Probabilistic Engineering Mechanics 71 (2023) 103400

Contents lists available at ScienceDirect

Probabilistic Engineering Mechanics

journal homepage: www.elsevier.com/locate/probengmech

Generic compressive strength prediction model applicable to multiple lithologies based on a broad global database

Herbert Muzamhindo [a], Maria Ferentinou [b,*]

[a] University of Johannesburg, South Africa
[b] Liverpool John Moores University, United Kingdom

ARTICLE INFO

Keywords:
ROCK/10/4025/
Compressional strength prediction
Multiple lithologies
ANFIS
ANN
Classification
SVM

ABSTRACT

In this study, we present a global database of ten parameters, which include measurements of rock index properties, strength stiffness and dynamic properties. Hoek–Brown constant m_i, is included, and was estimated, using Hoek and Brown proposed guidelines for determining m_i values for different rock types that can be used for preliminary design when triaxial tests are not available. This broad database is compiled from 96 studies and is labelled as "ROCK/10/4025", to describe the type of geomaterial, the number of the parameters, and the number of the data samples included. It consists of 35.4 % igneous, 54.8 % sedimentary, and 9.2% metamorphic rocks. The purpose of this paper is to propose a generic soft computing model applicable to multiple lithologies, that can become more reliable and perhaps more suitable for a specific site study when used in order to densify often limited similar site-specific data. To this end four broad samples of data were selected, and served as training data sets for developed machine learning models, to develop a generic compression strength prediction model applicable to multiple lithologies. The suggested algorithms in this study are Back-Propagation Artificial Neural Networks, Artificial Neuro-Fuzzy Inference Systems, Support Vector Machines, Nearest Neighbour classifiers and Ensemble Bagged Trees. According to the findings of this study, Artificial Neuro-Fuzzy Inference Systems model performance was found to be marginally superior, while Back Propagation Artificial Neural Networks, Support Vector Machines and Ensemble Bagged Trees models were found to have good performance. Constant m_i seems to be an important training parameter when training predictive models centred on data from multiple lithologies. As a result, we can suggest that these models are powerful tools that allow for a reliable estimation of compressive strength, based on the performance indicators. The performance was found to be 70%–82% when the problem of compressive strength prediction was approached as a classification problem (that is successful prediction of class from very weak to very strong), and 80%–96% when solved as a function approximation problem.

1. Introduction

The mechanical properties of intact rock, compressive strength (σ_c) and Young's modulus (E), are essential parameters required for most geotechnical projects. However, it is not always possible for direct determination of these parameters. Although the uniaxial compressive strength σ_c, of intact rock is a significant parameter in rock engineering projects, it is challenging to obtain a representative value due to the need for high-quality core samples, and/or limited budget, especially during the preliminary design stage of a project. There are in addition, many uncertainties associated with the sampling of data when predicting σ_c. Consequently, established transformation models need to be constantly updated when applied to multiple lithologies, or sites. Hence, the numerous empirical equations that are proposed in the literature seem to be inadequate in estimating the σ_c reliably, in a large application range, since the material properties are affected by many factors namely (the crystal particle size, the degree of interlock and angularity, the fabric that infers the degree of anisotropy, the porosity), and several site-specific ground conditions. There is, therefore, a need to propose a data-centred intelligent system supported by a broad data set, to allow for a more representative compressive strength estimation. Laboratory based direct compressive strength measurements may be problematic, as obtaining fresh samples is not always feasible. Since determination of such a parameter in the laboratory is not always cost and time effective, a plethora of empirical equations, were proposed in the literature to estimate a reliable and representative value for a specific project. These models are referred to as 'transformation models', in the geotechnical literature [1]. Table 1 presents, the transformation models that were proposed in the studies that were used to compile the compiled database for the purposes of the current study and Table 2 presents a data map of the data base.

* Corresponding author.
E-mail address: M.Ferentinou@ljmu.ac.uk (M. Ferentinou).

https://doi.org/10.1016/j.probengmech.2022.103400
Received 5 June 2022; Received in revised form 10 November 2022; Accepted 12 December 2022
Available online 21 December 2022

Table 1
Transformation models for σc in the case studies from the literature used to compile ROCK/10/4025.

