (TEs), and percentage of total errors (% TEs). The comparison is made between the proposed innovative model (i.e., ANN), and the approaches currently adopted (i.e., OESIL and SFSR), using the same sample of patients (considering both validation sub-samples, i.e., 1/5 and 1/10) and the available variables. In particular, the authors have calculated the performance indexes on the patients of the ANN validation sample. Only in the case of the SFSR, the scores are not available for all the subjects of the validation sample, so the number of patients does not correspond.

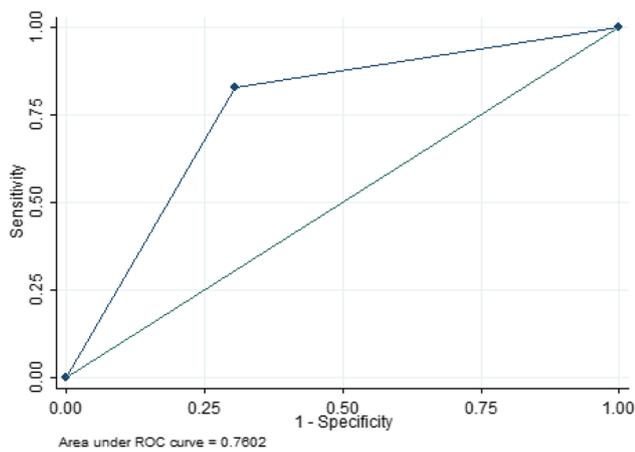
In terms of sensitivity, the ANN presents significantly higher values for both cases of validation (i.e., 93% and 100%), while, as for specificity, the ANN still performs well. These results are clearly due to the sensitivity–specificity algorithm, that is to say, the opportunity for physicians to set performance levels which that might be able to meet their expectations (i.e., minimize the number of false negatives). Looking at total errors, the ANN displays a lower percentage compared to the OESIL and SFSR methods when the validation is 1/5 of the whole sample, whereas the number of total errors is at its lowest when the proportion is 1/10. Obviously, these results are very sensitive to the size and quality of the data used in the training set. Indeed, when a smaller validation sample is considered, the performance of all the scores increases (i.e., proportion of the validation sample equal to 1/10).

In addition, Table 4 presents the values of the area under the curve (AUC), considering indexes and different validation proportions. These values indicate the area under the ROC (receiver-operating characteristics) curve, which represents the relation between specificity and 1 – sensitivity. In Fig. 3, the ROC curves are presented for both the ANN

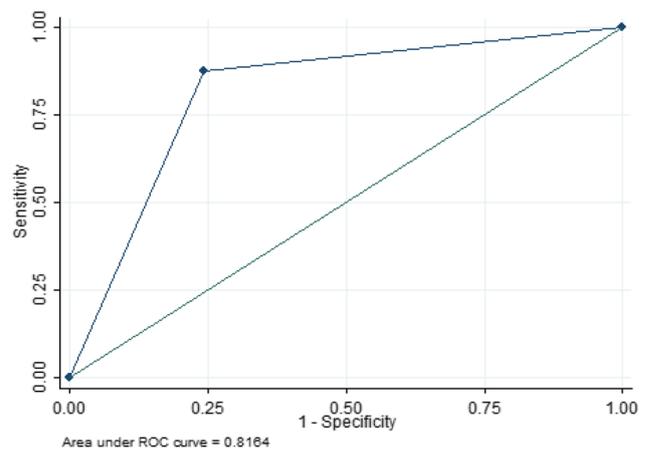
and the OESIL scores. The ROC curves are displayed in the same figures, distinguished only by the different validation proportion; instead, the SFSR is shown in separate graphs due to the different sample size (i.e., 495 instead of 1825). Note that the graphs are not strictly continuous curves but piecewise linear curves, since the outputs are dichotomous variables.

