ORIGINAL PAPER

Application of Artificial Neural Networks for Predicting the Bearing Capacity of Shallow Foundations on Rock Masses

M. A. Millán[1] · **R. Galindo[2]** · **A. Alencar[2]**

Abstract

Calculation of the bearing capacity of shallow foundations on rock masses is usually addressed either using empirical equations, analytical solutions, or numerical models. While the empirical laws are limited to the particular conditions and local geology of the data and the application of analytical solutions is complex and limited by its simplified assumptions, numerical models offer a reliable solution for the task but require more computational effort. This research presents an artificial neural network (ANN) solution to predict the bearing capacity due to general shear failure more simply and straightforwardly, obtained from FLAC numerical calculations based on the Hoek and Brown criterion, reproducing more realistic configurations than those offered by empirical or analytical solutions. The inputs included in the proposed ANN are rock type, uniaxial compressive strength, geological strength index, foundation width, dilatancy, bidimensional or axisymmetric problem, the roughness of the foundation-rock contact, and consideration or not of the self-weight of the rock mass. The predictions from the ANN model are in very good agreement with the numerical results, proving that it can be successfully employed to provide a very accurate assessment of the bearing capacity in a simpler and more accessible way than the existing methods.

Keywords Artificial neural network · Bearing capacity · Shallow foundation · Hoek and Brown failure criterion · Finite difference method

List of Symbols

σ_1	Major principal stress
σ_3	Minor principal stress
σ_c, UCS	Uniaxial compressive strength
m_0	Rock type
GSI	Geotechnical quality of the rock mass
ψ	Dilatancy angle
D	Depth of damage in the rock mass due to human actions
B	Foundation width
x_i	Input from each node in the previous layer in a layer of a neural network
w_{ji}	Adjustable connection weight in a neural network
θ_j	Threshold value or bias in a neural network
y_j	Output from a layer in a neural network
f	Nonlinear transfer function in a neural network
f_{hidden}, f_{outp}	Nonlinear transfer function at the hidden and output layers in a neural network
P_h	Bearing capacity of the shallow foundation

✉ M. A. Millán
miguelangel.millan@upm.es

1   Universidad Politécnica de Madrid, ETS Arquitectura, Avda. Juan de Herrera nº4, 28040 Madrid, Spain

2   Universidad Politécnica de Madrid, ETSI Caminos, C. y P., C/ Profesor Aranguren s/n, 28040 Madrid, Spain
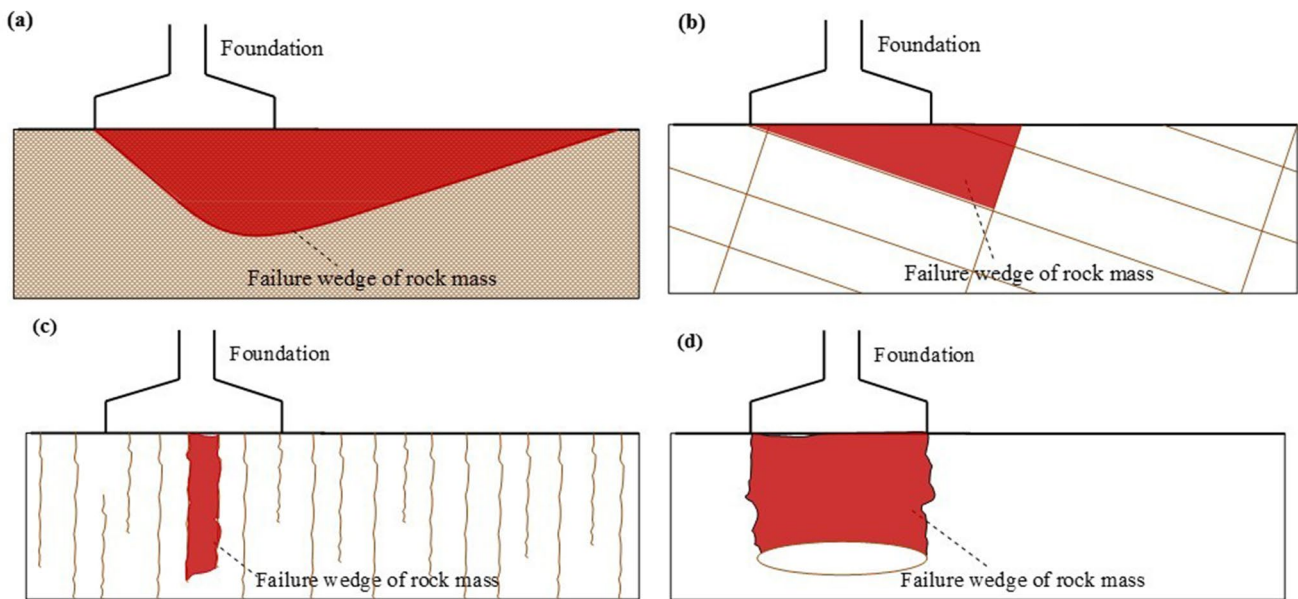
1 Introduction

Shallow foundations are the type of foundation more commonly used to transmit the applied loads to the underlying soil or rock from civil engineering or building structures. According to the properties of the rock mass and the layer beneath, the failure of rocks under applied loads may occur through several mechanisms (Sowers 1979; Paikowsky et al. 2010; Canadian Geotechnical Society 2006). These mechanisms are presented in Fig. 1 and correspond to (a) general shear failure; (b) local failure when discontinuity spacing is bigger than the foundation width; (c) failure of the underlying rock columns created by discontinuities spacing smaller than the foundation width; (d) punching shear failure following vertical joints existing bellow the contact zone or produced by shallow cavities.

Fig. 1 Failure mechanisms for shallow foundation: **a** general shear failure; **b** local failure through large spacing discontinuities; **c** columnar failure; **d** punching shear failure

These failure modes reveal the range of application of the different analytical or numerical formulations and the validity of empirical methods based on field trials. The local conditions are more representative of cases (b), (c), and (d), which acquire a great dependence on the spacing and the width of the foundation so that their adequate bearing capacity can be deduced from simplified mathematical models, field trials, and reduced models, where such local conditions can more easily be presented. However, the foundations of large civil engineering works, which include rock masses with varied distribution of discontinuities (type I, IV or V, Fig. 2), respond in a general way to a type of global failure, defined by the case (a). In this case, empirical experiences are more unpredictable due to their local condition and it is necessary to resort to analytical and/or numerical techniques.

The limitations under which analytical solutions can be established and the high computational costs to guarantee the convergence and stability of the solution in numerical methods (both are analyzed in detail in the following sections) makes advisable to offer simple formulations that allow obtaining a reliable solution of rapid application for the calculation of the bearing capacity under global failure in a rock mass.

Considering the intrinsic difficulty of the assessment of many of the engineering problems related to technical design, soft computing appeared as an alternative to the common analytic and numeric approaches. A type of soft computing technique is ANN. It learns from a dense enough data set and configures a black-box-type prediction

model that solves the problem in the form of a closed simple equation.

Many applications of ANN were presented in recent years, particularly in the Geotechnical Engineering field, as discussed by Shahin et al. (2001, 2008) in the review papers on the subject, expanding to almost every problem in the field. Regarding shallow foundations, the specific subject of this research, many applications of neural networks were presented.

The problem of shallow foundation settlement on cohesionless soils was addressed first, considering for example Sivakugan et al. (1998), Chen et al. (2006) and Shahin et al. (2002a, 2002b, 2003a, 2003b, 2004a, 2005a, 2005b). Later, some other researchers used perceptrons to obtain alternative approaches to the problem of shallow foundations, predicting the bearing capacity of strip footing on multi-layered cohesive soil (Kuo et al. 2009), or on cohesionless soils using neuro-fuzzy models (Provenzano et al. 2004; Padmini et al. 2008).

In the rock mechanics field, some advances were presented recently. Some of them are related to the analysis by ANN's of the parameters defining the rock behavior. Some researchers as Yang and Zhang (1997) used neural networks to identify the relative effect of each factor involved in a rock mechanic problem, as the stability of underground openings. Mert et al. (2011) and Gholami et al. (2013) presented an approach to assess the total RMR classification system using a simulation-based on neural networks. Ocak and Seker (2012) developed a neural network to estimate the elastic modulus of intact rocks, since its difficult determination in

Fig. 2 Rock mass types with global failures for shallow foundation (adapted from Hoek 1983)
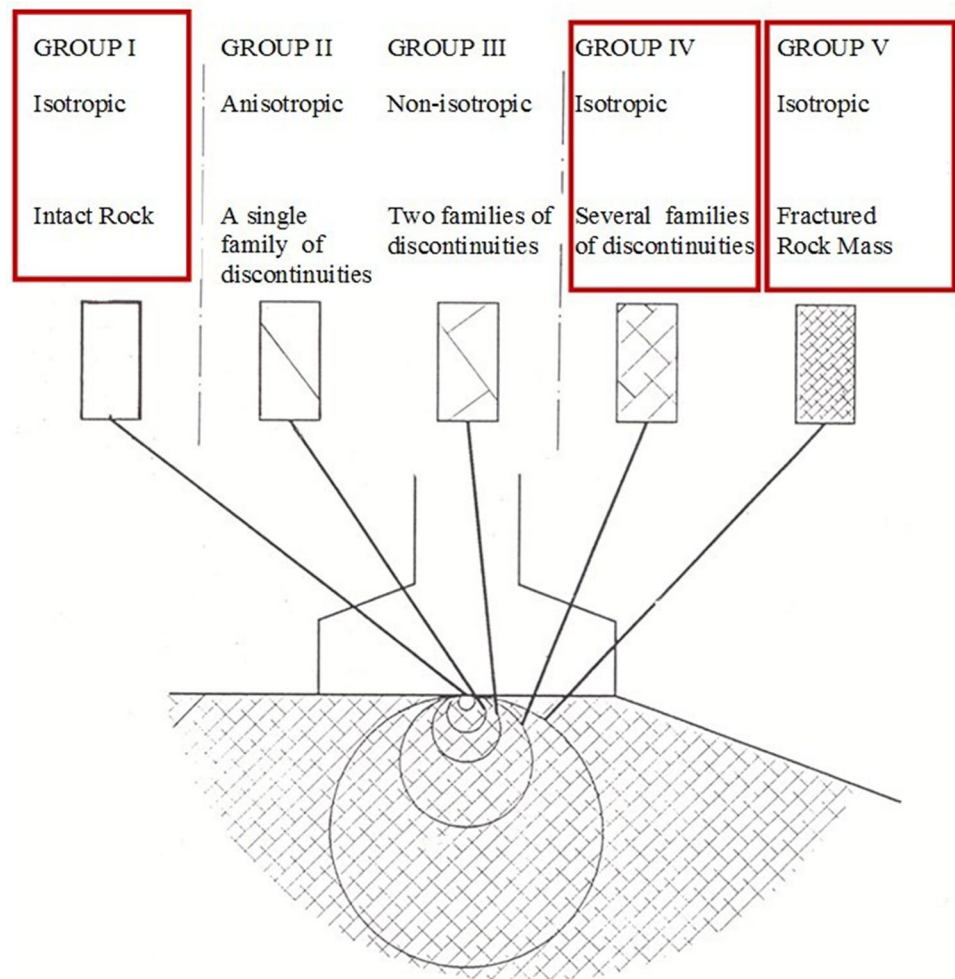
laboratory tests because high-quality cores are required. Yılmaz and Yuksek (2008) used ANN to indirectly estimate the rock parameters.

Very few results were presented related to the bearing capacity of shallow foundation on rocks, as Ziaee et al. (2015), based in a comprehensive database of tests and reduced model results, using four main parameters: rock mass rating, unconfined compressive strength of rock, ratio of joint spacing to foundation width, and angle of internal friction for the rock mass. Alavi and Sadrossadat (2016) proposed precise predictive equations derived from linear genetic programming (LGP) for the ultimate bearing capacity of shallow foundation resting on a jointed (non-fractured) rock. A comprehensive and reliable set of data including 102 previously published rock sockets, centrifuge rock sockets, plate load, and large-scaled footing load test results is collected to develop the models. These works, based on the results of load tests and small-scale tests, allow considering the incidence of local conditions to develop local failure mechanisms. However, they do not focus attention on the general shear failures generated in

sound or fractured rock masses, which can occur in large-scale civil works, where furthermore, for an adequate analysis of the bearing capacity, factors that are difficult to reproduce in tests on a small scale such as dilatation or roughness of the foundation contact must be incorporated. Then, more research is needed and this paper follows this line of exploration, introducing an enlarged set of parameters and adopting the widely accepted Hoek and Brown failure criterion for the rock.

As proved by the previous references, using ANN appears as an adequate alternate approach to the problem, overcoming the limitations of empirical, analytical, and numerical methods, provided an extended and accurate set of data is used to build the network.

A neural network (ANN) is proposed in this research to offer this alternative approach, introducing an enlarged set of parameters and adopting the widely accepted Hoek and Brown failure criterion for the rock. The network learns from a dense enough data set and configures a black-box-type prediction model that solves the problem in the form of a set of closed simple equations.

2 Objectives

The main objective of this research is to develop a neural network to address the calculation of the bearing capacity of shallow foundations on rock masses from a different approach than in common practice, where empirical, analytical, and numerical methods are used.

The reason for using this alternative approach is to make available a method of calculation that is simultaneously easy to use without having the limitations of other simple methods as empirical or analytical methods. That is also to say, having the accuracy of numerical methods without their complexity of use.

To reach this goal, it is necessary to consider a detailed analysis of the rock failure mechanism and of the analytical and numerical methods used to solve the problem of the bearing capacity, considering their respective limitations.