	Authors	Input	Nr of samples	Size of specimens	Lithology	Rock type	Model	Country
1	Abdi et al. [2]	n, γ, A_b, V_p	40	core samples no specific detail	sandstone, conglomerate, limestone, marl	Sedimentary	$\sigma_c = 15.36V_p + 29.62\gamma_d + 1.435A_b - 1.763n - 45.339$ $E = 1.95V_p + 8{:}07\gamma_d + 0.584A_b - 0.249n - 15.44$ MLP - ANN	Iran
2	Afolagboye et al. [3]	R_N, $I_{s(50)}$	50	40 mm x 40 mm x 60 mm	gneiss, granite, charnokite, quartzite	Metamorphic	MLP	Nigeria
3	Aggistalis et al. [4]	R_N, $I_{s(50)}$	93	Cylindrical D=62.2 mm	basalt, gabbro	Igneous	MLP analysis, $\sigma_c = -1.18\ R_N - 22.37$ WEATH + 25.9 (basalt) $\sigma_c = 8.63\ I_{s(50)} + 0.599R_N$	Greece
4	Akram and Bakar [5]	$I_{s(50)}$	8	Cylindrical	sandstone, limestone, siltstone, dolomite, marl	Sedimentary	$\sigma_c = 11.076I_s50$	Pakistan
5	Aliyu et al. [6]	ρ, V_p, V_s, Shore hardness, CAI	7	cylindrical, cuboidal	cryptocrystalline flint	Sedimentary	$\sigma_c = 17.6\ I_{s(50)} + 13.5$, $\sigma_c = 10.4\sigma_t + 18.2$, $\sigma_c = -47454.4 + 35905.6\rho - 6716.8\rho^{2.}$ $\sigma_c = 0.91V_p - 4500.6$	UK, France, Denmark
6	Jahed Armaghani et al. [7]	ρ_{dry}, V_p, Qtz, Kpr, Plg, Chl, Mica	45	core samples no specific detail	granite	Igneous	$\sigma_c = 69.505\rho_{dry} + 0.025V_p - 0.479$Qtz $- 1.439$Plg$- 158.796$; E= $-34.519\rho_{dry} + 0{:}019V_p + 1.217$Qtz $- 0.612$Plg $- 22.681$, ANFIS	Malaysia
7	Jahed Armaghani et al. [7]	R_L, V_p, $I_{s(50)}$	124	core samples no specific detail	granitic rocks	Igneous	NLMR ANN, ANFIS,	Malaysia
8	Armaghani et al. [8]	n, R_N, Vp, $I_{s(50)}$	71	core samples no specific detail	granite	Igneous	ANN - ICA $\sigma_c = -153.616n + 0.010V_P + 7.111$Is50 $+ 0.541R_n + 63.655$ $E = 16.573n + 0.011V_p + 4.560\ I_{s(50)} + 2.332R_n - 110.968$	Malaysia
9	Aydin and Basu [9]	n_e, nt, ρ, R_L, R_N	40	core samples	granite	Igneous	SLP, MLP $\sigma_c = 1.4459e^{0.0706RL,}$ $\sigma_c = 0.9165e^{0.0669RN}$ $E = 1.0405e^{0.058RL}$ $E = 0.7225e^{0.0548RN}$	Hong Kong
10	Azimian [10]	R_L, V_p, $I_{s(50)}$	30	Cylindrical D=54 mm	limestone	Sedimentary	SLP, MLP $\sigma_c = 0.009V_p^{1.105}$, $\sigma_c = 2.664\ I_{s(50)} - 35.22$, $\sigma_c = 1.530\ I_{s(50)} - 0.011(V_p) + 24.673$	Iran
11	Basu and Kamran [11]	$I_{s(50)}$	15	Cylindrical D=54 mm	schist	Metamorphic	$\sigma_c = 11.103\ I_{s(50)} + 37.659$	Pakistan
12	Basu et al. [12]	ρ, V_p, R_L, R_N	20	Cylindrical D=75 mm	granite	Igneous	$E = 0.1888\sigma_c + 30.234$	Brazil
13	Bell and Lindsay [13]	σ_{bt}, $I_{s(50)}$, Shore scleroscope hardness, R_L, E, ν	27	Cylindrical	sandstone	Sedimentary	Relationships between index, strength, hardness and elastic properties correlation	South Africa

(*continued on next page*)

The proposed models usually suffer from an inherent weakness, that they can be suitable only for the range of the conditions that exist in the calibration dataset and therefore could be dataset specific or site specific [96]. These models may not perform in a satisfactory way, if applied in another dataset originated from a different site. Most of these models are univariate or multivariate regression equations, but there are also several models based on soft computing methods. These models could not be characterized as generic since their performance beyond the calibration range is somehow questionable.

Dataset specific or site-specific models have their advantages when applied during the preliminary design stage, for a project, and adopted to estimate σ_c or E. They can serve as tools that allow, for initial estimations at preliminary design stage, but suffer from inherent limited and questionable application, outside the calibration range. Therefore, a generic transformation model, calibrated by a global rock database maybe necessary.

In this study a global dataset named ROCK/10/4025 is proposed which comprises 4025 data for ten rock parameters, namely: porosity (n), unit weight (γ), Schmidt hammer hardness (L-type), Block Punch Index (BPI), Brazilian tensile strength (σ_{bt}), Point Load strength index, (I_{s50}), P-Wave velocity (V_p), uniaxial compressive strength, (σ_c), Young's modulus, (E) and Hoek–Brown constant m_i.

Ching et al. [96], compiled a global rock database, labelled as ROCK/9/4069 which has a wider coverage than existing previously proposed transformation models. 14 out of 96 case studies, are common in both databases. These are original contributions that are presented

Table 1 (*continued*).