The last two rows of Table 4 report the negative likelihood ratio (LR –), i.e., the ratio between (1 – sensitivity)/specificity, and the positive likelihood ratio (LR +), obtained comparing sensitivity and (1 – specificity) [46, 47]. According to the current literature, a taxonomy of LR – and LR + identifies the predictive ability of the model [48, 49]. Considering the negative likelihood ratio, the lower the value, the more useful the model is. In particular, for results between 0.1 and 0.2, the test is moderately useful. Opposite considerations are drawn when taking LR + into account as, in this case, performance must exceed the unit value (i.e., 1 – “indifference cut-off”). In detail, results between 2 and 5 are only occasionally useful [48, 49]. To sum up, results in which LR + is greater than 1 are likely to yield the correct diagnosis; indeed, the bigger the number, the more convincingly the result points to a certain diagnosis [48]. Results of LR—varying between 0 and 1—do not support the diagnosis of interest; the closer LR is to 0, the less likely the disease. When the result is 1, the test is useless.

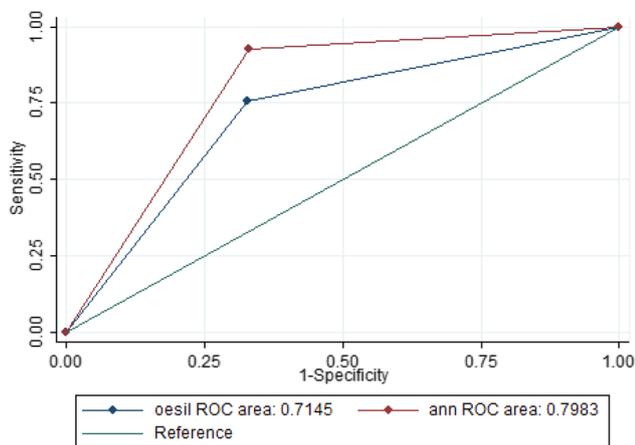
This study also looks at Garson’s Indexes, deriving from the analysis of the weight matrixes of the ANN (results shown in Table 5). Through an analysis of the weight matrixes obtained by running the neural network, a weight percentage is estimated for each variable included in the definition of the output (the healthy status). In detail, the most significant items (not in a statistical sense) are exertion, the absence of symptoms, and the patient’s gender. On the contrary, cardio-vascular history, hypertension, and age are the variables with the lowest impact on the determination of



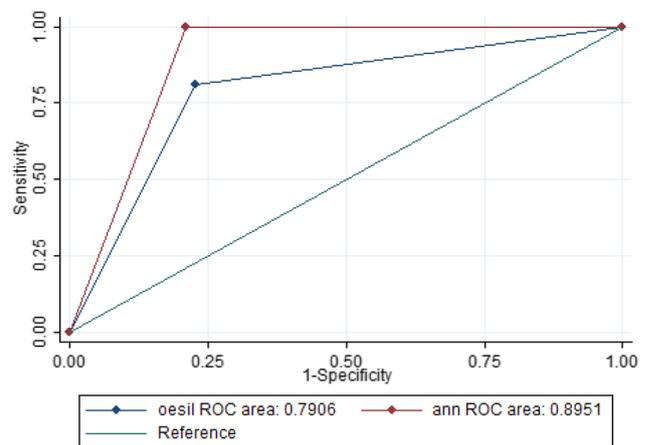
SFSR (validation equal to 1/5)



SFSR (validation equal to 1/10)



OESIL and ANN (validation equal to 1/5)



OESIL and ANN (validation equal to 1/10)

Fig. 3 ROC curves

Table 5 Garson’s indexes for the ANN

Variables	Weights (Garson’s Index) (%)
Sex	15.89
Age	5.53
Syncope during exertion	18.15
Trauma following syncope	13.04
Abnormal electrocardiography	6.40
History of cerebro-vascular disease	8.74
History of cardio-vascular disease	3.03
Absence of symptoms	17.97
Hypertension	4.92
History of syncope	6.35

the subject’s health status. These weights can be regarded as an indication of how much each variable contributes to determining severe outcomes.