The simplifications adopted in the analytical solution (Serrano et al. 2000) (plane strain, the associative flow rule, the coaxiality, the perfectly plastic yield surface, and weightless rock mass) and the relative complexity of numerical models, that considerably raises computational costs, invite to consider new approaches. These approaches should relate the bearing capacity of shallow foundations on rock masses with an extended set of parameters, including those related to the rock itself and those related to the problem geometry and characteristics.

The neural network is built based on an extended set of numerical calculations obtained using the commercial software FLAC, including those most influential parameters in the bearing capacity problem.

From this dataset, the network is trained and optimized, offering a fair rate between simplicity and accuracy. Finally, this generated ANN is converted into a set of simple equations of easy implementation and use.

3 Methods of Analysis

3.1 Identification of the Failure Mechanism Under Study

The heterogeneous and anisotropic nature that can characterize rock masses locally is determined by the presence of discontinuity families so that the study of the bearing capacity is highly conditioned by the distribution of discontinuities in the rock mass.

Simplified and semi-empirical solutions are available for specific rock mass configurations (Bishoni 1968; Goodman 1989) indicating that the bearing capacity of shallow foundations on jointed rock masses depends on the ratio of space between joints to foundation width, joint condition, rock type, and the condition of the underlying rock mass. Among this type of formulations, some stand out in regulations such as the Eurocode 7 (2004), which includes a rather simplified method, in which, as a function of the rock type, the ultimate bearing capacity is estimated depending on the uniaxial compressive strength and the spacing of the main joint set. However, these formulations do not take into account the important role of the rock type and its qualitative mass parameters such as rock quality designation (RQD), rock mass rating (RMR), or geological strength index (GSI) and therefore, they do not result useful for identifying general failure mechanisms, which cover a large amount of ground.

There were also proposed empirical methods that often establish a correlation between the bearing capacity and rock mass properties based on the observations and experimental test results such as (Bowles 1996) or the method by Carter and Kulhawy (1988), recommended by AASHTO (2007) and based on the lower bound solution. These equations consider the particular and local conditions of the experimentation but cannot take into account the geometry of the foundations or the spacing between joints in a general way.

The singularities induced by the distribution of discontinuities can be treated globally when the scale of the problem allows us to consider any of the situations indicated in Fig. 2, where the rock mass can be considered as a homogeneous and isotropic medium (cases I, IV, and V). This is the case in many of the main engineering works where large foundation surfaces are required, such as bridges, dams, or large building towers. This analysis corresponds to the study of the global shear failure of the ground and its analysis must be based on the choice of a suitable failure criterion for the rock mass.

Defining the value of the dilatancy angle is a very complex problem, being usual practice to define either null dilatancy or the associative flow rule. Considering the correlations proposed by Hoek and Brown (1997), the dilatancy is between 0º and a quarter of the friction angle, depending on the geotechnical quality of the rock mass. From this perspective, it is considered reasonable to study the extreme cases of the problem, from the conservative side (null dilatancy) to the more optimistic side (associated dilatancy).

3.2 Hoek and Brown Failure Criterion

In rock mechanic, the non-linear Hoek and Brown failure criterion (Hoek and Brown 1980, 1997; Hoek et al. 2002) is the most used and it is applicable for the rock mass with a homogeneous and isotropic behavior, that is to say, it has the same physical properties in all directions because of the inexistence or the abundance of discontinuities.

The Hoek and Brown failure criterion depends on the major principal stress (σ_1) and minor principal stress (σ_3) according to the following equation:

$$\frac{\sigma_1 - \sigma_3}{\sigma_c} = \left(m \cdot \frac{\sigma_3}{\sigma_c} + s \right)^a . \tag{1}$$

The uniaxial compressive strength (UCS) is σ_c, while the parameters m, s, and a can be evaluated by (2), (3) and (4) and depend on the rock type (m_0), geotechnical quality of the rock mass (GSI) and damage in the rock mass due to human actions (D) that in shallow foundations is usually equal to zero:

$$m = m_0 \cdot e^{\frac{GSI-100}{28-14 \cdot D}}, \tag{2}$$

$$s = e^{\frac{GSI-100}{9-3 \cdot D}}, \tag{3}$$

$$a = \frac{1}{2} + \frac{1}{6} \cdot \left(e^{-GSI/15} - e^{-20/3} \right). \tag{4}$$

This criterion defines the failure envelope for a rock mass and allows the study of the mechanism of general shear failure in the ground for the problem of bearing capacity of shallow foundations.

3.3 Calculation Methods of the Global Shear Failure Mechanism

Over the years several methods were used to study the bearing capacity under the global failure of the ground: limit equilibrium method (Terzaghi 1943; Meyerhof 1951), slip line method (Sokolovskii 1965), limit analysis method (Sloan 1988; Sloan and Kleeman 1995), and the numerical method (Griffiths 1982; Merifield et al. 2006).

The traditional analytical solutions to estimate the ultimate bearing capacity in soils (Terzaghi 1943; Brinch-Hansen 1970) were developed for the linear Mohr–Coulomb failure criterion that depends on the cohesion and internal friction angle of the material. Since the development of the non-linear Hoek and Brown failure criterion for rock masses (Hoek and Brown 1980, 1997; Hoek et al. 2002), the equivalent strength parameters of the Mohr–Coulomb failure criterion (cohesion and friction angle) were assessed for the corresponding stress level of the rock mass, to introduce them in traditional formulations. However, some research from different authors as Merifield et al. (2006) observed that the use of Mohr–Coulomb equivalent strength parameters overestimates the bearing capacity by up to 157% in the case of a good quality rock mass (GSI about 75).

To calculate the bearing capacity, Yang and Yin (2005) applied the multi-wedge translation failure mechanism and the tangential line technique. They used the upper bound limit theory for strip foundation based on a modified Hoek and Brown failure criterion deducing the equivalent parameters of the Mohr–Coulomb failure criterion. Saada et al. (2008) also proposed another method to calculate the bearing capacity based on the limit theories by applying the Hoek and Brown failure criterion and deducing the equivalent parameters of the Mohr–Coulomb method that provided a better fit than the results obtained in Yang and Yin (2005). However, these methods do not claim to be general for consideration in the presentation of results and their implementation requires a specific analysis.

Thus, a specific study is necessary that expressly considers the markedly non-linear character of the rock mass, for which analytical and numerical methods have been developed.

3.3.1 Analytical Method

The analytical method that solves the internal equilibrium equations combined with the failure criterion was proposed by Serrano and Olalla (1994) and Serrano et al. (2000) applying the Hoek and Brown failure criterion (Hoek and Brown 1997) and the modified Hoek and Brown failure criterion (Hoek et al. 2002) respectively. It is based on the characteristic line method (Sokolovskii 1965) with the hypothesis of plane strain, associative flow rule, coaxiality, perfectly plastic yield surface, and weightless rock mass.

According to this analytical formulation, the ground surface that supports the foundation is composed of two sectors (Fig. 3): boundary 1 with inclination α where the load acting on a surface is known (for example, the ground load on the foundation level or the load from installed anchors) (acting with the inclination i_1); and the boundary 2, where the bearing capacity of the foundation should be determined (acting with the inclination i_2).

The failure criterion of Hoek and Brown (1) is nonlinear, then the inclination of the curve on the diagram shear stress
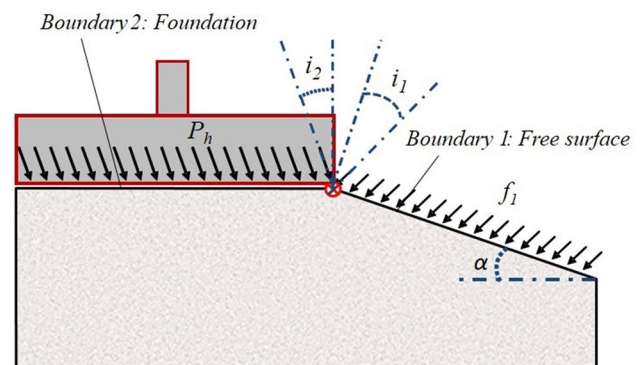

Fig. 3 Mathematical model of the bearing capacity of the strip footing

(τ)—normal stress (σ) (which in rock mechanics is associated with the instantaneous friction angle of the material) depends on the stress state.

The analytical solution is based on the characteristic line method with the equation of the Riemann invariants fulfilled along the characteristic line (Serrano et al. 2000). From the overload on the surface, it can be obtained the instantaneous friction angle, and the direction of the principal stress in this boundary 1 (see Fig. 3), and using Riemann invariants, both the instantaneous friction angle below the foundation (boundary 2, see Fig. 3) and the ultimate bearing capacity can be estimated.

3.3.2 Numerical Methods

Finite element limit analysis has been one of the possibilities to obtain collapse loads. Different calculations using the finite element method under lower and upper bound theorem (Sloan 1988; Sloan and Kleeman 1995) were used by Zheng et al. (2000) and Sutcliffe et al. (2004) to determine the bearing capacity of the fractured rock and jointed rock mass. Later, Merifield et al. (2006) applied the limit theorems (upper and lower bound), as an extension of the formulation developed by Lyamin and Sloan (2002a, b) to determine the bearing capacity on a fractured rock mass whose behavior is a Hoek and Brown type. More recently, a lower-bound finite elements limit analysis in combination with either semidefinite programming (SDP) or nonlinear optimization respectively (Kumar and Khatri 2011; Kumar and Mohapatra 2017; Chakraborty and Kumar 2015) have been applied to solve stability problems involving a modified Hoek and Brown (Hoek et al. 2002) yield criterion in the rock mass, in particular the bearing capacity for both circular and strip footings.

Finite element limit analysis computes the upper o lower limit load using optimization techniques rather than time stepping and increasing the system load to a collapse load, as conventional non-linear finite element techniques do. They may use linear programming to consider the common non-linear behavior of yield surfaces or second-order cone programming to address them directly (Makrodimopoulos and Martin 2006). However, as reported in Smith and Gilbert (2007), the solutions obtained are often highly sensitive to the geometry of the original finite element mesh, though the problem may be overcome by adaptive remeshing techniques at the cost of a higher complex procedure.

Similar limitations were found using other numerical techniques as the Finite Difference method. For example, some authors reported the difficulty of obtaining stable results in conditions of highly fractured rock masses with high values of m_0 and low values of GSI (Merifield et al. 2006). Contrary to the case of the linear Mohr–Coulomb criterion used in soil mechanics, the application of the Hoek

Table 1 Parameters considered for evaluating the bearing capacity of a shallow foundation on a rock mass

Geomechanical	Rock type (m_0)
	Uniaxial compressive strength (UCS)
	Rock mass quality index (GSI)
	Dilatancy (ψ)
	Self-weight
Geometric	Shape: plane strain/axisymmetric
	Foundation width (B)
Soil-structure interaction	Contact roughness

Table 2 List of inputs for the neural network and adopted values

Inputs parameters	Adopted values
m_0	5/12/20/32
UCS (MPa)	5/10/50/100
GSI	10/50/85
B (m)	4, 5/11/16, 5/22
Dilatancy	Null/associated
Shape	Plane strain/axisymmetric
Contact roughness	Roughless/rough
Self-weight	Weigthless/self-weigthted

and Brown non-linear criterion has produced results discrepancies among researchers when evaluating the bearing capacity of a foundation (Alencar et al. 2019) due to the need for an exhaustive numerical study.

4 Generation of the Bearing Capacity Results Database

To apply the ANN technique, it is necessary to have a database with reliable results, which is extensive enough to cover the range of variation of the parameters identified as essential for defining the global bearing capacity of the rock mass and that allows training and properly validating the developed neural network.

The first step is to identify the parameters that influence the evaluation of the bearing capacity of a foundation on a rock mass. Thus, the geomechanical, geometric, and soil-structure interaction parameters indicated in Table 1 are considered.

Next, the calculation cases have been established by varying the parameters throughout their range of variation to have the set of data used to build and calibrate the ANN model. Table 2 shows the values adopted for the parameters so that a series of 2762 calculation simulations is obtained.

Finally, it is necessary to be able to generate reliable results for each simulation case of the bearing capacity of shallow foundations on rock masses. For this, the

geotechnical commercial programs FLAC (Itasca Consulting Group 2012) has been used. FLAC applies the finite differences method (FDM) to solve geotechnical problems. The software contains different constitutive models like Mohr–Coulomb and Hoek and Brown, using the tangent linear approximation in this last failure criterion. In FDM it is assumed that the bearing capacity is reached when the continuous medium does not support more load because an internal failure mechanism was formed. However, an accurate calculation requires an extensive and laborious convergence analysis of the different numerical parameters, which becomes costlier in computational terms due to the markedly non-linear character of the Hoek and Brown failure criterion that governs rock masses. To guarantee the precision of the solution, the following have been numerically controlled for each of the calculation simulations: (1) the linearization of the Hoek and Brown criterion of the rock mass; (2) the nodal distance; and (3) the velocity condition applied to the boundary nodes.

In the case of the linearization of the failure criterion, it should be noted that an adaptive discretization has been adopted for each area of the numerical model so that, in each calculation step, the stress in the area is defined as the point of failure on the Hoek and Brown criterion (that is, the failure line that is tangent to that point on the failure surface). This way of associating the linearization of the failure criterion is extended to all the areas of the numerical model and is carried out until the stresses of all the areas stabilize.

In the case under study, the vertical load was obtained through a constant velocity condition (low enough to reproduce quasi-static load conditions) applied to the boundary nodes, and the bearing capacity was determined from the relation between stresses and displacements of one of the nodes; in this case, the central node of the foundation was considered.

The numerical control of the velocity of the nodes and the nodal distance implies carrying out several similar calculations with different values of each numerical parameter until the solution is stabilized. Thus, Fig. 4 shows a 2D finite-difference model used to calculate the different cases, applying a plane strain condition to a symmetric model, where only half of the strip footing is represented; while in Fig. 5 the convergence control is presented with both numerical parameters in one of the numerical simulations carried out with FLAC, which corresponds to the values: $m_0 = 12$; UCS $= 50$ MPa; GSI $= 50$; $B = 4.5$ m; associated dilatancy; plane strain; rough contact and weightless rock.