	Authors	Input	Nr of samples	Size of specimens	Lithology	Rock type	Model	Country
14	Bieniawski [14]	$I_{s(50)}$	12	Cylindrical 54mm, 42mm, 21.5 mm	sandstone, quartzite, norite	Mixed	σ_c = 24Is(50) 54 mm sample	South Africa
15	Bilgin et al. [15]	ρ, V_p, Sv, Edyn, v, R_L(10), R_L(3)	24	Cubic	coal	Sedimentary	$\sigma_c = 0.9346e^{0.062RL}$ Highest 3 Schmidt Hammer Values (SH3)	Turkey
16	Briševac et al. [16]	ρ, n, $I_{s(50)}$, RL, Vp	30	Cylindrical	mudstone and Whakestone carbonates	Sedimentary	σ_c = −106.2093 −0.04868 ρ + 11.5110 $I_{s(50)}$ + 0.052 Vp σ_c = −240.0109 + 1.5087 n + 11.5916 $I_{s(50)}$+ 0.0522 V_p, Bagging, random forests, regression forests variable importance V_p	Croatia
17	Bruno et al. [17]	R_L	95		carbonate rocks	Sedimentary	$\sigma_c = 1/(-0.022+1.41/R_L)$ $\sigma_c = \exp(-4.04+2.29\ \ln(R_L))$	Italy
18	Cargill and Shakoor [18]	$I_{s(50)}$, ρ , R_L, L.A Abrasion, Slake D Index	14	Cylindrical	sandstone, limestones, dolomite, syenite gneiss	Sedimentary	$\sigma_c = 23Is_{(54)}$ + 13 ln σ_c = 4.3 * 10^{-2}(R x ρ_d) + 1.2 sandstones ln σ_c = 1.8 * 10−2 (R x ρ_d) + 2.9 carbonates	USA
19	Çelik [19]	n, γ_d V_p, R_L	90	Cubic, 7, 9, 11 cm	carbonate rocks	Sedimentary	$\sigma_c = 2.683e^{0.549Vp,}$ $\sigma_c = 15.14\gamma_d$ + 2.88R_L−446.3, σ_c = −6.53L + 3.63V_p + 3.45R_L−50.68, LS-SVM	Turkey
20	Ceryan et al. [20]	n, ne, I_d, V_p, V_m	55	cubic	carbonate rocks	Sedimentary	ANN	Turkey
21	Cheshomi et al. [21]	SCI (single compression strength index) D particle diameter	10	cylindrical	sandstone	Sedimentary	σ_c = (−0.31.92) SCSI + 1.24D6.72)	Iran
22	Çobanoğlu and Çelik [22]	w, $I_{s(50)}$, V_p, R_L	75	cylindrical	sandstone, limestone, cement mortar	Sedimentary	σ_c = 4.14 $Is_{(50)}$ + 29.8 V_p + 0.54 R_L −116	Turkey
23	Dehghan et al. [23]	n, $I_{s(50)}$, V_p, R_L	30		travertine	Sedimentary	MLP, ANN, GRNN	Iran
24	Demirdag et al. [24]	R_L	59	Cubic	metamorphic, sedimentary, and igneous rock samples	Mixed	For low-porous sedimentary rocks: σ_c/R_L = 6.6333D−0.428 For high-porous sedimentary rocks: σ_c/R_L = 9.0233D−0.708 For metamorphic rocks: σ_c/R_L = 6.9364D−0.554 For igneous rocks: σ_c/R_L = 6.52 88D−0.434	Turkey
25	Diamantis et al. [25]	β, γ_{dry}, γ_{sat}, WA, n, V_p, V_s, $I_{s(50)}$	32	Cylindrical D = (50–55)	serpentine	Igneous	$\sigma_c = 10.61(I_{s(50)})$ + 6.87(10^{-2}) −339.48	Greece
26	Dinçer et al. [26]	n, γ_{dry}, R_N,SH, V_p, $I_{s(50)}$	24	Cubic (0.25*0.25*0.2)	basalts, tuffs	Igneous	σ_c = 2.75 R_L−36.83, E=0.47R_L−6.25	Turkey
27	Dinçer et al. [27]	n, γ_{dry}, R_N, V_p, $I_{s(50)}$	21	Cylindrical	caliche	Sedimentary	σ_c = −6.319 + 4.418 9 $10^{-3}V_p$ + 0.427γ E_av = 0.944 + 5.899 9 10^{-4} V_p −3.17 9 10^{-2}n	Turkey
28	Ersoy and Acar [28]	n, γ_{dry}, γ_{sat}, V_p, V_s	9	Cylindrical	granitic rocks	Igneous	σ_c = 116 −33.8V_p	Turkey
29	Fakir et al. [29]	n, γ_{dry}, V_p, σ_{bt}, $I_{s(50)}$,	29	Cylindrical	granitoid rocks	Igneous	σ_c = 15.939 $I_{s(50)}$ + 37.235 σ_c = 0.0673V_p −257.39 σ_c = 11.564σ_t −13.1 σ_c = 0.0142(R_L)2.3559	South Africa
30	Fener et al. [30]	$I_{s(50)}$, R_L, ISI	11	Cylindrical D = 38 mm	igneous, metamorphic	Mixed	σ_c = 9.08 $I_{s(50)}$ + 39:32, $\sigma_c = 4.24e^{0:059Rn}$, σ_c = 4.26ISI −204.33	Turkey

(*continued on next page*)

for the first time in the current paper. The proposed database was compared to other transformation models and was adopted to calibrate the bias and variability of proposed transformation models. The investigation of the named database suggested that most probably the proposed in literature transformation models were not dependent on the rock type, and that the models predicting E as opposed to the models that predict σ_c suffer from smaller transformation uncertainties. It was also found that the $I_{s(50)}$ is the most effective parameter into predicting the σ_c while S_h (shore scleroscope hardness) and V_p (shear wave velocity) are the most effective at estimating E.

Table 1 (*continued*).