Limits

Some limitations of this study should be acknowledged. First, the database used was compiled for different purposes several years ago (i.e., 2004 and 2008) and it includes different end points and clinical variables. Moreover, there is heterogeneity in health system organizations, data collection methods, and ECG interpretation. For this reason, although our data are encouraging, the use of ANNs would require more investigation before they can be extensively applied in clinical practice. Further studies, comparing ANNs with

both existing risk stratification tools and clinical judgment, are needed to detect any potential hidden weaknesses of this innovative technique [15].

Conclusions

The need to classify patients is a critical issue and complex systems can be a useful methodology [50, 51]. This is even more important in the healthcare sector and, particularly, in EDs. The primary goal of physicians working in these units is to discriminate between individuals at low risk, who can be safely discharged, and patients at high risk, who require prompt hospitalization to be monitored and given appropriate treatment [4, 15–17]. Considering patients in emergency departments after a syncope event, this work proposes a comparative analysis between the stratification rules currently used (which are based on multivariate logistic regression models) and artificial neural networks (ANNs), highlighting differences in the correct classification of severe outcomes at 10 days (from the index event). According to our results, there are good opportunities to successfully implement alternative risk stratification tools based on ANNs, adopting smart solutions (e.g., a mobile app) to customize the stratification of patients admitted to emergency departments after a syncope event. The adoption of these smart solutions would make it possible to customize the stratification risk according to the specific clinical case and the physician's decision-making process. Obviously, these opportunities may have significant impacts in terms of appropriateness of treatment and, consequently, in terms of hospital management and costs. Indeed, the correct classification of patients is fundamental for subsequent hospital admission and medical treatment. A decision-making process based on ANNs might represent an improvement for the healthcare sector, ameliorating the management and rate of admission to emergency departments of syncope patients, reducing the cost of inappropriate treatments and hospitalization, and ensuring a more effective use of resources.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Statement of human and animal rights This is a retrospective analysis on anonymous data, previously collected by the hospitals, without formal ethical approval according to national law.

Informed consent Patients' informed consents have been collected by the hospitals before treatments and clinical procedures.

References

1. Cournane S, Conway R, Creagh D, Byrne DG, Silke B (2015) Consultant duration of clinical practice as a cost determinant of an emergency medical admission. *Eur J Health Econ* 2015(16):561–567. <https://doi.org/10.1007/s10198-014-0619-z>
2. Eriksen BO, Førde OH, Kristiansen IS et al (2000) Cost savings and health losses from reducing inappropriate admissions to a department of internal medicine. *Int J Technol Assess Health Care* 16(4):1147–1157
3. Lee A, Clarke B, Hazlett SC et al (2003) How to minimize inappropriate utilization of Accident and Emergency Departments: improve the validity of classifying the general practice cases amongst the A&E attendees. *Health Policy* 66(2):159–168. [https://doi.org/10.1016/S0168-8510\(03\)00023-X](https://doi.org/10.1016/S0168-8510(03)00023-X)
4. Casagrande I, Costantino G, Falavigna G, Furlan R, Ippoliti R (2016) Artificial Neural Networks and risk stratification models in Emergency Departments: the policy maker's perspective. *Health Policy* 120(1):111–119
5. Weingart S, Wyer P (2006) *Emergency medicine decision making: critical choices in chaotic environments*. McGraw-Hill, New York
6. McDonagh MS, Smith DH, Goddard M (2000) Measuring appropriate use of acute beds. A systematic review of methods and results. *Health Policy* 53(3):157–184. [https://doi.org/10.1016/S0168-8510\(00\)00092-0](https://doi.org/10.1016/S0168-8510(00)00092-0)
7. Fellin G, Apolone G, Tampieri A et al (1995) Appropriateness of hospital use: an overview of Italian studies. *Int J Qual Health Care* 7(3):219–225
8. Propper C, Burgess S, Green K (2004) Does competition between hospitals improve the quality of care? Hospital death rates and the NHS internal market. *J Public Econ* 88(7–8):1247–1272
9. Chua CL, Palangkaraya A, Yong J (2010) A two-stage estimation of hospital quality using mortality outcome measure: an application using hospital administrative data. *Health Econ* 19:1404–1424. <https://doi.org/10.1002/hec.1560>
10. Ippoliti R, Falavigna G (2012) Efficiency of medical care industry: evidence from the Italian regional system. *Eur J Oper Res* 217(2012):643–652. <https://doi.org/10.1016/j.ejor.2011.10.010>
11. Quaglio GL, Karapiperis T, Van Woensel L et al (2013) Austerity and health in Europe. *Health Policy* 113(1):13–19. <https://doi.org/10.1016/j.healthpol.2013.09.005>
12. Ippoliti R, Falavigna G, Grosso F, Maconi A, Randi L, Numico G (2018) The economic impact of clinical research in an Italian public hospital: the malignant pleural mesothelioma case study. *Int J Health Policy Manag* 7(8):728–737. <https://doi.org/10.1517/ijhpm.2018.13>
13. De Vos P, Orduñez-García P, Santos-Peña M et al (2010) Public hospital management in times of crisis: lessons learned from Cienfuegos, Cuba (1996–2008). *Health Policy* 96(1):64–71. <https://doi.org/10.1016/j.healthpol.2010.01.005>
14. Ippoliti R, Allievi I, Falavigna G, Rizzi S, Moda G (2018) The sustainability of a community nurses programme aimed at supporting active ageing in mountain areas. *Int J Health Plan Manag*. <https://doi.org/10.1002/hpm.2591> (forthcoming)
15. Costantino G, Falavigna G, Solbiati M, Furlan R, Ippoliti R (2017) Neural networks as a tool to predict syncope risk in the Emergency Department. *Europace* 19(11):1891–1895. <https://doi.org/10.1093/europace/euw336>