5 Overview of Back-Propagation Neural Networks

A comprehensive description and analysis of the type and use of back-propagation perceptrons is above the scope of this paper and can be found somewhere else (for example Zurada 1992; Fausett 1994).

A multilayer perceptron (MLP) consists of multiple layers of computational units (nodes or neurons) interconnected, in this case, in a feed-forward way. The first layer is formed by the input neurons, connected to one or more hidden layers of neurons, and followed by the final layer of output neurons.
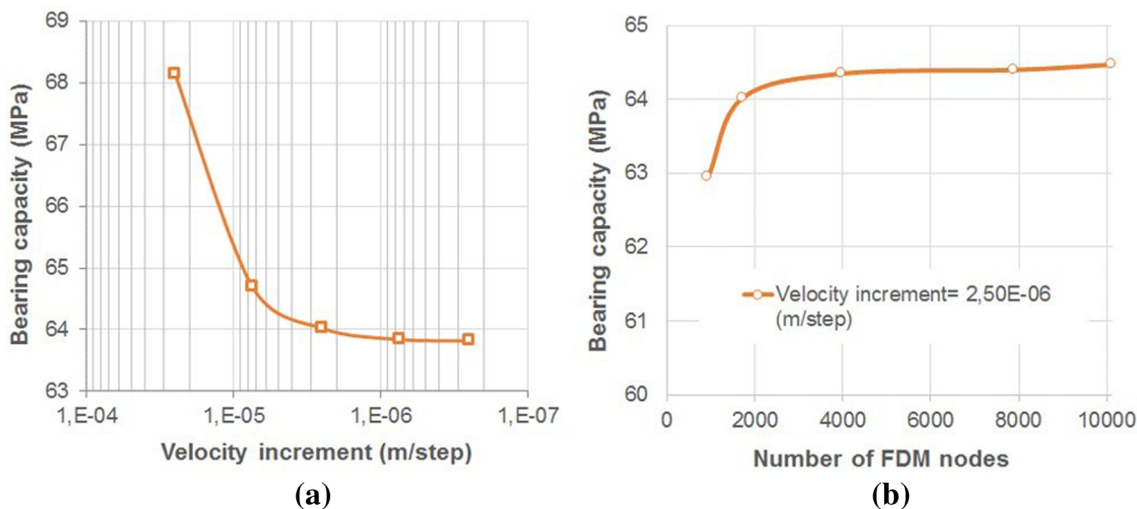
Fig. 4 2D model used in the calculation through FDM
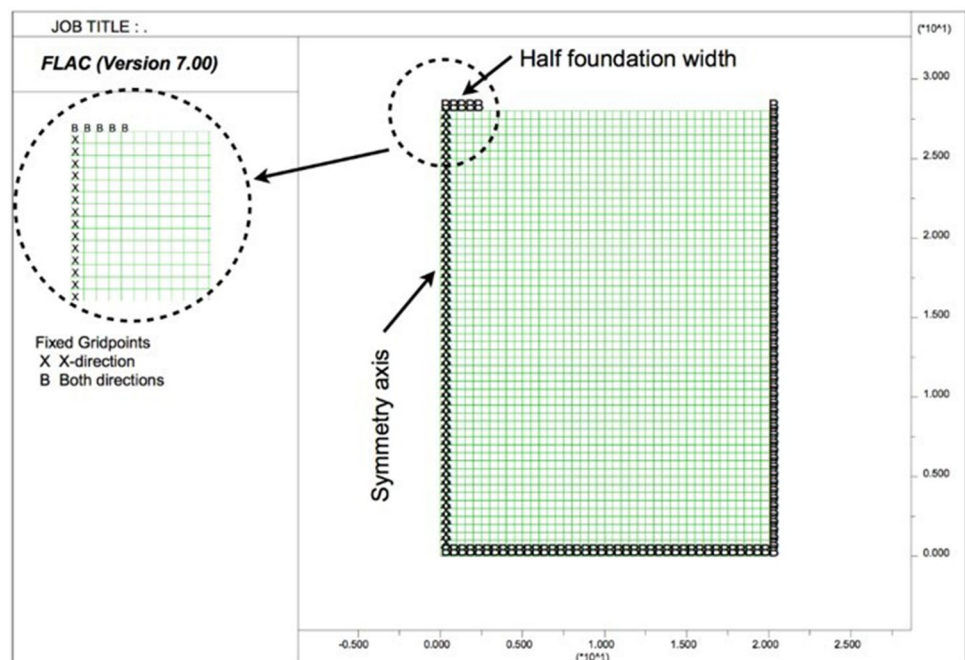
Fig. 5 Numerical parameters considered for evaluating the bearing capacity of a shallow foundation on a rock mass for the case $m_0 = 12$; UCS = 50 MPa; GSI = 50; $B = 4.5$ m; associated dilatancy; plane strain; rough contact and weightless: **a** FDM convergence analysis according to the velocity increment; **b** FDM convergence analysis according to the density of nodes

Each neuron in one layer has directed weighted connections to the neurons of the subsequent layer.

The input from each node in the previous layer (x_i) is multiplied by an adjustable connection weight (w_{ji}) when arriving at the node j in the actual layer, and a threshold value or bias (θ_j) is added or subtracted. This combined input (I_j) is then passed through a non-linear transfer function (sigmoidal or tanh function), and the result becomes the input for the next layer. This process is summarized in Figs. 6 and 7, and Eqs. 5 and 6:

$$I_j = \sum \left(w_{ji} \cdot x_j \right) + \theta_j, \tag{5}$$

$$y_j = f(I_j). \tag{6}$$

This process starts from the input layer, fed by a set of data presented to the network for the learning process. The final output obtained from this data is compared against the known results, then a measure of some predefined
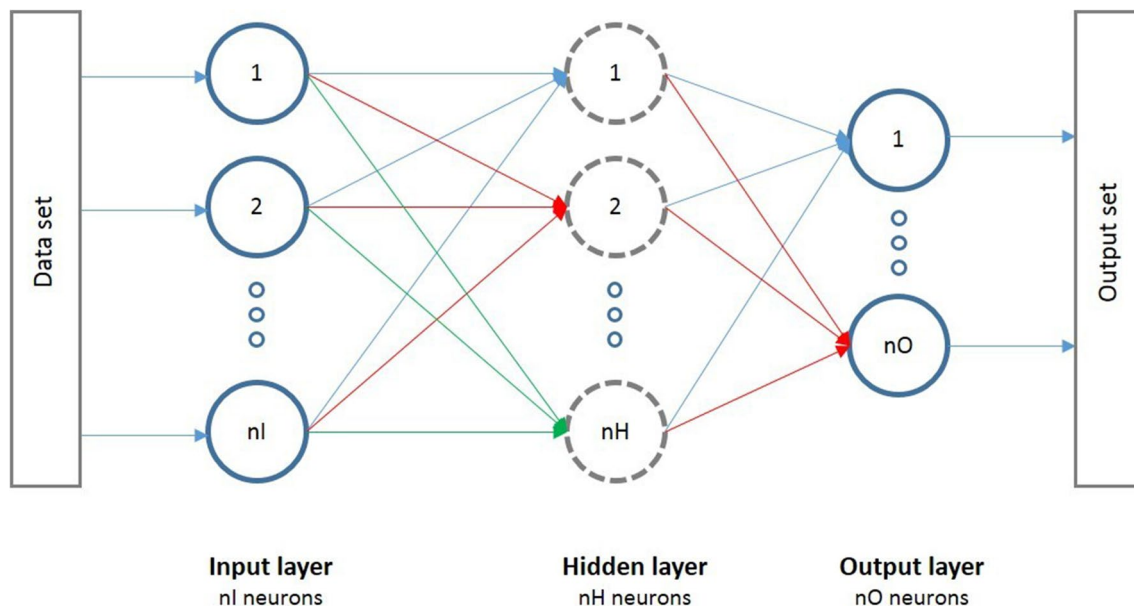

Fig. 6 A schematic illustration of a multilayer perceptron network
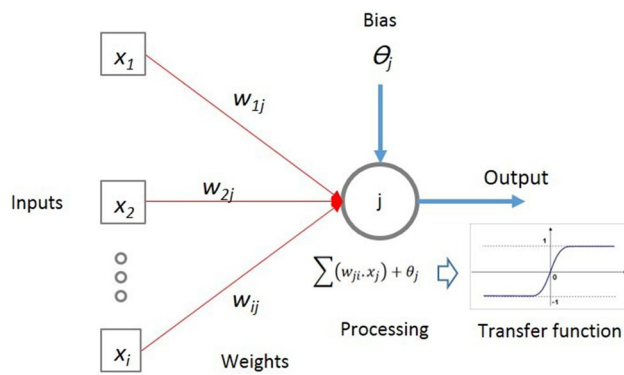
Fig. 7 Input–processing–output system in an artificial neuron

error-function is obtained. Using different optimization techniques, the error is fed back through the network, and then the algorithm adjusts the weights of each connection to reduce the value of the error function by some small amount. Repeating this process, a large enough number of training cycles, the network eventually converges to a state having a small error, meaning that the network has learned the function.

An excessive number of training cycles may produce what is called overfitting, giving smaller errors for the training set but reducing the net capacity to predict and generalize the results from other sets. A stopping criterion is needed to avoid overfitting, being the cross-validation technique the most accepted in the field, and will be introduced in the next section.

6 ANN Model Development

In this research, the artificial neural network for predicting the bearing capacity of shallow foundations on rock masses is developed using the Matlab computer software utilities (Mathworks Inc. 2019a). The data used to build and calibrate the model is obtained from a series of 2762 numerical simulations of the problem using the FLAC software based on finite differences, changing the different parameters affecting the model behavior.

6.1 Selection of the Inputs and Outputs

In this particular case, as in most cases in Geotechnical Engineering, the parameter selection for the neural network is based on the knowledge of the physical problem underlying the behavior of the system.

The main parameters defining the bearing capacity of a shallow foundation on a rock mass are: three parameters characterizing the rock mass as rock type (m_0), uniaxial compressive strength (UCS), and geological strength index

(GSI), and the foundation width (B). However, as introduced before in Table 2, there are also some other additional parameters influencing the results that allow expanding the limitations of analytical and empirical formulations and are also considered as inputs: the dilatancy on the failure surface (assigning a 0 value when the dilatancy angle is $\psi = 0$ and 1 for $\psi = \rho$), bidimensional or axisymmetric problem (assigning a 0 and 1 values respectively), rough or roughless foundation-rock contact (assigning a 0 and 1 values respectively), and weightless or self-weighted rock mass (assigning a 0 and 1 values respectively).

Therefore, the total number of inputs is 8 and the bearing capacity of the foundation is the single output variable.

6.2 Data Division and Statistical Analysis

When creating a neural network, one of the basic issues is training the net up to a point before losing its capacity for accurate predictions with a different set of inputs. This situation is known as over-fitting and can be avoided if training and checking of prediction capabilities are developed simultaneously using different subsets of the data input.

The total number of available inputs is divided into three sets, training, testing, and validation. Using this division, cross-validation or the net is allowed, as proposed by Stone (1974). The training set is used to adjust the weight of the net nodes, and the testing set is used to check the performance of the model at different stages of the training process and determine when to stop to avoid over-fitting. The validation set allows an independent check of the network when the training process is finished and the different weights of each connection are already defined. The percentage of the inputs assigned to each set is (75-15-15), meaning that 75% is assigned to the training set (1658 cases), 15% to the testing set (332 cases), and 15% to the validation set (332 cases).

The above-mentioned division should represent statistically the whole population of inputs (following Masters 1993). Then, the statistical properties of the different subsets (e.g. mean, standard deviation, and range) should match those of the complete set. Consequently, all patterns contained in the original set of inputs should be reproduced in each of the subsets. Since the division is performed randomly, an iterative process is adopted until a group of training, test, and validation sets consistent with the statistical properties is achieved.

The statistical properties considered are mean, standard deviation, minimum, maximum, and range, as suggested by Shahin et al. (2004b). Table 3 shows the results obtained for the different input parameters. It is clear that, as the input population is large, there is no problem in the statistical matching among the different subsets.

The validity of the network is limited by the range of parameters included in the training data. Consequently, the

Table 3 Input and output statistics for the different data sets (75-15-15 case)

Model variable	Dataset	Mean	Standard deviation	Minimum	Maximum	Range
m_0	Training	16.88	9.91	5	32	27
	Testing	16.96	9.85	5	32	27
	Validation	17.73	10.26	5	32	27
B (m)	Training	13.50	6.44	4.5	22	17.5
	Testing	13.39	6.53	4.5	22	17.5
	Validation	13.72	6.72	4.5	22	17.5
UCS (MPa)	Training	42.15	38.46	5	100	95
	Testing	38.03	37.31	5	100	95
	Validation	39.31	37.77	5	100	95
GSI	Training	47.68	30.27	10	85	75
	Testing	49.13	30.43	10	85	75
	Validation	52.59	29.34	10	85	75
Dilatancy	Training	0.66	0.47	0	1	1
	Testing	0.69	0.46	0	1	1
	Validation	0.70	0.46	0	1	1
Plane strain/axisymmetric	Training	0.49	0.50	0	1	1
	Testing	0.53	0.50	0	1	1
	Validation	0.49	0.50	0	1	1
Contact roughness	Training	0.69	0.46	0	1	1
	Testing	0.67	0.47	0	1	1
	Validation	0.66	0.47	0	1	1
Self-weight	Training	0.52	0.50	0	1	1
	Testing	0.52	0.50	0	1	1
	Validation	0.50	0.50	0	1	1
P_h	Training	141.36	265.56	0.18	1745	1744.82
	Testing	127.25	240.16	0.21	1700	1699.79
	Validation	139.96	234.71	0.2075	1718	1717.7925

performance is better when no extrapolation is made over that range.