	Authors	Input	Nr of samples	Size of specimens	Lithology	Rock type	Model	Country
31	Ferentinou and Fakir [31]	γ_{dry}, σ_t, $I_{s(50)}$	56	Cylindrical	granitoid, granite, sandstone, dolerite	Mixed	ANN	South Africa
32	Ghasemi et al. [32]	n, γ, R_L, V_p, SD	10	Cubic	mixed	Mixed	model tree approach, MLP algorithm	Turkey
33	Gomez-Heras et al. [33]	n, ρ,V_p, LHD	29	Cubic/ Cylindrical samples	carbonate porous, siliceous sandstone porous, siliceous crystalline, carbonate crystalline	Sedimentary	$\sigma_c = 10^{-3.945}LHD^{2.126}$, $\sigma_c = 10^{-5.376}LHD^{2.050}V_p^{0.450}$, $\sigma_c = 10^{-1.833} \cdot 1.396 \cdot e^{-3.138 \cdot n}$, $\sigma c = 10^{-2.604}LHD^{1.444}V_p^{0.172}e^{-2.807 \cdot n}$	Spain
34	Gonz'lez et al. [34]	n, ρ_{dry}, ρ sat, V_p(dry), V_p(sat), Vs(dry), Vs(sat)	13	Cylindrical D=30.9 mm	saturated limestone	Sedimentary	$\sigma_c = -185 + 55.8Vp - 24.87n$, $\sigma_c = 2.0137e0.794Vp - 0.401n$	Chile
35	Guney and Altindag [35]	$I_{s(50)}$, Bending Str, R_N, V_p	7	Cubic	sedimentary, igneous	Mixed	$\sigma_c = 4$ 124AM34 33	Turkey
36	Hebib et al. [36]	n, ρ, R_L	19	Cubic	sedimentary	Sedimentary	$\sigma_c = 32.904\ln(d/n) + 85.268$	Algeria
37	Heidari et al. [37]	$I_{s(50)}$	15	Cylindrical	gypsum	Sedimentary	$\sigma_c = 7.56\ I_{s(50)} + 23.68$	Iran
38	Heidari et al. [38]	R_L, BPI, $I_{s(50)}$, V_p	98	Cylindrical	grainstone, wackestone-mudstone, boundstone, gypsum, silty marl	Sedimentary	$\sigma_c = 43.898I_{S(50)} - 57.134$, $\sigma_c = 8.9217BPI - 1.2334$, $\sigma_c = 5.3466RN - 99.878$, Sugeno-type fuzzy algorithm	Iran
39	Ince and Fener [39]	n, ρ, V_p, $I_{s(50)}$, σ_t, Abrasion R, Slake durability	10	Cylindrical D=42 mm	pyroclastic	Sedimentary	% loss in $\sigma_c = 39.91 + 42.25Vp - 12.50\ I_{s(50)} + 95.29\rho_d + 2.76Id_4$	Turkey
40	İnce et al. [40]	n, $I_{s(50)}$, ρ_{dry}, ρ_{sat},	50	Cylindrical D=38 mm	pyroclastic	Sedimentary	(MRA), gene expression programming (GEP)	Turkey
41	Jalali et al. (2017)	$I_{s(50)}$, R_L, BPI, $I_{s(50)}$, V_p	106	Cylindrical	Grain-stone, wacke-stone, mudstone, bound-stone, gypsum, silty marl	Sedimentary	$\sigma_c = 1.277$ SHN + 2.86 BPI + 16.41($I_{S(50)}$) + 0.011Vp −82.436 ANN, ANFIS, FIS	Iran
42	Jamshidi et al. [41]	σb_t, R_L	22	Cylindrical D=54 mm	sandstone	Sedimentary	$\sigma_c = 70.417\ \ln(R_L) - 206.04$	Iran
43	Kahraman and Gunaydin [42]	$I_{s(50)}$	52	Cylindrical D =	andesite, granite, limestone, marble, travertine, serpentine, quartzite, metagabbro, anhydrite, gneiss, granodiorite	Mixed	$\sigma_c = 8.20\ (I_{s(50)}) + 36.43$ igneous, $\sigma c = 18.45\ (I_{s(50)}) - 13.63$ metamorphic, $\sigma_c = 29.77Is - 51.49$ sedimentary	Turkey
44	Kahraman [43]	$I_{s(50)}$, RL, Vp, Impact str	48	Cylindrical D=33 mm	dolomite, sandstone, limestone, marl, diabase, serpentine, haematite	Mixed	$\sigma_c = 4 * 10^{-10}\ I_{s(50)}^{5.87}$	Turkey
45	Kahraman et al. [44]	ρ, σ_{bt}, R_L, Impact pen rate, $I_{s(50)}$, Vp	27	Cylindrical	marl, clay, tuff, sandstone	Sedimentary	regression conical and spherical bit-tooth and rock properties	Turkey
46	Kahraman [45]	n, ρ_{dry}, ρ_{sat}, $I_{s(50)}$ dry, $I_{s(50)}$ sat	32		pyroclastic rocks	Igneous	$\sigma_c = 2.27e^{1.04IS(50)}$	Turkey
47	Kahraman et al. [46]	n, γ_{dry},γ_{sat}, n, BPI	28	Cylindrical D =38 mm	pyroclastic rocks	Igneous	$\sigma_c = 2.8\ BPI^{1.02}$	Turkey

(*continued on next page*)

Engineers were traditionally screening data from previous studies on similar or adjacent sites, to support geotechnical analysis on a new site. This practice was more common, in the case of limited ground investigation. It is not rare for geotechnical practitioners to inform their analysis based on experience and knowledge of the regional and site geology and engineering geological ground conditions, as well-informed assumptions. It is interesting to refer to [97] suggestion on characteristic value selection. *"Parameter values are gathered in statistical*

Table 1 (*continued*).