16. Costantino G, Sun B, Barbic F et al (2015) Syncope clinical management in the ED: a consensus from the first international workshop on syncope risk stratification in the ED. *Eur Heart J*. <https://doi.org/10.1093/eurheartj/ehv378>
17. Sun B, Costantino G, Barbic F et al (2014) Priorities for emergency department syncope research. *Ann Emerg Med* 64(6):649–655. <https://doi.org/10.1016/j.annemergmed.2014.04.014>
18. Quinn J, McDermott D, Stiell I, Kohn M, Wells G (2006) Prospective validation of the San Francisco Syncope Rule to predict patients with serious outcomes. *Ann Emerg Med* 47(5):448–454
19. Costantino G, Furlan R (2013) Syncope risk stratification in the emergency department. *Cardiol Clin* 31(2013):27–38. <https://doi.org/10.1016/j.ccl.2012.10.003>
20. Quinn JV, Stiell IG, McDermott DA, Sellers KL, Kohn MA, Wells GA (2004) Derivation of the San Francisco Syncope Rule to predict patients with short-term serious outcomes. *Ann Emerg Med* 43(2):224–232
21. Quinn J, McDermott D (2011) Electrocardiogram findings in emergency department patients with syncope. *Acad Emerg Med* 18(7):714–718
22. Dipaola F, Costantino G, Perego F et al (2010) San Francisco syncope rule, Osservatorio epidemiologico sulla sincope nel lazio risk score, and clinical judgment in the assessment of short-term outcome of syncope. *Am J Emerg Med* 2010(28):432–439
23. Costantino G, Perego F, Dipaola F et al (2008) Short and long-term prognosis of syncope, risk factors, and role of hospital admission. *J Am Coll Cardiol* 51(30):276–283
24. Del Rosso A, Ungar A, Maggi R et al (2008) Clinical predictors of cardiac syncope at initial evaluation in patients referred urgently to a general hospital: the EGSYS score. *Heart* 94:1620–1626
25. Efron B, Tibshirani RJ (1994) *An introduction to the bootstrap*. CRC Press, Boca Raton
26. Trichakis I, Nikolos I, Karatzas GP (2011) Comparison of bootstrap confidence intervals for an ANN model of a karstic aquifer response. *Hydrol Process* 25(18):2827–2836
27. Falavigna G (2008) New contents and perspectives in the risk analysis of enterprises. *Int J Bus Perform Manag* 10(2/3):136–173
28. Heckerling PS, Canaris GJ, Flach SD, Tape TG, Wigton RS, Gerber BS (2007) Predictors of urinary tract infection based on artificial neural networks and genetic algorithms. *Int J Med Inform* 76(4):289–296
29. Tenório JM, Hummel AD, Cohrs FM, Sdepanian VL, Pisa IT, de Fátima Marin H (2011) Artificial intelligence techniques applied to the development of a decision–support system for diagnosing celiac disease. *Int J Med Inform* 80(11):793–802
30. Ellenius J, Groth T (2000) Methods for selection of adequate neural network structures with application to early assessment of chest pain patients by biochemical monitoring. *Int J Med Inform* 57(2–3):181–202
31. Parmanto B, Deneault LG, Denault AY (2001) Detection of hemodynamic changes in clinical monitoring by time-delay neural networks. *Int J Med Inform* 63(1–2):91–99