6.3 Data Scaling

In most ANN's, the input parameters are scaled to a (0, 1) or (−1, 1) range before the training to eliminate their dimension and to assure all of them are treated homogeneously during training the network, working well when the data distribution is more or less homogeneous. However, that scaling procedure, used alone, does not produce good results in this research.

One of the more important problems found in building the network depends on the distribution of the output values of bearing capacity (P_h) among the data range, which is extremely concentrated around the small values of P_h (Fig. 8a). Since most of the P_h values are very small compared to the whole range, and even the global error of the ANN approximation was very reduced, it created very important percentage errors for the smaller values when using only regular scaling.

Adding a logarithmic scaling (natural log) to the P_h targets before the regular scaling (0, 1) solves the problem since it produces a more homogeneous distribution of the data along with the P_h range. Data distributions without scaling, using logarithmic scaling and the final step, with log scaling and (0, 1) scaling simultaneously applied, are shown in Fig. 8b, c, respectively. It is important to note that the log scaling is exclusively applied to the output values P_h, and not to any other input value.

6.4 Determination of Network Architecture and Internal Parameters

As introduced in Sect. 5, the neural network is a multilayer set of neurons (input layer-hidden layers-output layer), interconnected by inputs from the previous layer and output to the next layer. At each neuron, the inputs are changed into outputs by affecting them with weights and biases and applying a transfer function that changes their values. Building the net means defining the model architecture (layers and neurons) and training it (weights and biases).
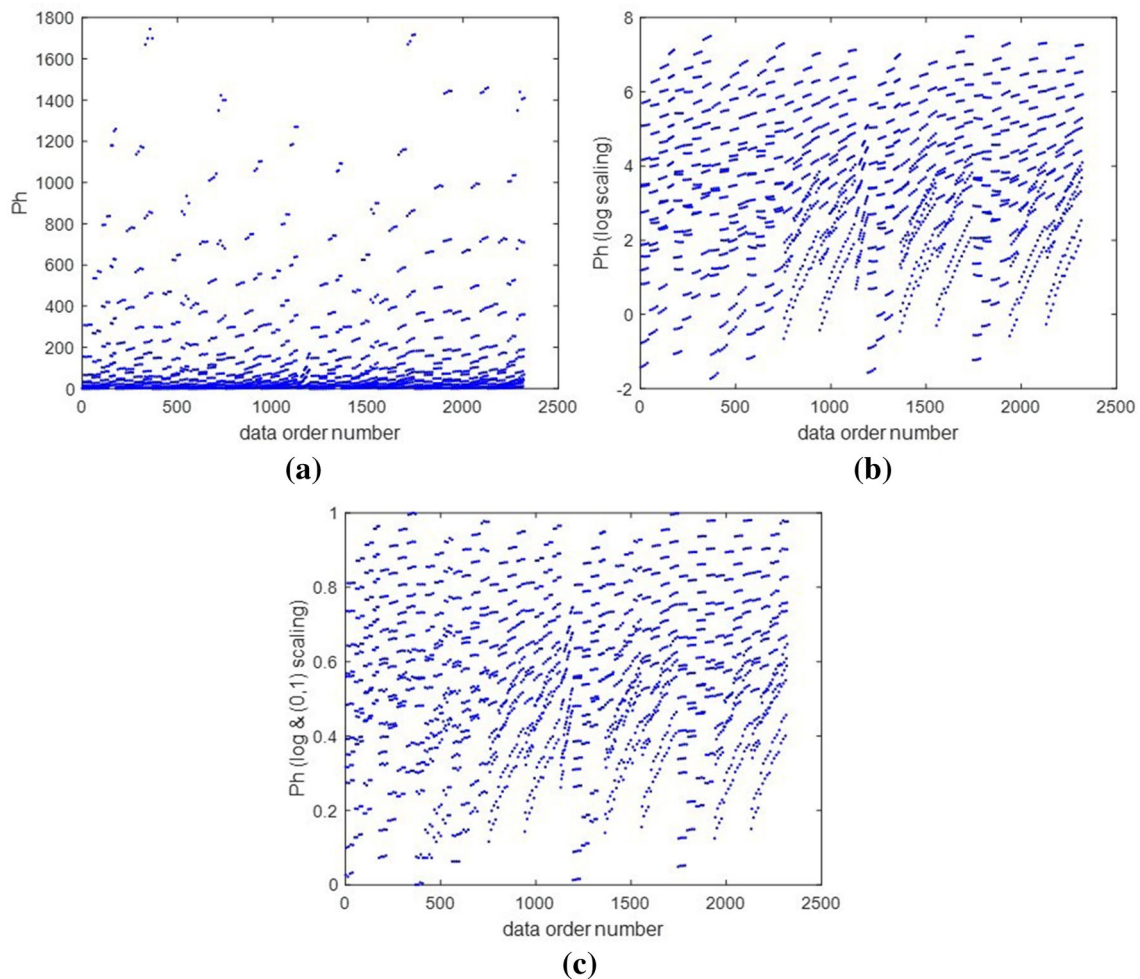
Fig. 8 Different stages of normalization of the target data P_h: **a** Original data, **b** log scaling, and **c** (0,1) scaling applied to the previous log-scaled data

Although the input layer and output layer are predefined by the problem under study and the intended results, the hidden layers are not. Model architecture requires the selection of the optimal number of hidden layers in the network and the determination of the optimal number of neurons in each layer. Defining a net includes a process of defining many networks of different complexity and select the optimum.

There is no unified theory to obtain the optimal number of layers in a network and that task is performed by trial-and-error procedures. Besides, a single hidden layer can reproduce successfully any continuous function, as stated by Hornik et al. (1989). The number of hidden nodes $(2I + 1)$ (being I the number of inputs) is used with success in the literature of ANN for geotechnical problems.

The single-hidden layer network is tried using several nodes starting from 3 to a maximum of 25, checking that over $(2I + 1) = 17$ (being $I = 8$ the number of inputs), no improvement is achieved, as suggested by Caudill (1988). The transfer functions used at each layer to change from

input to output (see Sect. 5) can be also chosen among a variety of them. In the single hidden layer model, different combinations of transfer functions were tried (sigmoidal, hyperbolic tangent, and linear for the output layer) to identify those giving the most accurate predictions. Finally, a sigmoidal transfer function is adopted for the hidden and a linear one for the output layers.

The networks performance, concerning the number of hidden layer nodes, is evaluated using as performance measures the correlation coefficient r, the root mean square percentage error, RMSPE, and mean absolute percentage error, MAPE. These indicators are chosen because their measurements of the errors are relative and expressed as a percentage, allowing their immediate and straightforward evaluation. The expression for each one is included in Table 4. The results for the different indexes are shown in Table 5.

The RMSPE and MAPE are also plotted versus the number of hidden layer nodes in Fig. 9 and Fig. 10, respectively. As can be seen in the figure, performance quickly

Table 4 Performance measures of error

Correlation coefficient	Root mean square percentage error	Mean absolute percentage error
$r = \dfrac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\cdot\text{Var}(Y)}}$	$\text{RMSPE} = \dfrac{100}{n}\sqrt{\dfrac{\sum_{i=1}^{n}(x_i-y_i)^2}{n}}$	$\text{MAPE} = \dfrac{100}{n}\sum_{i=1}^{n}\left|\dfrac{(x_i-y_i)}{x_i}\right|$

X, Y are the compared series of data, x_i and y_i are the elements of the series. Cov(X,Y) is the covariance between X and Y, and Var(X) the variance of X

Table 5 Performance results of ANN models for different number of hidden layer nodes

| Number hidden layer nodes	Performance measures								
	Training set			Testing set			Validation set		
	r	RMSPE (%)	MAPE (%)	r	RMSPE (%)	MAPE (%)	r	RMSPE (%)	MAPE (%)
1	0.961	51.1	38.9	0.962	49.4	37.3	0.966	50.1	37.4
3	0.989	17.1	13.5	0.987	18.1	13.7	0.991	18.2	14.5
5	0.998	7.5	5.7	0.999	7.2	5.4	0.999	7.8	5.7
7	0.999	4.9	3.6	0.999	5.3	3.9	0.999	4.9	3.4
9	0.999	3.5	2.7	0.999	3.9	2.9	0.999	3.9	2.9
11	1.000	3.3	2.5	1.000	3.8	2.7	1.000	3.2	2.4
13	1.000	2.3	1.7	1.000	2.5	1.9	1.000	2.7	2.0
15	1.000	2.1	1.5	1.000	2.6	1.8	1.000	2.4	1.8
17	1.000	2.2	1.6	1.000	2.3	1.6	1.000	2.5	1.8
19	1.000	1.9	1.4	1.000	2.5	1.7	1.000	2.5	1.8
21	1.000	1.8	1.3	1.000	2.5	1.7	1.000	2.1	1.6
23	1.000	1.7	1.2	1.000	2.2	1.4	1.000	2.0	1.4
25	1.000	1.7	1.3	1.000	2.1	1.5	1.000	2.1	1.5

Fig. 9 Root mean square percentage error versus the number of hidden layer nodes
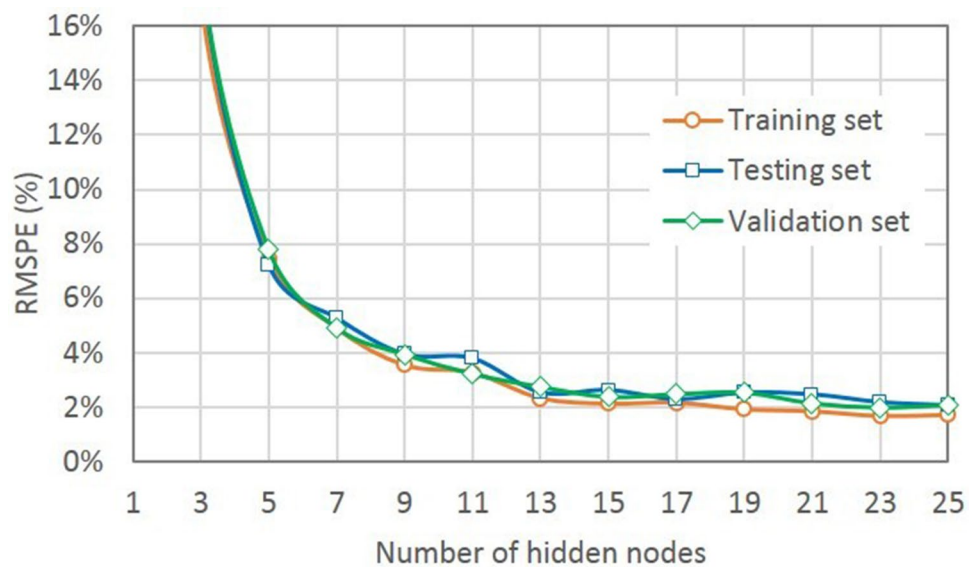
improves when increasing the number of nodes, highly improving before 5 nodes, and stabilizing over 13 nodes. A network with seven hidden layers' nodes is selected as the optimum model, balancing that the performance measures are in all cases around or below 5% and that the number of nodes is not very high. It is the authors' opinion that reducing the performance to 2% does not compensate for increasing the number of hidden nodes to 13 or 14.

**Fig. 10** Mean absolute percentage error versus the number of hidden layer nodes
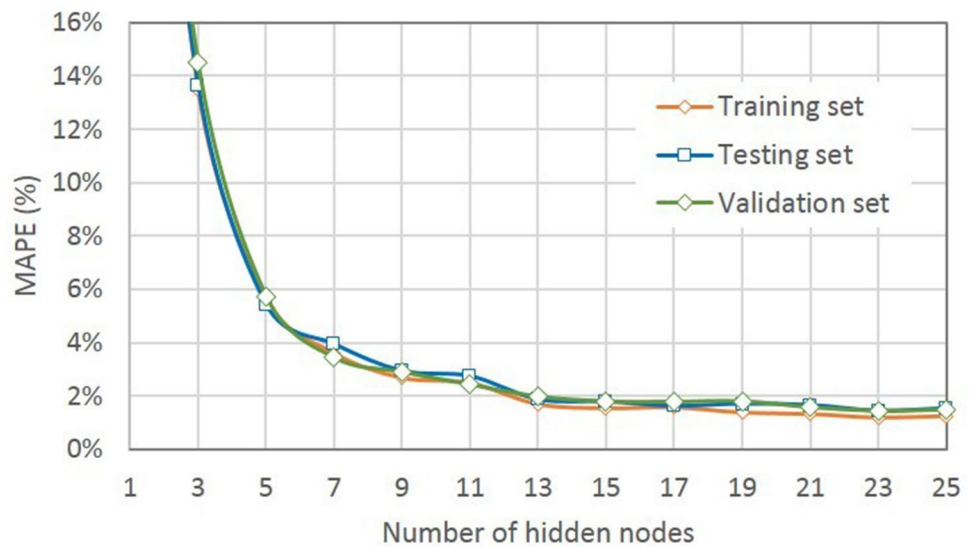
The structure of the optimum neural network is included in Fig. 11 showing an input layer with 8 neurons, a hidden layer of 7 neurons, and an output layer of 1 neuron.

The weights and biases of the network are detailed in Sect. 5.7.

The comparison between targets and predictions is superimposed in Fig. 12a, where a very good agreement is observed. The histogram of error values obtained predicting P_h with a 7-hidden-nodes ANN is also shown in Fig. 12b. The error is below 10% in most cases, never reaching a maximum of 20%.