	Authors	Input	Nr of samples	Size of specimens	Lithology	Rock type	Model	Country
48	Kahraman et al. [47]	ρ, n, σb_t	9	Cylindrical D = 47.6 mm	granites	Igneous		Turkey
49	Kainthola et al. [48]	V_p, γ, σ_{bt}	199	Cylindrical D=54 mm	limestone, quartzite, slate, quartz	Mixed	$\sigma_c = 0.032(Vp) - 19.78$, limestone, $\sigma c = 0.077(Vp) - 241.2$, quartzite, $\sigma c = 0.037(Vp) - 36.12$ slate, $\sigma c = 0.021(Vp) - 4.911$ Quartz	India
50	Kamani and Ajalloeian [49]	$I_{s(50)}$	40	Cylindrical D=54 mm	carbonate aggregates	Sedimentary	Aggregate degradation tests AIV = 0.763Is(50)−10.67PLI + 54.248	Iran
51	Karakus et al. [50]	n, ρ, $I_{s(50)}$, R_L, V_p	10	Cylindrical	limestone	Sedimentary	Et = −12.766 −((4.06*10^{-7})exp(n) + (29.055 ln (Is(50) + 29.055 ln(Is(50))−(7.48 * 10^{-5}RL)−(6.44 * 10$^{-3}\sigma_c$nV $_p$) −(1.19 * 10^{-9} σ_c)	Turkey
52	Karakus [51]	n, V_p, $I_{s(50)}$, R_L	19	Cylindrical	granitic rocks	Igneous	genetic programming	Turkey
53	Karaman et al. [52]	σ_{bt}, $I_{s(50)}$, R_L	37	Cylindrical D=38 mm	pyroclastic rocks	Igneous	$\sigma_{bt} = 3.34$ Is(50) −3.4, $\sigma_{bt} = 0.72$ R −16.6, $\sigma_c = 33.3e3\ \sigma_t/R$, $\sigma_{bt} = 10.8\ \sigma_c/R_L$ −8.85	Turkey
54	Karaman and Kesimal [53]	R_L	47	Cylindrical	sedimentary, igneous, metamorphic	Mixed	$\sigma_c = 0.044R^{2.0043}$, $\sigma_c = 0.1383RT1^{1.743}$	Turkey
55	Kasim and Shakoor [54]	G_s, w, ρ, LA	22	Cylindrical	limestone, dolomite, sandstone, granite, marble	Mixed	$\sigma_c = 246.6 + (-2.8)(DE)$	Turkey
56	Khandelwal and Singh [55]	V_p, σ_{bt}, R_L, ρ	12	Cylindrical	coal, shale, sandstone	Mixed	$\sigma_c = 0.1333Vp - 227.19$, E=4.9718Vp−7151	India
57	Kılıç and Teymen [56]	n, σ_{bt}, $I_{s(50)}$, SHR, RL, V_p,	19	Cylindrical D=42 mm	diorite, quartzite, limestone, sandstone, tuff, marble	Mixed	$\sigma_c = 2.304Vp^{2.4315}$, $\sigma_c = 0.159SHR^{0.6269}$, $\sigma_c = 100\ln(Is50) + 13.9$, $\sigma_c = 0{,}0137R_L^{2.2721}$, $\sigma_c = 147.16e^{-0.0835n}$	Turkey
58	Korkanç and Solak [57]	γ_{dry}, γ_{sat}, n, SDI, V_p, $I_{s(50)}$, NPI	20	Cylindrical	tuffs	Igneous	$\sigma_c = 96.255xGMR - 6.3669$, $\sigma_c = -187.87GMR + 103.42$	Turkey
59	Mahdiabadi and Khanlari [58]	BPI, CPI, $I_{s(50)}$	80	Cylindrical	calcareous mudstones	Sedimentary	$\sigma_c = -6.479 + 3.425BPI + 0.639\ CPI + 7.889\ I_{s(50)}$, $E = 0.709 + 0.202\ BPI - 0.066\ CPI + 0.579\ I_{s(50)}$ ANN, ANFIS	Iran
60	Madhubabu et al. [59]	ρ, n, V_p,$I_{s(50)}$	19	Cubic (0.2*0.2*0.2)	carbonate	Sedimentary	E = −43.214−2.867n+ 1.384 Is(50)−127.411 ν + 18.251ρ − 0.0162 Vp, $\sigma_c = -11.813 - 2.572n + 23.665\ I_{s(50)}$+ 41.654 v 12.197$\rho$ − 0.001V_p	
61	Mahmoodzadeh et al. [60]	n, R_L, V_p,$I_{s(50)}$	170	Cylindrical	limestone, slate, quartzite, dolomite, granite, schist	Mixed	Soft computing models: long short-term memory (LSTM), deep neural networks (DNN), nearest neighbour (KNN), Gaussian process regression (GPR), support vector regression (SVR), decision tree (DT)	Iraq
62	Ludovico-Marques et al. [61]	n, ϵ	35	Cylindrical D=75 mm	granite, sandstone	Mixed	analytical model to describe σ_c with ϵ, and n	Portugal

(*continued on next page*)

populations of 'samples'. Different types of statistical population are distinguished depending on the way the populations are built up. In a local population, the sample test results, or the derived values are obtained from tests at the site of or very close to the geotechnical structure being designed. In the case of regional populations, the sample test results are obtained from tests on the same ground formation extending over a large area and collected, for example in a data bank. If a sufficiently large local population is available, it will be used primarily to select the characteristic value of the parameter considered; however, if no or only little local information is available, the selection of the characteristic value may be mainly based on results of regional sampling or other relevant experience". This challenge is called "site challenge" [98] or site recognition challenge [99].

Table 1 (*continued*).

	Authors	Input	Nr of samples	Size of specimens	Lithology	Rock type	Model	Country
63	Martins et al. [62]	n, ρ, V_p	55	Cylindrical	granite Begonha (1997) and Begonha and Sequeira Braga (2002)	Igneous	ANN, SVM	Portugal
64	Mehrabi Mazidi et al. [63]	ρ, q_u	23	Cylindrical	limestone	Sedimentary	$\sigma_c = 280\ (q_u)^2 - 152\ q_u + 64.25$	Iran
65	Mishra and Basu [64]	n, ρ, BPI, $I_{s(50)}$, R_L, V_p	20	Cylindrical	granite, schist, sandstone	Mixed	σ_c = EXP(0.0066) BPI+ 0.030 $I_{s(50)}$ + 0.01SRH 0.00019Vp + 2:807, σ_c = EXP 0.0029 BPI + 0.161 $I_{s(50)}$ + 0.013 SRH + 0.00018 V_p + 1.469), σ_c = EXP(0.024BPI + 0.047 $I_{s(50)}$ + 0.032 SRH + 0.00043 Vp + 0.872), σ_c = EXP (0.011 BPI + 0.065 $I_{s(50)}$ + 0.029 SRH + 0.000012 V_p + 2.157, FIS	India
66	Mohamad et al. [65]	ρ, σ_{bt}, $I_{s(50)}$, V_p	40	Cylindrical	shale, old alluvium, iron pan	Mixed	$\sigma_c = 0.032Vp + 44.227$, $\sigma_c = 12.291\ I_{s(50)} + 5.892$, $\sigma_c = 15.361\ \sigma_{bt} + 10.303$	Malaysia
67	Momeni et al. [66]	ρ, V_p, R_L, $I_{s(50)}$	66	Cylindrical	granite, limestone	Mixed	Particle swarm optimization-ANN	Malaysia
68	Naijib et al. (2015)	V_p, V_s, ρ	45	Cylindrical	limestone	Sedimentary	σ_c=11.05$E_S^{0.66}$; σ_c=12.8 (Ed/10) $^{1.32}\sigma_c = 3.67Vp^{2.14}$	Iran
69	Nefeslioglu [67]	w, γ, V_p	66	Cylindrical	claystone, mudstone	Sedimentary	$\sigma_c = 0.499138e^{1.575529(Vp)}$, $E = 0.006272e^{\ 2.979848(Vp)}$	Turkey
70	Ng et al. [68]	Gs, ρ_d, n, V_p, $I_{s(50)}$, R_L	145	Cylindrical D=54 mm	granite Grade III	Igneous	σ_c = a1(Is50) + a2 ($e^{0.0386RL}$) + a3($e^{0.0004Vp}$) + a4 ($\ln(n_c)$) +a5(Gs), σ_c = 5.01(Is50) + 5.52 $^{e0.0004Vp}$+ 3.53	China
71	Palchik and Hatzor [69]	E, c, d, n	12	Cylindrical D=52 mm	porous chalk	Sedimentary	$\sigma c = ae^{-bn}$	Israel
72	[70]	n, R_N	33	Cylindrical	carbonate rocks	Sedimentary	N = (σ_c) 0.2329 + 15.7244, N = (Et) 0.5155 + 17.4880, Et = (σ_c) 0.3752 + 4.479	Greece
73	Saldaña et al. [71]	n, γ, V_p, Vs	29	Cylindrical	travertine	Sedimentary	$\sigma_c = 115.32n^{-0.2116,}$ $\sigma_c = -13648 + 2407V_p + 5623\rho + 322n - 982Vp\ \rho - 10V_p - 115\rho d$	Chile
74	Salehin et al. [72]	ρ, $I_{s(50)}$, σ_{bt}, V_p, Vs, v,	39	Cylindrical	marl	Sedimentary	Adaptive boosting algorithm	Iran
75	Sarkar et al. [73]	V_p, $I_{s(50)}$, ρ , SDI	40	Cylindrical	limestone, slate, quartzite, quartz, mica schist	Mixed	ANN	Iran
76	Sengun et al. [74]	n, V_p, R_L, SHR	11	Cubic (11*11*11)	limestone, marble, onyx travertine	Sedimentary	–	Turkey
77	Shalabi et al. [75]	γ,v, R_L, SHR	58	Cylindrical D=54 mm	dolomite, dolomitic limestone, and shale rocks	Sedimentary	σ_c = 2.23HT + 7.23, σ_c = 3.201HR−46.59, σ_c = 97.221HA + 15.031, σ_c = 3.326Sh−79.76	USA
78	[76]	V_p, ISI, SDI	48	cylindrical	sandstone, limestone shale, basalt, schist, coal, phyllite	Mixed	$\sigma_c = 0.0642V_p + 117.99$	India
79	Sharo and Al-Shorman [77]	$I_{s(50)}$	85	Cylindrical	Sandstone, limestone shale, Basalt, Schist, coal, phyllite	Sedimentary	σ_c = 8.86+14.45 $I_{s(50)}$, chalky limestone, σ_c = 24.32+7.63 $I_{s(50)}$ basalt	Jordan