32. Baxt WG, Shofer FS, Sites FD, Hollander JE (2002) A neural network aid for the early diagnosis of cardiac ischemia in patients presenting to the emergency department with chest pain. *Ann Emerg Med* 40(6):575–583
33. Harrison RF, Kennedy RL (2005) Artificial neural network models for prediction of acute coronary syndromes using clinical data from the time of presentation. *Ann Emerg Med* 46(5):431–439
34. Bektaş F, Eken C, Soyuncu S et al (2008) Artificial neural network in predicting craniocervical junction injury: an alternative approach to trauma patients. *Eur J Emerg Med* 15(6):318–323. <https://doi.org/10.1097/MEJ.0b013e3282fcef7af>
35. Hornik K, Stinchcombe M, White H (1989) Multilayer feed-forward networks are universal approximators. *Neural Netw* 2:359–366
36. Kim KJ (2003) Financial time series forecasting using support vector machines. *Neurocomputing* 55(1):307–319
37. Min JH, Lee YC (2005) Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Syst Appl* 28(4):603–614
38. Patuwo E, Hu MY, Hung MS (1993) Two-group classification using neural networks. *Decis Sci* 24(4):825–845
39. Nath R, Rajagopalan B, Ryker R (1997) Determining the saliency of input variables in neural network classifiers. *Comput Oper Res* 24(8):767–773
40. Chauhan N, Ravi V, Karthik Chandra D (2009) Differential evolution trained wavelet neural networks: application to bankruptcy prediction in banks. *Expert Syst Appl* 36(4):7659–7665
41. Salchenberger LM, Cinar E, Lash NA (1992) Neural networks: a new tool for predicting thrift failures. *Decis Sci* 23(4):899–916
42. Olmeda I, Fernández E (1997) Hybrid classifiers for financial multicriteria decision making: the case of bankruptcy prediction. *Comput Econ* 10(4):317–335
43. Garson GD (1991) Interpreting neural-network connection weights. *AI Expert* 6(4):46–51
44. Kohonen T (1990) The self-organizing map. *Proc IEEE* 78(9):1464–1480
45. Falavigna G (2012) Financial ratings with scarce information: a neural network approach. *Expert Syst Appl* 39(2):1784–1792. <https://doi.org/10.1016/j.eswa.2011.08.074> (ISSN: 0957-4174)
46. Reid MC, Lane DA, Feinstein AR (1998) Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 104(4):374–380
47. Gardner M, Altman DG (2000) *Statistics with confidence: confidence intervals and statistical guidelines*. BMJ Books, London
48. McGee S (2002) Simplifying likelihood ratios. *J Gen Intern Med* 17(8):647–650. <https://doi.org/10.1046/j.1525-1497.2002.10750.x>
49. Henderson MC, Tierney LM, Smetana GW (2012) *The patient history*, 2nd edn. McGraw-Hill, New York, p 30 (ISBN 978-0-07-162494-7)
50. Kunene KN, Weistroffer HR (2008) An approach for predicting and describing patient outcome using multicriteria decision analysis and decision rules. *Eur J Oper Res* 185(3):984–997
51. West D, Mangiameli P, Rampal R, West V (2005) Ensemble strategies for a medical diagnostic decision support system: a breast cancer diagnosis application. *Eur J Oper Res* 162(2):532–551