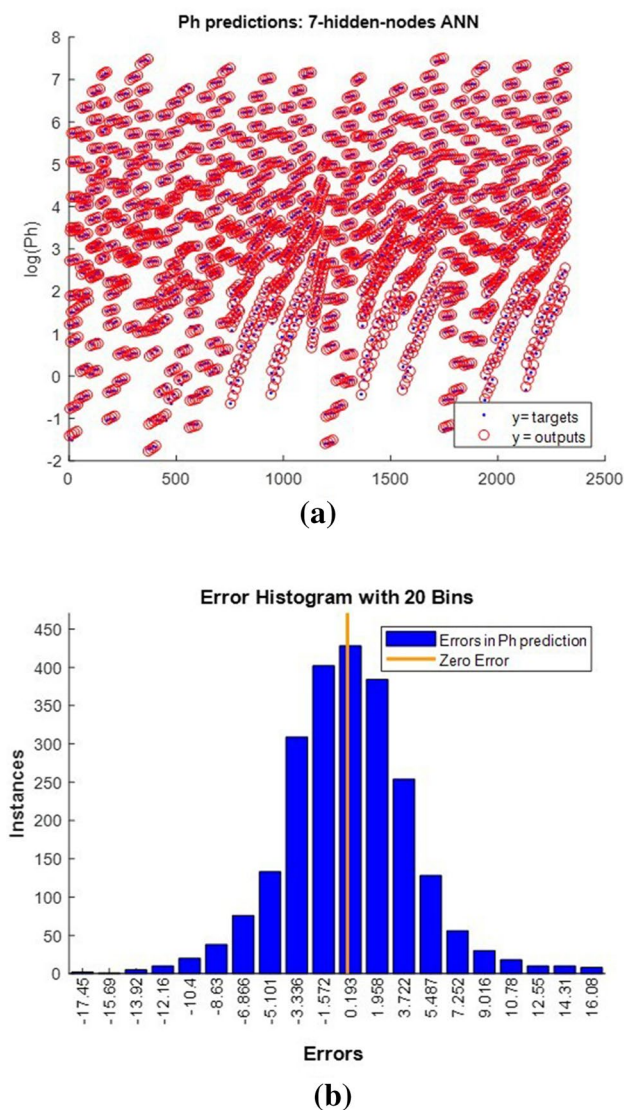**Fig. 11** Structure of the proposed optimum neural network
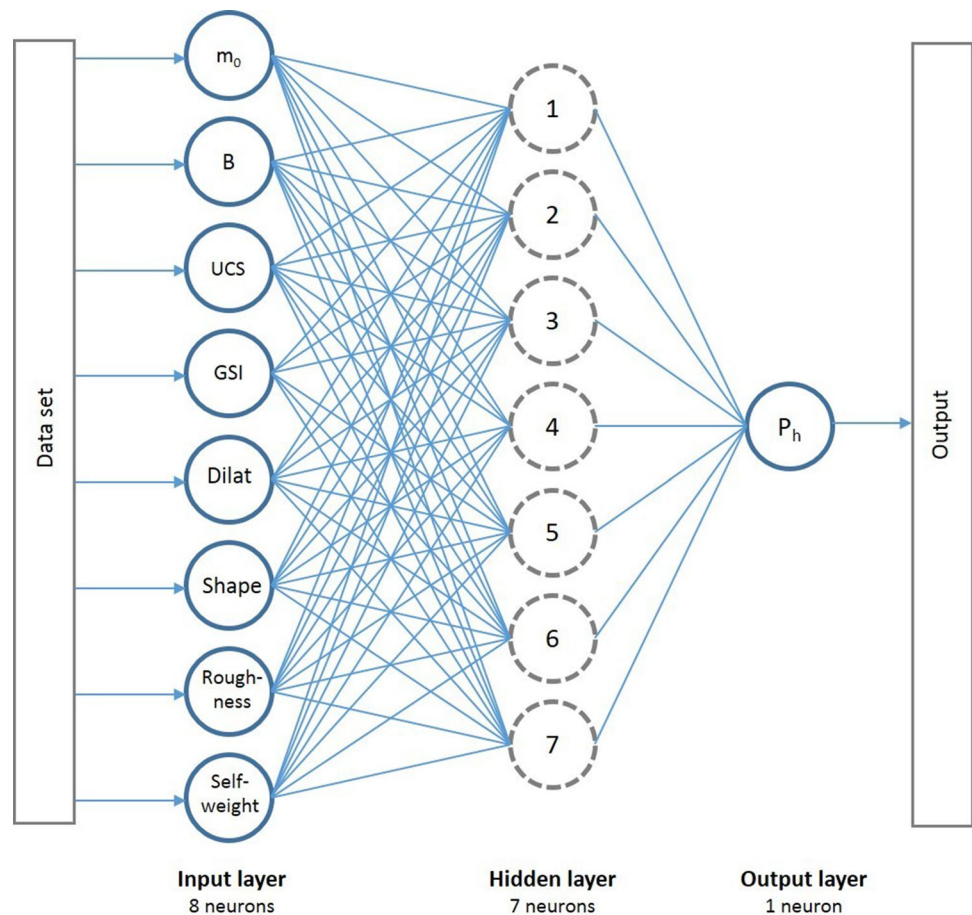
Fig. 12 P_h predictions (**a**) and histogram of error values (**b**) using an ANN with 7 hidden nodes

The plot of predicted versus calculated P_h bearing capacity with respect to the data in the training, testing, and validation sets, and all the sets simultaneously, are shown in Fig. 13a–d, respectively, where the solid line indicates equality and the regression coefficient is indicated above each sub-figure. According to the value of R, always higher than 0.999, it can be concluded that there exists a very good correlation between the model predictions and the target calculated values.

6.5 Model Optimization

Optimizing the network is addressed by the process of training or learning that can begin once the network weights and biases are initialized. It requires a set of

examples of the correct network behavior (network inputs and target outputs to compare to in successive training steps).

The training process consists of tuning the values of the weights and biases of the network to optimize network performance which is checked using the performance function. Training in feed-forward neural networks commonly uses the back-propagation algorithm (Rumelhart et al. 1986), which involves performing computations backward through the network. Its details can be found in every basic reference on neural networks (Fausett 1994).

The performance function used in this procedure is the mean square error (the average squared error between the series of network outputs x_i and the series of target outputs y_i: $\mathrm{MSE} = \frac{1}{N} \sum_{i=1}^{n} (x_i - y_i)^2$), though there are some other options. At each step in training, the gradient of the network performance with respect to the network weights is calculated. The training process will stop when one of the stopping criteria is reached, as explained in Sect. 5.6.

Different training algorithms may be applied to the network. It is difficult to know in advance the fastest method for a given problem, then trial and error seems to be the appropriate decision criterion. Among the algorithms, Quasi-Newton, resilient backpropagation, scaled conjugate gradient, and the Levenberg–Marquardt optimization schemes, offered as Matlab algorithms, were tried in this research. The Levenberg–Marquardt algorithm (Marquardt 1963) was selected since it produced faster and more accurate predictions in this problem. As a drawback, it requires more memory than other algorithms but that was not a limitation in this particular model.

6.6 Stopping Criteria

Different stopping criteria are used to decide when finishing the training of the network (Shahin et al. 2008). Using the Matlab built-in Levenberg–Marquardt procedure, the training stops when any of these conditions occur (MathWorks 2019b):

- The maximum number of steps is reached.
- The maximum amount of time is exceeded.
- The performance function is minimized to the goal.
- The performance gradient falls below a pre-fixed minimum.
- The momentum exceeds a pre-fixed maximum. In this algorithm, the momentum does not have a constant value but is decreased after each successful step from an initial value (0.01), aiming to a final value of 0 (in this case, the value of 1e−8 was reached). Using this procedure, the performance function is always reduced.
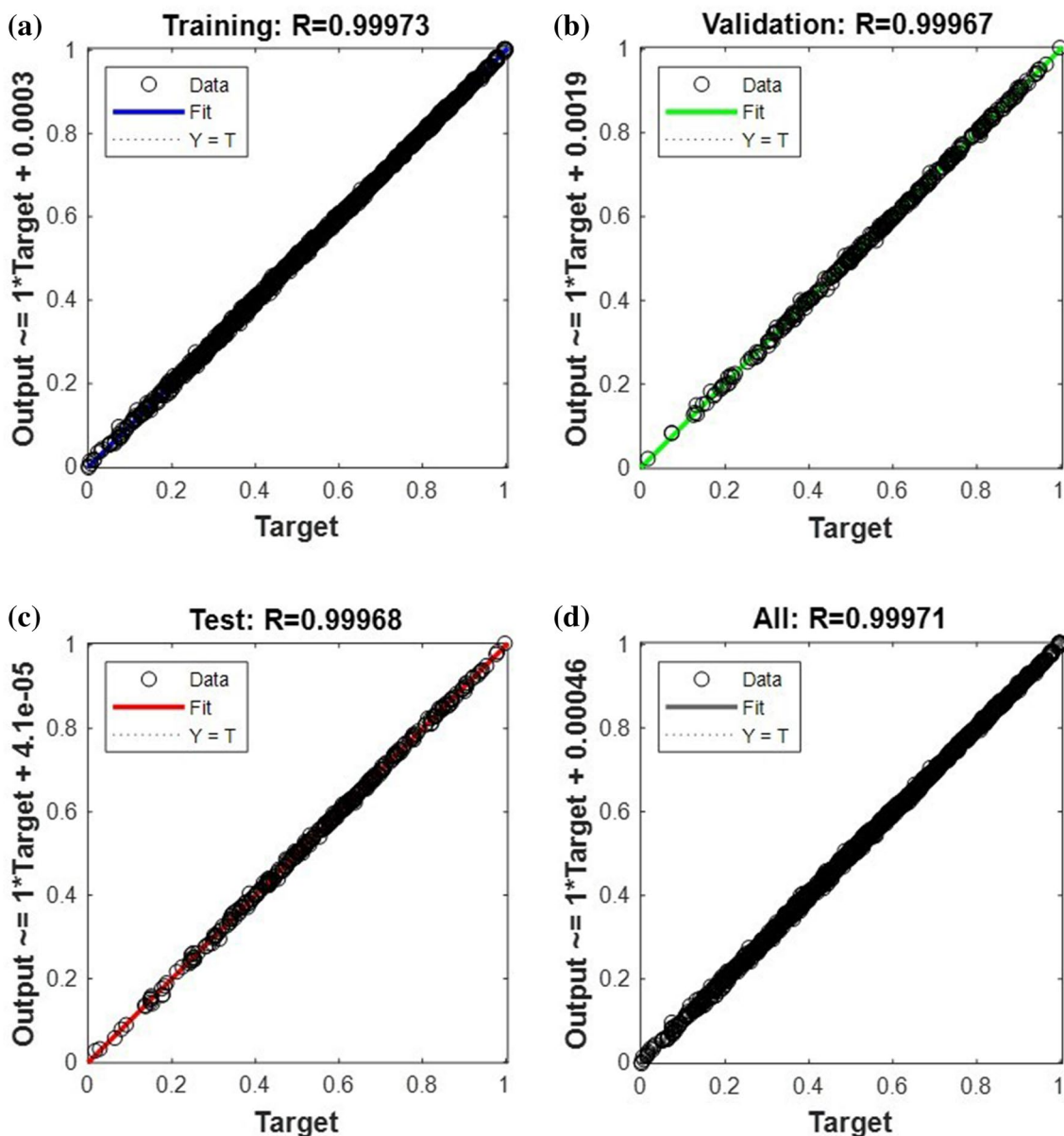
Fig. 13 Measured versus calculated P_h (target) for the optimal ANN model with respect to: **a** the training set, **b** validation set, **c** testing set, and **d** complete set

– The validation performance has increased more than a pre-fixed number of times since the last time it decreased (when using validation).

In this research, the input data is divided into training, testing, and validation subsets and cross-validation criteria (Stone 1974) can be used. This is the most valuable criteria to avoid over-fitting (Smith 1996). From the training and testing subsets, the performance of the network is checked at each step and used to adjust the connection weights. The network is then validated using the independent validation subset, examining if the performance improves as expected

or, on the contrary, deteriorates (which means over-fitting), in which case the training stops. To avoid false minima, the training continues several steps before concluding the process.

The performance evolution (mean square error) of the net through the training process is shown in Fig. 14a, as well as the gradient, the momentum μ, and the validation checks in Fig. 14b. The optimum performance was reached with 613 epochs (iterations), unless the training continued for 150 more epochs to avoid false minima (as the one found around epoch 400, as can be seen in the validation check figure at the bottom).
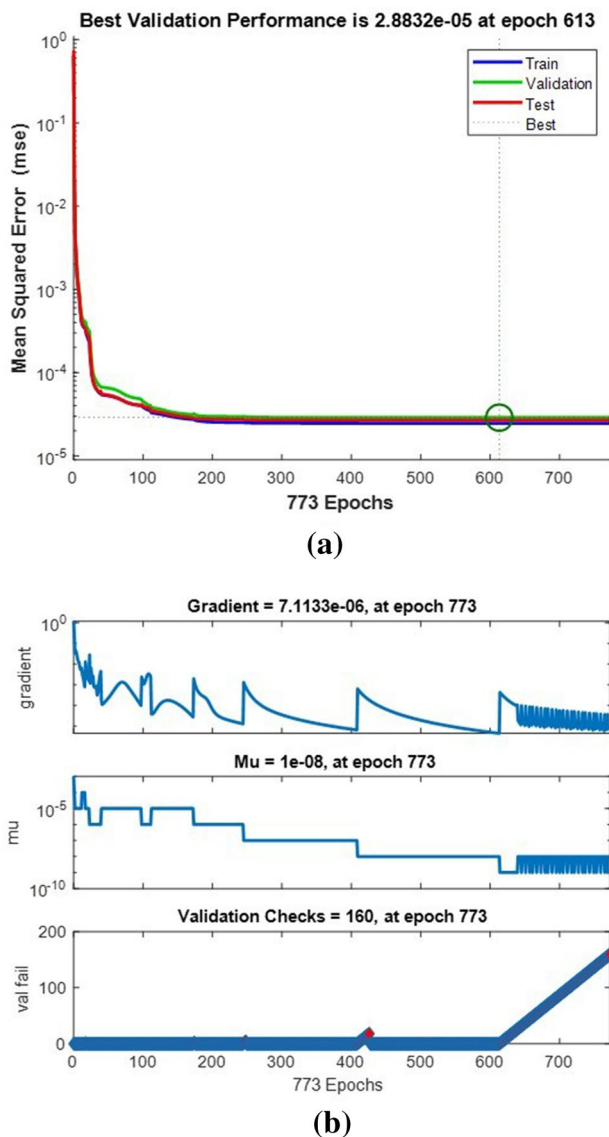
(a)

(b)

Fig. 14 Performance evolution during the training process. **a** Variation of the MSE with the number of epochs. **b** Evolution of variables involves in the training process

6.7 Weights Analysis of the ANN

The weights in the ANN represent the amount of each input signal that will be transmitted to the output. The neural network weight matrix obtained for the studied case (Table 6) can be used to estimate the relative importance of the various input variables on the output variable. The relative importance of input variables is calculated using Garson's algorithm (GA) (Garson 1991) of common use in the neural network literature. Garson's algorithm employs the absolute value of weights for calculation.