(*continued on next page*)

In the current study we want to develop a model that can reliably predict σ_c applicable in multiple lithologies. We believe that this is a first step towards exploring further the idea of densifying a limited local or site-specific population, with a larger more generic but similar population, to augment the available sample population and estimate reliably the compressional strength or the Young's modulus of intact rocks on a specific site. To this end we compiled a new generic database and applied intelligent bioinspired computational algorithms: Back-Propagation Artificial Neural Networks (BP-ANN), Artificial Neuro-Fuzzy Inference Systems (ANFIS), Support Vector Machines (SVM),

Table 1 (*continued*).

	Authors	Input	Nr of samples	Size of specimens	Lithology	Rock type	Model	Country
80	[78]	Is(50)	11	Cylindrical D = 54 mm	quartzite, sandstone, amphibolite, gabbro, khondalite, epidiorite, limestone, dolomite, shale, rock salt	Mixed	σ_c = (21–24) PLI harder rocks, σ_c = (14–16) PLI softer rocks	India
81	Sulukcu and Ulusay [79]	σ_{bt}, BPI, Is(50)	disc specimens were tested	Cylindrical D=50 mm	Mixed	Mixed	$\sigma_c = 5.1BPI - 3.3$, $\sigma_c = 15.31I_{s(50)}$	Turkey
82	Tandon and Gupta [80]	R_L, Is(50), V_p	60	Cylindrical	quartzites, granites, granitic gneiss, meta-basic (meta-amphibolite, meta-dolerite)	Metamorphic	$\sigma_c = 1.910SHR-10.30$, $\sigma_c = 3125PLI + 40.08$	India/ Himalaya
83	Teymen [81]	ρ, n,Is(50), V_p	28	Cubic (71*71*71)	tuffs	Igneous	$\sigma_c = 19.5I_{s(50)}$, $\sigma_c = 4.53Vp^{2.23}$	Turkey
84	Teymen and Mengüç [82]	ρ, Is(50), V_p, σ_{bt}, R_L, SRL	94	Cylindrical	Mixed	Mixed	Simple regression, multiple regression, ANN, ANFIS Genetic expression programming	Turkey
85	Tiryaki [83]	n, d	44	Cylindrical	Mixed	Mixed	$\sigma_c = 0.88 + \rho 2.24 + SHR0.22+CI0.8$, $E = 0.01 + \rho 5.72 + SHR0.29 + CI0.67$, regression trees, ANN	Turkey
86	Torabi M. Ataei M. Javanshir [84]	R_L	41	Cubic (0.2*0.2*0.2)	coal	Sedimentary	$\sigma_c = 0.0465R_L^2 - 0.1756R_L + 27.682$	Iran
87	Tumac and Hojjati [85]	RQD, R_L, IBR	21	Cylindrical	Mixed	Mixed	IBR = (0.763 RQDr) + (0.649 σcr) − 6.183	Turkey
88	Ulusay et al. [86]	γ, n, Is(50), v	15	Cylindrical	Mixed	Mixed	$\sigma_c = 19.5IS(50) + 12.7$	Turkey
89	Vasanelli et al. [87]	V_p, R_L	23	microcores and standard cubic samples	fine-grained calcarenite	Sedimentary	$\sigma_c = 0.0159x - 27$	Italy
90	Wang and Wan [88]	RL	20	Cylindrical	Mixed	Mixed	$\sigma_c = (6222/(88.15-R_L)) - 70.38$, simulated annealing-gene expression programming (SA-GEP)	China
91	Wen et al. [89]	ρ, Is(50), V_p, ν	40	Cylindrical	dolomitic limestone	Sedimentary	$\sigma_c = 0.034Vp - 86.36$, $E = 0.013Vp - 30.71$, $\sigma_c = 20.91\ Is - 4.79$	China
92	Yagiz [90]	σ_c, σ_{bt}, PSI, DPW	138	Cylindrical	granitoid	Igneous	ROP = 1.19 + 0.0247 Is 50, ROP = 1.413 + 0.0042 σ_c, ROP = 1.66 + 0.040 σ_{bt}	USA
93	Yarali and Soyer [91]	σ_{bt}, R_N, SRL, Is(50)	32	Cylindrical	sedimentary, igneous, metamorphic rock samples	Mixed	DRI = −0.2641UCS +87.049, DRI = $-5.872Is_{(50)}$ +86.297	Turkey
94	Yavuz et al. [92]	n, ρ, Vp, R_L	12	Cubic (0.7*0.7*0.7)	carbonate rocks	Sedimentary	I = βo+B1Io+β2n	Turkey
95	Yavuz et al. [93]	n, ρ, R_L, σ_{bt}, V_p, K	11	Cubic (0.5*0.5*0.5)	carbonate rocks	Sedimentary	$K = 0.560/\sigma_c^{0.604}$	Turkey

Nearest Neighbour classifiers (KNN) and Ensemble Bagged Trees Model (EBTM) to develop models that can learn from existing paradigms and subsequently make reliable predictions of rock compressive strength for a site of concern. Based on the developed model's performance indicators on testing data, we could present some confidence, that the applied machine learning (ML) algorithms, generalize successfully and can be used to predict successfully in future the σ_c for multiple rock types. As far as we are concerned there are no studies that propose a common regression equation or a soft computing algorithm that can predict the compressional strength of intact rock σ_c based on physical and index properties for multiple rock types. Rahman and Sarkar [100] concluded that a common regression equation cannot be used to predict the σ_c from for multiple rock types, based on their study for σ_c prediction from P-wave velocity values of 12 different rocks. This paper further investigates the performance of machine learning algorithms (BP-ANN, ANFIS, SVM, KNN, EBGM), that can predict σ_c and in future can be adopted to develop quasi-specific models [101].

Table 2
Data map for ROCK/10/4025 and available in the literature.

Authors	Nr of data groups	γ	n	V_p	R_L	Is (50)	σ_{bt}	BPI	σ_c	E
Abdi et al. [2]	40	x	x	x					x	x
Afolagboye et al. [3]	50				x	x			x	
Aggistalis et al. [4]	93				x	x			x	x
Akram and Bakar [5]	9					x			x	
Aliyu et al. [6]	7	x		x		x	x		x	x
Armaghani et al. [8]	124			x	x	x			x	
Jahed Armaghani et al. [7]	45	x		x					x	
Jahed Armaghani et al. [94]	71		x	x	x				x	x
Aydin and Basu [9]	40	x	x	x	x				x	x
Azmian (2017)	30			x	x				x	
Basu and Kamran [11]	15					x			x	
Basu et al. [12]	20	x		x	x				x	x
Bieniawski [14]	12					x			x	
Bilgin et al. [15]	23	x		x	x				x	
Bell and Lindsay [13]	27	x	x		x	x	x		x	
Briševac et al. [16]	30	x	x	x	x	x			x	
Bruno et al. [17]	97				x				x	
Cargill and Shakoor [18]	14	x			x	x			x	x
Çelik [19]	90	x	x	x	x				x	
Ceryan et al. (2012)	55	x	x	x					x	
Cheshomi et al. [21]	10	x	x						x	
Çobanoğlu and Çelik [22]	75			x	x	x			x	
Dehghan et al. [23]	30		x	x	x	x			x	
Demirdag et al. [24]	59		x	x					x	
Diamantis et al. [25]	32	x	x	x		x			x	
Dinçer et al. [26]	24	x			x				x	x
Dinçer et al. [27]	19	x	x	x	x	x			x	x
Ersoy and Acar [28]	9	x		x					x	
Fakir et al. [29]	29	x	x	x	x	x	x		x	
Fener et al. [30]	11				x	x			x	
Ferentinou and Fakir [31]	47	x				x	x		x	
Ghasemi et al. [32]	10	x	x	x	x				x	x
Gomez-Heras et al. [33]	29	x	x	x					x	
Gonz'lez et al. [34]	13	x	x	x					x	
Guney and Altindag [35]	7			x	x	x			x	
Hebib et al. [36]	19	x	x		x				x	
Heidari et al. [37]	30					x	x		x	
Heidari et al. [95]	15					x			x	
Heidari et al. [38]	98			x	x	x		x	x	
Ince and Fener [39]	10	x	x	x		x	x		x	
İnce et al. [40]	50	x	x			x			x	
Jamshidi et al. [41]	22				x		x		x	
Jalali et al. (2017)	106			x	x	x		x	x	
Kahraman and Gunaydin [42]	52					x			x	
Kahraman [43]	48			x	x	x			x	
Kahraman [45]	32	x	x			x			x	
Kahraman et al. [44]	22	x		x	x	x	x		x	x
Kahraman et al. [46]	28	x	x					x	x	
Kahraman et al. [47]	9	x	x				x		x	
Kainthola et al. [48]	199	x		x					x	
Kamani and Ajalloeian [49]	40					x			x	
Karakus [51]	19		x	x	x	x			x	
Karakus et al. [50]	9	x	x	x	x	x			x	
Karaman and Kesimal [53]	47				x				x	
Karaman et al. [52]	37				x	x	x		x	
Kasim and Shakoor [54]	22	x							x	
Khandelwal and Singh [55]	12	x		x			x		x	x
Kılıç and Teymen [56]	19		x	x	x	x	x		x	
Korkanç and Solak [57]	20	x	x	x		x			x	
Madhubabu et al. [59]	13	x	x	x		x			x	x
Mahdiabadi and Khanlari [58]	80					x		x	x	x
Mahmoodzadeh et al. [60]	170		x	x	x	x			x	
Ludovico-Marques et al. [61]	35		x						x	
Martins et al. [62]	55	x	x	x					x	x
Mehrabi Mazidi et al. [63]	23	x							x	
Mishra and Basu [64]	60	x	x	x	x	x		x	x	x
Mohamad et al. [65]	40	x		x		x	x		x	
Momeni et al. [66]	66	x		x	x	x			x	
Naijib et al. (2015)	45	x		x					x	x
Nefeslioglu [67]	66	x		x					x	

(*continued on next page*)

2. Database ROCK/10/4025

A generic global database ROCK/10/4025 is compiled from published case studies, in the literature, and includes 9 measured intact rock parameters, and a 10th parameter Hoek–Brown constant m_i, which was estimated for all the data, to indirectly include the lithology in the database. The current database is labelled, (material type)/(number of parameters of interest)/(number of data points) following generic

Table 2 (*continued*).