Figure 15 shows the relative importance of the different inputs using Garson's algorithm. GA shows a higher

contribution from m_0, UCS, and GSI. These results will be confirmed in the next section, where a sensitivity study is carried out.

6.8 Model Validation and Performance Measures

After training the model, a validation step should be performed to ensure, on one hand, its ability to generalize between the limits defined by the training data and, on the other hand, its robustness under a wide range of conditions (Shahin et al. 2008).

The simplest measure of performance is addressed using the already introduced statistical parameters as the correlation coefficient, r, the root mean square percentage error, RMSPE, and mean absolute percentage error, MAPE. This analysis was commented on in a previous section and is not repeated here.

However, validation against performance alone results in a network that may produce accurate predictions when using similar situations to the training data but may not be robust under different conditions. Shahin et al. (2005c) proposed to perform a sensitivity analysis to check the changes in the outputs when the inputs change. The robustness of the model is then analyzed examining how well model predictions are in agreement with the known underlying physical processes, that is to say, the changes in the predicted bearing capacity of shallow foundations on rock masses when changing the different parameters should agree with what is expected from the experience and other analytical and numerical methodologies.

The sensitivity analysis investigates the response of the neural network to a set of hypothetical input data that are generated over the ranges of the minimum and maximum data used to train the model. One variable at a time is changed over those ranges while all others are fixed to a certain chosen value. The process is repeated for each variable and different sets of synthetic inputs are generated by this procedure.

The UCS parameter is chosen as a reference for all the analysis since it is one of the most influential parameters on the foundation response and its influence is clearly known from the analytical formulation, which predicts that the bearing capacity grows proportionally to the UCS (Serrano et al. 2000).

Variation of results with m_0 and GSI against UCS are shown in Fig. 16, for the case of null dilatancy (Fig. 16a, c) and associated dilatance (Fig. 16b, d) keeping other parameters constant. It reveals that bearing capacity is highly dependent on the compressive strength of the rock UCS, increasing almost linearly as UCS increases. The growth of the bearing capacity with the UCS becomes more pronounced with a higher value of the parameters m_0 (Fig. 16a) and GSI (Fig. 16c) being the result much more sensitive

Table 6 Weights and biases corresponding to the optimal ANN

| Hidden layer nodes	Weight from node i in the input layer to node j in the hidden layer w_{ij}								Hidden layer bias θ_j
	m_0	B (m)	UCS (MPa)	GSI	Dilatancy	Plane strain	Roughness	Self-weight	
	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$	
$j=1$	0.0904	0.3748	−1.0865	−1.8340	−0.3503	−0.2675	0.0152	−5.1612	−0.4305
$j=2$	0.5123	0.0294	−0.0416	−0.0688	−0.0356	−0.0377	−0.0118	0.0456	−0.2945
$j=3$	−0.1519	−0.0658	0.1049	0.3028	0.0420	0.0591	−0.0074	−0.0906	−0.1662
$j=4$	−0.4427	0.1000	−0.1222	0.3550	−2.6524	0.3938	2.6979	0.0360	−2.0397
$j=5$	−0.0081	0.0160	4.2855	−0.0651	−0.0129	−0.0089	−0.0034	0.0333	2.0280
$j=6$	−0.2113	−0.0215	0.0236	−2.4740	−0.0189	−0.0722	−0.0587	−0.0719	−1.1319
$j=7$	0.2910	0.0462	−0.0620	−0.1653	−0.0247	−0.0436	−0.0001	0.0652	−0.0723
Output layer node	Weight from node j in the hidden layer to node k in the output layer w_{jk}								Output layer bias θ_k
	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$		
$k=1$	−0.1109	−6.1851	8.0002	0.0891	4.8032	−0.7105	15.3881		−4.6804
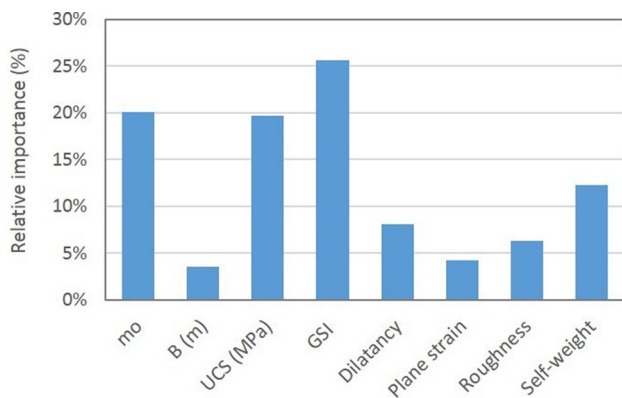
Fig. 15 Relative importance of the various input variables on the output variable

to the variation of the GSI than of the m_0, as reported in previous research (Serrano and Olalla 1994; Serrano et al. 2000; Merifield et al. 2006; Saada et al. 2008). The results presented for the case of null dilatancy (Fig. 16b, d) show a significant reduction of the bearing capacity relative to the associated dilatancy case (Fig. 16a, c). The reduction is more important for m_0 (Fig. 16b), reaching a 37% reduction for $m_0 = 32$, while for GSI (Fig. 16d) the reduction is about 24% for GSI = 85. Those results are consistent with the literature (Alencar et al. 2019).

The influence of B and self-weight is presented in Fig. 17 since they are closely related. They are obtained for the reference case: $m_0 = 12$, GSI = 10, associated dilatancy, 2D case, rough contact. When a weightless rock is considered (Fig. 17a), no difference is shown in the results for different values of B, as expected from the analytical solution of the problem. If self-weight is introduced

(Fig. 17b), a significant variation on the results appears, more evident for larger values of UCS, proving an important influence of B for the selected case (it is smaller for higher values of GSI).

AI figures show a coherent and smooth variation with the parameters, proving that the neural network offers a robust response in the input domain where it was trained.

The influence of the different hypotheses on the model (dilatancy, shape, roughness, and self-weight) are presented in Fig. 18, where an intermediate value of the parameters is chosen as reference ($m_0 = 10$, $B = 11$ m, GSI = 50) and each of the others is changed between 0 and 1 alternatively to address its influence on the results. In general, starting from the 2D case, with associate dilatancy, rough contact, and weightless rock (Dilat = 1, Sh = 0, $R = 1$, Wh = 0), the figure shows an important reduction in P_h when dilatancy changes to null, as can be expected. A smaller reduction is obtained when the contact changes from rough type to roughness type since a rough interface rigidizes the rock under the foundation, and only a slight increment is observed when the self-weight of the rock is considered, showing the increment of work necessary to move the wedge of failure. On the contrary, an important increase in P_h appears when an axisymmetric foundation is considered instead of the 2D one, which is coherent with the well-known fact in geotechnical engineering that the bearing capacity is smaller in the 2D one as theoretical results show. All results are consistent with previous research on the subject (Alencar et al. 2019).

The previous results confirm that the optimum network performs well and can be successfully used to coherently predict the bearing capacity of shallow foundations on rock masses.
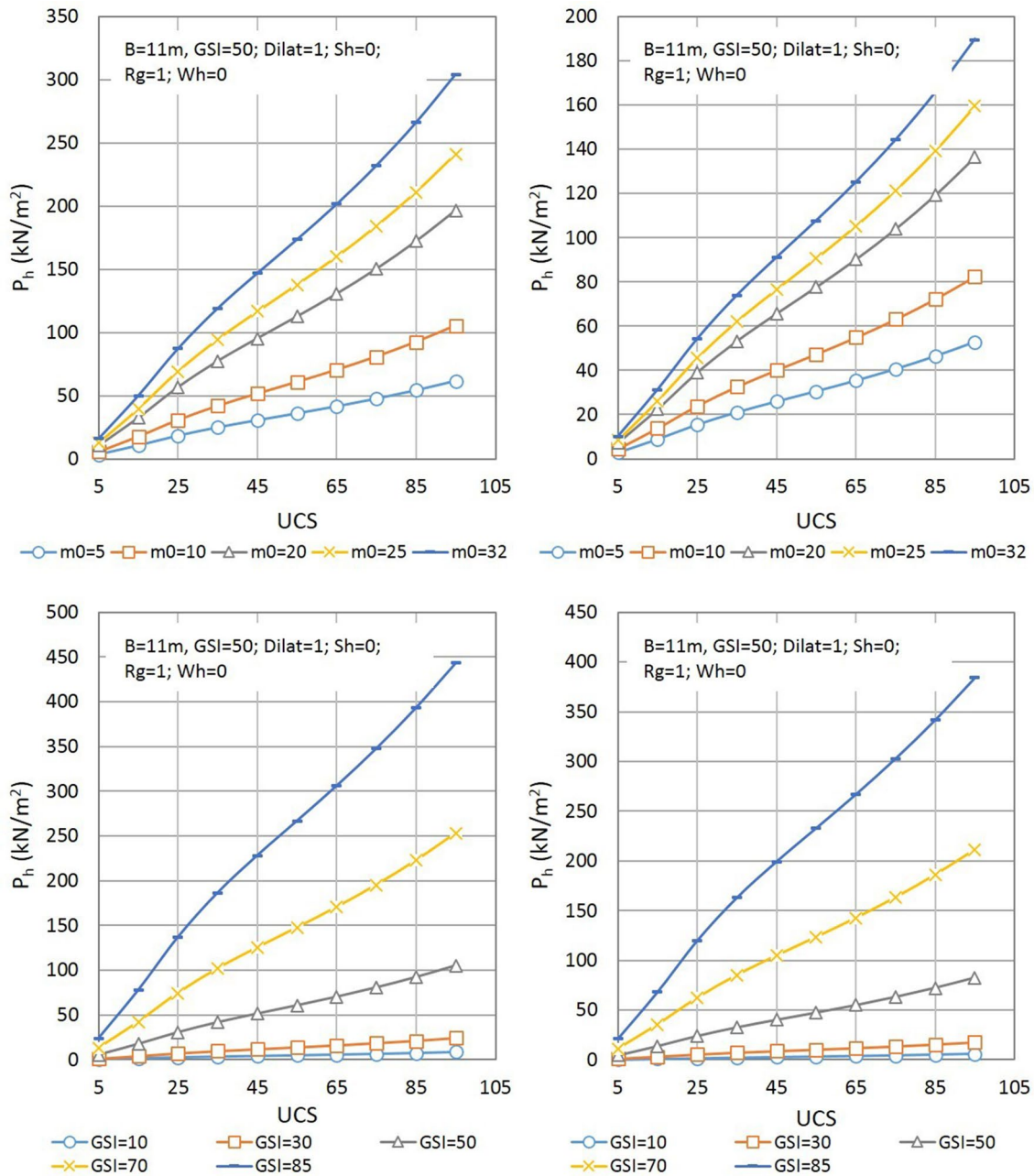
Fig. 16 Variation of the bearing capacity P_h. Relation between m_0 and UCS for the associated (**a**) and null dilatancy (**b**) cases. Relation between GSI and UCS for the associated (**c**) and null dilatancy (**d**) cases. Variable codes are dilat: dilatancy, Sh: shape, Rg: roughness, Wh: self-weight

7 Results and Discussion

The main result of this research is the ANN model that is conveniently converted into a set of simple equations to facilitate its use, being easily implemented in a spreadsheet. The detailed equations are presented in the following, as well as some discussion about its scope and limitations.

The optimal ANN previously developed has 8 inputs, 7 neurons in the hidden layer, and only one output, as shown schematically in Fig. 11.

The equations are obtained from the weights and biases provided by the Matlab software (applying to the inputs and one output the convenient scaling and re-scaling process), using the following expression (Ranasinghe et al. 2017):

$$y_n = f_{\text{outp}} \left\{ \theta_k + \sum_1^7 \left\{ w_{kj} \cdot f_{\text{hidden}} \left[\sum_1^8 \left(w_{ji} \cdot x_{ni} \right) + \theta_j \right] \right\} \right\},$$
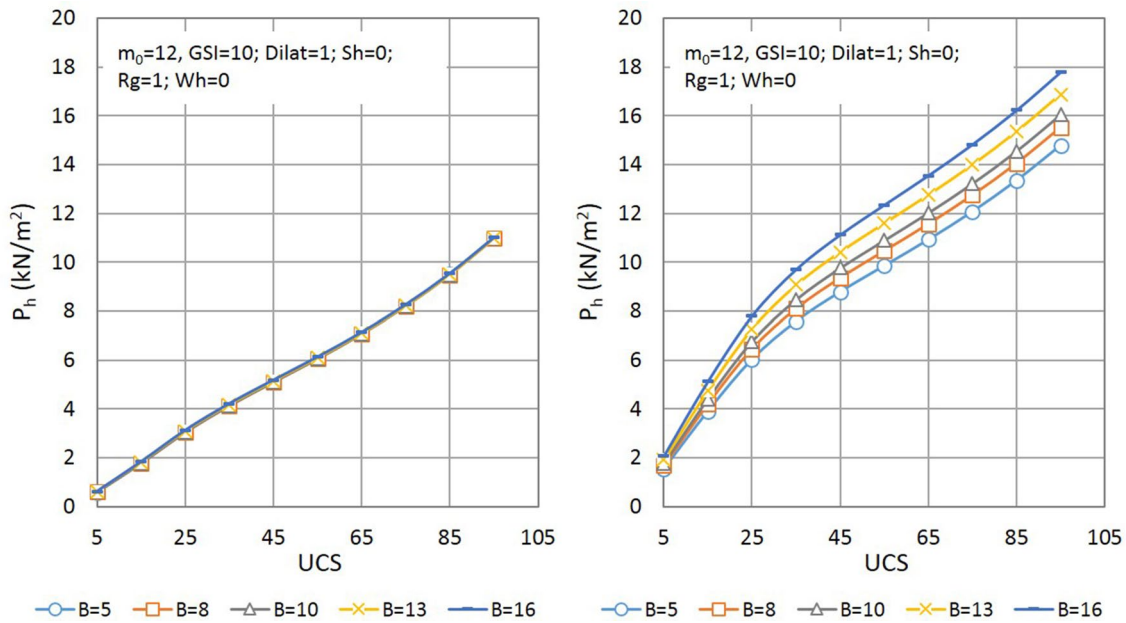
(7)

Fig. 17 Variation of the bearing capacity P_h relation between B (m) and UCS for weightless (**a**) and self-weighted rock (**b**) for the reference case: $m_0 = 12$, GSI = 10, Dilat = 1, Sh = 0, Rg = 1. Relation. Variable codes are dilat: dilatancy, Sh: shape, Rg: roughness, Wh: self-weight
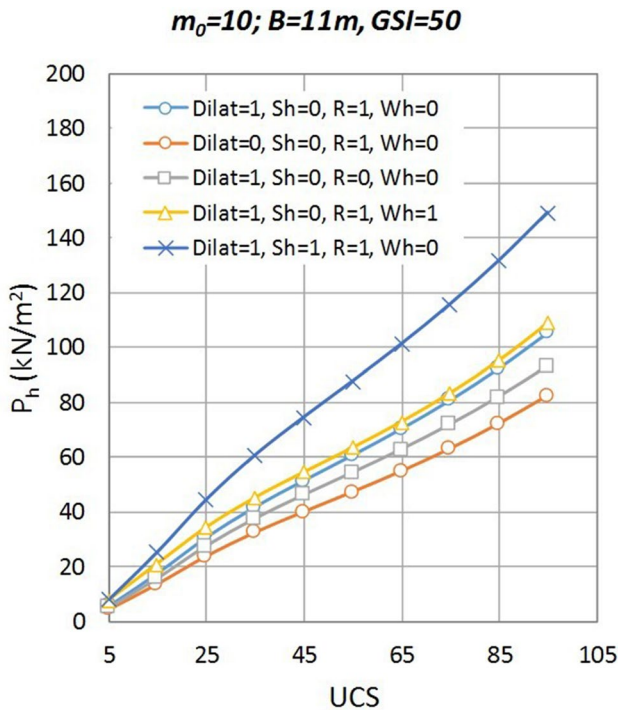

Fig. 18 Variation of the bearing capacity P_h with respect to the different hypotheses in the model for the reference case: $m_0 = 10$, $B = 11$ m, GSI = 50

where y_n is the single normalized output variable, representing the bearing capacity; (θ_k) is the threshold value at the output layer and (w_{kj}) is the connection weight between the

jth node in the hidden layer and the kth node in the output layer; (θ_j) is the threshold value of the jth hidden node and (w_{ji}) is the connection weight between the ith input node and the jth hidden node; (x_{ni}) is the ith normalized input variable; and (f_{outp}) and (f_{hidden}) are the transfer functions at the hidden and output nodes respectively.

The previous expression (7) can be implemented recursively, considering that f_{hidden} correspond to the tangent-sigmoid function $\left(\frac{2}{1+e^{-2 \cdot t}}\right) - 1$, and f_{outp} to the linear function, as:

$$y_n = f_{outp}\left\{\theta_k + \sum_{1}^{7}\{w_{kj} \cdot T_j\}\right\} = \theta_k + \sum_{1}^{7}\{w_{kj} \cdot T_j\}, \quad (8)$$

$$T_j = f_{hidden}\left\{\theta_j + \sum_{1}^{8}\{w_{ji} \cdot x_{ni}\}\right\} = \frac{2}{1 + e^{-2\left[\theta_j + \sum_{1}^{8}\{w_{ji} \cdot x_{ni}\}\right]}} - 1. \quad (9)$$

Note that the inputs x_{ni} in Eq. (9) are the normalized values. The weights and biases provided by the Matlab software corresponding to the optimal ANN are detailed in Table 6.

It is important to remember that reverse normalization should be applied to Eq. (8) to obtain the desired value of bearing capacity P_h. Considering that the variables are normalized in the range $(h_{min}, h_{max}) = (0,1)$ the following expressions should be used:

– Direct normalization of input values x_i to normalized values x_{ni}:

$$x_{ni} = \frac{(x_i - x_{\min-i})}{x_{\max-i} - x_{\min-i}}(h_{\max} - h_{\min}) + h_{\min}$$

$$= \frac{(x_i)}{x_{\max-i} - x_{\min-i}}. \tag{10}$$

– Reverse normalization of the normalized output value y_n to the output value y:

$$y = (y_n - h_{\min}) \cdot \frac{(y_{n-\max} - y_{n-\min})}{(h_{\max} - h_{\min})} + y_{n-\min}$$

$$= 9.179(y_n) - 1.715. \tag{11}$$

– Reverse log conversion of the output value y to obtain the final limit load:

$$P_h = e^y = e^{9.179(y_n) - 1.715}. \tag{12}$$

Being $x_{\max-i}$, $x_{\min-i}$ the maximum and minimum value in the ith-parameter input data (x_i) used to train the neural network that can be found in Table 3.

From the results obtained in the previous section, the equation representing the optimal ANN (already including the double normalization previously explained) can be expressed as:

– Parameters with a numerical value: m_0, UCS (MPa), GSI, B (m) can be given any value between inputs limits. For example, for m_0 the adopted values in the net were (5/12/20/32) but any value between 5 and 32 can be used as input for new predictions. The use of (5/12/20/32) for training assures a higher precision.

– Parameters that cannot be related homogeneously to a numerical value:

o Self-weight. Although a 0/1 value was adopted for this parameter, it is related to 0 and 26 kN/m³ real values. However, interpolation between those values is not recommended, since the predictions may not be accurate since only those two extreme cases were used to train the net.

o Dilatancy. This case is different from self-weight because associated dilatancy does not correspond to a unique value but it depends on the stress level (angle of the tangent to the Hoek and Brown criterion at the point of failure). Interpolation for dilatancy values other than 0–1 may give inaccurate results.

– No interpolation corresponds to shape and contact roughness.

$$P_h = e^{9.179(-0.1109T_1 - 6.1851T_1 + 8.0002T_1 + 0.0891T_1 + 4.8032T_1 + -0.7105T_1 + 15.3881T_1 - 4.6804) - 1.715}, \tag{13}$$

where the different terms T_1 to T_7 are obtained using Eq. (9). As an example, the term T_1 is shown in full, being x_{ni} the normalized (not original) input values:

From the previous discussion, it can be followed that the case of a shallow foundation on a saturated rock with the phreatic level at the ground surface can be also studied using

$$T_1 = -1 + \frac{2}{1 + e^{-2(0.0904x_{n1} + 0.3748x_{n2} - 1.0865x_{n3} - 1.8340x_{n4} - 0.3503x_{n5} - 0.2675x_{n6} + 0.0152x_{n7} - 5.1612x_{n8} - 0.4305)}}. \tag{14}$$

Using the previous equations, the trained ANN is easily reproduced and a very complex problem that would demand an elaborate model to be solved is straightforwardly addressed.

Some parameters as those corresponding to the modified Hoek and Brown criterion or dilatancy are not easily available in some software solutions, making the proposed ANN especially powerful and useful.

It is important to remark that the scope of the proposed ANN is restricted between the limits adopted for the input parameters. Predictions made outside those limits may not be accurate.

Although interpolation between input limits is available and the results will produce relatively accurate results, it should not be used for non-homogeneous inputs that not are associated with a real numerical value. A more detailed explanation is for the different cases:

the network by assigning to the self-weight parameter the specific weight of the submerged rock.

8 Summary and Conclusions

This research deal with the calculation of the bearing capacity of shallow foundations on rock masses due to general shear failure. Different tools can be used to address the problem, having either excessive simplification (empirical equations, analytical solutions) or are quite complex and requires high expertise (numerical models). A different approach is available nowadays, neural networks and soft computing, that can overcome both limitations.

An artificial neural network (ANN) is created to predict the aforementioned bearing capacity of shallow foundations

Table 7 Data normalization in the proposed example

	m_0	B (m)	UCS (MPa)	GSI	Dilatancy	Plane strain	Roughness	Self-weight
	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$
Original inputs	20	11	50	85	1	0	1	1
Normalized inputs	0.5556	0.3714	0.4737	1	1	0	1	1

on rock masses, based on a large number of numerical calculations used as input data for training the net.

Since based on complex numerical calculations using FLAC software, it will have the same capacity that the substituted method if uses the same main representative parameters and assures a good accuracy of predictions by adequate building and training.

The inputs used in the ANN are three parameters characterizing the rock mass as rock type (m_0), uniaxial compressive strength (UCS), and geological strength index (GSI), the foundation width (B), the dilatancy on the failure surface, bidimensional or axisymmetric problem, rough or roughless foundation-rock contact, and weightless or self-weighted rock mass. The only output is the bearing capacity (Ph). Most professional foundation problems can be addressed properly using the previous set of parameters, even covering a wider range (as considering dilatancy).

Once inputs and outputs are defined, both building and training the net are required. The structure and the optimal size of the network are obtained by comparing the results for a net with a different number of hidden nodes (from 3 to 25), training functions, and transfer functions. Each architecture is trained and its predictions compared to the target values, analyzing the network performance, comparing predictions using the ANN with the target values, using the correlation coefficient, r, the root mean square percentage error, RMSPE, and mean absolute percentage error, MAPE. The final optimal architecture is a single hidden layer with 7 nodes, balancing that the previous global performance measures are in all cases around or below 5% and that the number of nodes is not very high. The individual prediction error is below 10% in most cases, never reaching a maximum of 20%.

This network can deal with similar problems as an elaborate numerical model and assures good accuracy. A sensitivity analysis is developed to ensure de robustness of the proposed model. Both show a very good agreement with the expected results proving that the model can be used to predict the bearing capacity. Its only limitation is the need to use input values within the limits defined to train the network as explained in the paper.

Finally, the neural network is converted into a set of simplified design equations to facilitate the use of the model through hand or spreadsheet calculations. By using these equations, no specific training or experience is needed to solve the problem (unlike with numerical methods) beyond the necessary knowledge to properly define the input parameters and correctly interpret the obtained bearing capacity.

Appendix: Illustrative numerical example

To illustrate the detailed use of the previous bearing capacity equation, a simple case of a shallow foundation on a rock mass having $m_0 = 20$, B (m) $= 11$, UCS $= 50$ MPa, and GSI $= 85$. We also consider dilatancy 1 (associated), plane strain (0), rough contact (1), and self-weighted rock (1).

The following steps should be followed:

The input data should be scaled to the (0, 1) range, using Eq. (10):

$$x_{ni} = \frac{(x_i)}{x_{\max-i} - x_{\min-i}}. \tag{15}$$

Being x_{ni} and x_i the obtained normalized value and the original value, respectively; x, $x_{\max-i}$ and $x_{\min-i}$ the original, maximum and minimum values of the original data, respectively.

The original and normalized values for the studied case are shown in Table 7.

Calculation of the intermediate terms corresponding to the hidden-neuron outputs, T_1 to T_7, following Eq. (14), (15) and similar expressions:

$$
\begin{aligned}
T_1 &= -1 + 2/\left\{1 + \exp\left[-2 \times \left(0.0904 \times x_1 + 0.3748 \times x_2 - 1.0865 \times x_3 - 1.8340 \times x_4\right.\right.\right. \\
&\quad \left.\left.\left. -0.3503 \times x_5 + -0.2675 \times x_6 + 0.0152 \times x_7 - 5.1612 \times x_8 - 0.4305\right)\right]\right\} \\
&= -1 + 2/\{1 + \exp[-2 \times (0.09043 \times 0.5556 + 0.37487 \times 0.3714 - 1.0865 \times 0.4737 \\
&\quad -1.8340 \times 1 + -0.3503 \times 1 - 0.2675 \times 0 + 0.01512 \times 1 - 5.1612 \times 1 - 0.43045)]\} \\
&= -1.0000
\end{aligned}
$$

Similar results are obtained for the rest of the terms: $T_2 = -0.0891$; $T_3 = 0.0215$; $T_4 = -0.9536$; $T_5 = 0.9993$; $T_6 = -0.9991$, and $T_7 = -0.0477$.

Obtain the final limit load from Eq. (13) (this step includes the reverse normalization of the normalized output value y_n, from the final output neuron to the output value y, and the reverse log conversion of the output value y to obtain the final limit load):

$$\begin{aligned}
P_h &= \exp\big[9.17 \times (-0.1108 \times T_1 - 6.1851 \times T_2 \\
&\quad + 8.0001 \times T_3 + 0.089 \times T_4 \\
&\quad + 4.8032 \times T_5 + -0.7105 \times T_6 \\
&\quad + 15.388 \times T_7 - 4.6803) - 1.715\big] \\
&= \exp[9.17 \times (-0.1108 \times -1.0000 \\
&\quad -6.1851 \times -0.0891 + 8.0001 \times 0.0215 \\
&\quad + 0.089 \times -0.9536 + 4.8032 \times 0.9993 \\
&\quad -0.7105 \times -0.9991 \\
&\quad + 15.388 \times -0.0477 - 4.6803) - 1.715] \\
&= 416.3 \text{ kN.}
\end{aligned}$$

The predicted result offers a 3.4% error compared to the target value, equal to 402.7 kN.

Author contributions Conceptualization: MAM and RG; methodology: MAM; formal analysis and investigation: RG; numerical calculations: AA, ANN development: MAM; writing—original draft preparation: MAM; writing—review and editing: R. Galindo.

Availability of data and material (data transparency) Not applicable.

Code availability (software application or custom code) Not applicable.

Declarations

Conflict of interest There is no conflict of interest to declare that are relevant to the content of this article.

References

AASHTO (2007) LRFD bridge design specifications, 4th edn. AASHTO, Washington DC

Alavi AH, Sadrossadat E (2016) New design equations for estimation of ultimate bearing capacity of shallow foundations resting on rock masses. Geosci Front 7:91–99

Alencar A, Galindo R, Melentijevic S (2019) Bearing capacity of foundation on rock mass depending on footing shape and interface roughness. Geomech Eng 18(4):11–21. https://doi.org/10.12989/gae.2019.18.4.391

Bishoni BL (1968) Bearing capacity of a closely jointed rock. Ph.D. Dissertation. Institute of Technology, Georgia

Bowles JE (1996) Foundation analysis and design, 5th edn. McGraw-Hill Inc, New York

Brinch-Hansen JA (1970) Revised and extended formula for bearing capacity. Bulletin Nº 28, Danish Geotechnical Institute Copenhagen, pp 5–11

Canadian Geotechnical Society (2006) Canadian Foundation Engineering Manual (Montreal, Quebec), 4th edn

Carter P, Kulhawy FH (1988) Analysis and design of foundations socketed into rock. Report No. EL-5918. Empire State Electric Engineering Research Corporation and Electric Power Research Institute, New York

Caudill M (1988) Neural networks primer, part III. AI Expert 3(6):53–59

Chakraborty M, Kumar J (2015) Bearing capacity of circular footings over rock mass by using axisymmetric quasi lower bound finite element limit analysis. Comput Geotech 70:138–149. https://doi.org/10.1016/j.compgeo.2015.07.015

Chen Y, Azzam R, Zhang F (2006) The displacement computation and construction pre-control of a foundation pit in Shanghai utilizing FEM and intelligent methods. Geotech Geol Eng 24(6):1781–1801. https://doi.org/10.1007/s10706-006-6807-6

Eurocode 7 (CEN) (2004) Geotechnical design—part 1: general rules

Fausett LV (1994) Fundamentals neural networks: architecture, algorithms, and applications. Prentice-Hall, Englewood Cliffs

Garson GD (1991) Interpreting neural-network connection weights. AI Expert 6(7):47–51

Gholami R, Rasouli V, Alimoradi A (2013) Improved RMR rock mass classification using artificial intelligence algorithms. Rock Mech Rock Eng 46:1199–1209. https://doi.org/10.1007/s00603-012-0338-7

Goodman RE (1989) Introduction to rock mechanics, 2nd edn. Wiley, New York

Griffiths DV (1982) Computation of bearing capacity factors using finite elements. Geotechnique 32(3):195–202. https://doi.org/10.1680/geot.1982.32.3.195

Hoek E (1983) Strength of jointed rock masses. Geotechnique 33(3):187–222

Hoek E, Brown ET (1980) Empirical strength criterion for rock masses. J Geotech Eng Div ASCE 106(9):1013–1035

Hoek E, Brown ET (1997) Practical estimates of rock mass strength. Int J Rock Mech Min Sci 34(8):1165–1186

Hoek E, Carranza-Torres C, Corkum B (2002) Hoek-Brown failure criterion—2002 edition. In: Hammah R, Bawden W, Curran J, Telesnicki M (eds) Proceedings of NARMS-TAC, mining innovation and technology, Toronto, 10 July 2002, pp 267–73

Hornik K, Stinchcombe M, White H (1989) Multilayer feed forward networks are universal approximators. Neural Netw 2:359–366

Itasca Consulting Group, Inc FLAC (2012) Fast Lagrangian analysis of continua, Version 7.0 Fluid-Mechanical Interaction. Minneapolis

Kumar J, Khatri VN (2011) Bearing capacity factors of circular foundations for a general c–ϕ soil using lower bound finite elements limit analysis. Int J Numer Anal Meth Geomech 35(3):393–405

Kumar J, Mohapatra D (2017) Lower-bound finite elements limit analysis for Hoek-Brown materials using semidefinite programming. J Eng Mech 143(9):04017077. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001296

Kuo YL, Jaksa MB, Lyamin AV, Kaggwa WS (2009) ANN-based model for predicting the bearing capacity of strip footing on multi-layered cohesive soil. Comput Geotech 36:503–516

Lyamin AV, Sloan SW (2002a) Lower bound limit analysis using non-linear programming. Int J Numer Methods Eng 55(5):573–611. https://doi.org/10.1002/nme.511

Lyamin AV, Sloan SW (2002b) Upper bound limit analysis using linear finite elements and non-linear programming. Int J Numer Anal Methods Geomech 26(2):181–216. https://doi.org/10.1002/nag.198

Makrodimopoulos A, Martin CM (2006) Lower bound limit analysis of cohesive-frictional materials using second-order cone programming. Int J Numer Methods Eng 6:604–634. https://doi.org/10.1002/nme.1567

Marquardt D (1963) An algorithm for least-squares estimation of non-linear parameters. SIAM J Appl Math 11(2):431–441

Masters T (1993) Practical neural network recipes in C++. Academic, San Diego

Mathworks Inc (2019a) Matlab, R2019a (computer program)

MathWorks Inc (2019b) Deep learning toolbox: user's guide (R2019a). Trainlm. Levenberg-Marquardt backpropagation. https://es.mathworks.com/help/deeplearning/ref/trainlm.html. Accessed 30 Sept 2020

Merifield RS, Lyamin AV, Sloan SW (2006) Limit analysis solutions for the bearing capacity of rock masses using the generalised Hoek-Brown criterion. Int J Rock Mech Min Sci 43:920–937. https://doi.org/10.1016/j.ijrmms.2006.02.001

Mert E, Yilmaz S, Inal M (2011) An assessment of total RMR classification system using unified simulation model based on artificial neural networks. Neural Comput Appl 20:603–610. https://doi.org/10.1007/s00521-011-0578-6

Meyerhof GG (1951) The ultimate bearing capacity of foundations. Geotechnique 2(4):301–332. https://doi.org/10.1680/geot.1951.2.4.301

Ocak I, Seker SE (2012) Estimation of elastic modulus of intact rocks by artificial neural network. Rock Mech Rock Eng 45:1047–1054. https://doi.org/10.1007/s00603-012-0236-z

Padmini D, Ilamparuthi K, Sudheer KP (2008) Ultimate bearing capacity prediction of shallow foundations on cohesionless soils using neurofuzzy models. Comput Geotech 35:33–46

Paikowsky S, Cannif M, Lensy K, Aloys K, Amatya S, Muganga R (2010) NCHRP Report 651: LRFD design and construction of shallow foundations for highway bridge structures. Transportation Research Board of the National Academies

Provenzano P, Ferlisi S, Musso A (2004) Interpretation of a model footing response through an adaptive neural fuzzy inference system. Comput Geotech 31(3):251–266

Ranasinghe RATM, Jaksa MB, Kuo YL, Pooya F (2017) Application of artificial neural networks for predicting the impact of rolling dynamic compaction using dynamic cone penetrometer test results. J Rock Mech Geotech Eng 9(2):340–349. https://doi.org/10.1016/j.jrmge.2016.11.011

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representation by error propagation. In: Rumelhart DE, McClelland JL (eds) Parallel distributed processing. MIT Press, Cambridge

Saada Z, Maghous S, Garnier D (2008) Bearing capacity of shallow foundations on rocks obeying a modified Hoek-Brown failure criterion. Comput Geotech 35:144–154. https://doi.org/10.1016/j.compgeo.2007.06.003

Serrano A, Olalla C (1994) Ultimate bearing capacity of rock masses. Int J Rock Mech Min Sci Geomech Abstr 31:93–106. https://doi.org/10.1016/0148-9062(95)00081-X

Serrano A, Olalla C, González J (2000) Ultimate bearing capacity of rock masses based on the modified Hoek-Brown criterion. Int J Rock Mech Min Sci 37:1013–1018. https://doi.org/10.1016/S1365-1609(00)00028-9

Shahin MA, Jaksa MB, Maier HR (2001) Artificial neural network application in geotechnical engineering. Aust Geomech 36(1):49–62

Shahin MA, Jaksa MB, Maier HR (2002a) Artificial neural network based settlement prediction formula for shallow foundations on granular soils. Aust Geomech 37(4):45–52

Shahin MA, Maier HR, Jaksa MB (2002b) Predicting settlement of shallow foundations using neural networks. J Geotech Geoenviron Eng ASCE 128(9):785–793

Shahin MA, Jaksa MB, Maier HR (2003a) Neurofuzzy networks applied to settlement of shallow foundations on granular soils. In: Proceedings of the 9th international conference on applications of statistics and probability in civil engineering, ICASP9, San Francisco, Millpress, Rotterdam, pp 1379–1383

Shahin MA, Jaksa MB, Maier HR (2003b) Settlement prediction of shallow foundations on granular soils using B-spline neurofuzzy models. Comput Geotech 30(8):637–647

Shahin MA, Jaksa MB, Maier HR (2004a) Application of neural networks in foundation engineering. In: International e-conference on modern trends in foundation engineering: geotechnical challenges & solutions. http://www.civil.iitm.ac.in

Shahin MA, Maier HR, Jaksa MB (2004b) Data division for developing neural networks applied to geotechnical engineering. J Comput Civ Eng ASCE 18(2):105–114

Shahin MA, Jaksa MB, Maier HR (2005a) Neural network based stochastic design charts for settlement prediction. Can Geotech J 42(1):110–120

Shahin MA, Jaksa MB, Maier HR (2005b) Stochastic simulation of settlement of shallow foundations based on a deterministic neural network model. In: Proceedings of the international congress on modeling and simulation, MODSIM 2005, Melbourne (Australia), pp 73–78

Shahin MA, Maier HR, Jaksa MB (2005c) Investigation into the robustness of artificial neural network models for a case study in civil engineering. In: Proceedings of the international congress on modeling and simulation, MODSIM 2005, Melbourne (Australia), pp 79–83

Shahin M, Jaksa M, Maier H (2008) State of the art of artificial neural networks in geotechnical engineering. Electron J Geotech Eng 8:1–26

Sivakugan N, Eckersley J, Li H (1998) Settlement predictions using neural networks. Aust Civ Eng Trans 40:49–52. https://search.informit.com/doi/10.3316/informit.208179908890899

Sloan SW (1988) Lower bound limit analysis using finite elements and linear programming. Int J Numer Anal Methods Geomech 12:61–77. https://doi.org/10.1002/nag.1610120105

Sloan SW, Kleeman PW (1995) Upper bound limit analysis using discontinuous velocity fields. Comput Methods Appl Mech Eng 127:293–314. https://doi.org/10.1016/0045-7825(95)00868-1

Smith M (1996) Neural networks for statistical modeling. International Thomson Computer Press, Boston

Smith CC, Gilbert M (2007) Application of discontinuity layout optimization to plane plasticity problems. Proc R Soc A 463(2086):2461–2484

Sokolovskii VV (1965) Statics of soil media. Butterworths Science, London (**Translator R. Jones & A. Schofield**)

Sowers GF (1979) Introductory soil mechanics and foundations: geotechnical engineering, 4th edn. MacMillan, New York

Stone M (1974) Cross-validatory choice and assessment of statistical predictions. J R Stat Soc B 36:111–147

Sutcliffe D, Yu HS, Sloan SW (2004) Lower bound solutions for bearing capacity of jointed rock. Comput Geotech 31(1):23–36. https://doi.org/10.1016/j.compgeo.2003.11.001

Terzaghi K (1943) Theoretical soil mechanics. Wiley, New York

Yang Y, Zhang Q (1997) A hierarchical analysis for rock engineering using artificial neural networks. Rock Mech Rock Eng 30:207–222. https://doi.org/10.1007/BF01045717

Yang X, Yin J (2005) Upper bound solution for ultimate bearing capacity with a modified Hoek-Brown failure criterion. Int J Rock Mech Sci 42:550–560. https://doi.org/10.1016/j.ijrmms.2005.03.002

Yılmaz I, Yuksek A (2008) An example of artificial neural network (ANN) Application for indirect estimation of rock parameters. Rock Mech Rock Eng 41:781–795. https://doi.org/10.1007/s00603-007-0138-7

Zheng X, Booker JR, Carter JP (2000) Limit analysis of the bearing capacity of fissured materials. Int J Solids Struct 37(8):1211–1243. https://doi.org/10.1016/S0020-7683(98)00282-0

Ziaee SA, Sadrossadat E, Alavi AH, Shadmehri DM (2015) Explicit formulation of bearing capacity of shallow foundations on rock masses using artificial neural networks: application and supplementary studies. Environ Earth Sci 73:3417–3431. https://doi.org/10.1007/s12665-014-3630-x

Zurada JM (1992) Introduction to artificial neural systems. West Publishing Company, St. Paul

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.