Authors	Nr of data groups	γ	n	V_p	R_L	Is (50)	σ_{bt}	BPI	σ_c	E
Ng et al. [68]	145	x	x	x	x	x			x	
Palchik and Hatzor [69]	12	x	x						x	x
Sachpazis [70]	33				x				x	x
Saldaña et al. [71]	29		x	x					x	
Salehin et al. [72]	39	x		x		x	x		x	x
Sarkar et al. [73]	40	x		x		x			x	
Sengun et al. [74]	11		x	x	x				x	
Shalabi et al. [75]	58	x			x				x	x
Sharma and Singh [76]	48			x					x	
Sharo and Al-Shorman [77]	85					x			x	
Singh et al. [78]	12					x			x	
Sulukcu and Ulusay [79]	23					x		x	x	x
Tandon and Gupta [80]	60			x	x	x			x	
Teymen and Mengüç [82]	93	x		x	x	x	x		x	
Teymen [81]	27	x	x	x		x			x	
Tiryaki [83]	44	x	x						x	x
Torabi M. Ataei M. Javanshir [84]	41				x				x	
Tumac and Hojjati [85]	21				x				x	
Ulusay et al. [86]	15	x	x			x			x	x
Vasanelli et al. [87]	23			x	x				x	
Wang and Wan [88]	20				x				x	
Wen et al. [89]	40	x		x		x			x	x
Yagiz [90]	138						x		x	
Yarali and Soyer [91]	32				x	x	x		x	
Yavuz et al. [92]	12	x	x	x	x				x	
Yavuz et al. [93]	11	x	x	x	x				x	

databases compiled for clays [102–105], sands [106], rocks [96], and rock mass [101].

The rock database comprises igneous, metamorphic, and sedimentary rock classes with 40 various lithologies. The geographical extend of the database covers 23 countries, including Algeria, Brazil, Chile, Croatia, Denmark, Egypt, France, Greece, India, Iran, Iraq, Israel, Italy, Jordan, Malaysia, Nigeria, Pakistan, Portugal, South Africa, Spain, Turkey, Ukraine, USA and expands, across five continents.

ROCK/10/4025 is populated with 54.8% sedimentary, 35.4% igneous, 9.2% metamorphic and 0.6% unclassified samples. The index parameters of intact rock samples are unit weight (γ; 4.41–35.41 kN/m^3), porosities (n; 0.06%–37.8%), Schmidt Hammer Hardness (R_L: 10.0–72.0), Point Load Index (I_{s50}: 0.05–30.55) MPa, P-wave velocity (V_p; 0.38–9.91 km/s), Brazilian tensile strength (σ_{bt}; 0.48–49.24 MPa), Block Punch index (BPI: 1.89–38.98 MPa), Hoek–Brown constant (m_i; 6–33), Young's modulus (E; 0.03–183.30 GPa), and (σ_c;0.27–560.31 MPa). The values of σ_c cover the full range of weak to (σ_c < 20 MPa), medium (20< σ_c<100 MPa), and strong rocks, (σ_c>100 MPa), based on classification by [107]. Most of the samples have σ_c ranging from 50 MPa–250 MPa, corresponding to strong and very strong. Only a few samples have σ_c <5 MPa and σ_c > 250 MPa, which are the very weak, extremely weak, and extremely strong samples. Most of the test samples are cylindrical of diameter in the range (30–75) mm, and some are cubical (253–753) mm^3.

There are 4025 records in the database. Each data set is stored in Excel worksheet that consists of a set of values measured for the same intact rock sample. The resulting database is not a genuine multivariate database in the way that all the 10 parameters are populated for all the rows in the worksheet, (i.e each row is not in full, completed with values). The percentage of completeness, of the data base defined as "(number of filled values)/[(number of parameters) × (number of rows)]. The percentage of completeness is 50% for ROCK/10/4025, while the percentage of completeness for ROCK/9/4069 is 34.2% [101]

Ten parameters commonly measured and identified in the literature to indirectly estimate σ_c and E, are included in the data base. The parameters of interest are γ, n, R_L, BPI, I_{S50}, σ_{bt}, σ_c, E, V_p, m_i. They can be categorized in the following groups:

1. Index properties: porosity (n), unit weight (γ), L-type Schmidt hammer hardness (R_L), and Block Bunch Index (BPI). The Schmidt rebound hammer is a practical non-destructive test method for the evaluation initially of rock strength and, subsequently, rock quality [108]. The surface hardness of rock is measured using a portable device with a non-destructive application called the Schmidt rebound hammer [109]. The ISRM has standardized the recommended test procedure for the Schmidt rebound hammer test. Researchers have indicated that the Schmidt rebound hammer could be a good indicator for determining the σ_c of rock [108,109]. There are two types of Schmidt hammer: L-type and N-type. In ROCK/10/4025, R_L data dominate (96% are R_L) because the L-type hammer is recommended for rocks as per [110]. There are studies where both R_L and R_N were measured Aydin and Basu [9], Basu et al. [12], Bilgin et al. [15], but in ROCK/10/4025, only $R_L$was included.
2. Strengths: Brazilian tensile strength (σ_{bt}), point load strength index (I_{s50}), and uniaxial compressive strength (σ_c) strengths. Index tests like the I_{s50}, σ_{bt}, have helped researchers develop models to predict the σ_c of rocks because index tests require relatively small data set samples. Grasso et al. [111] state that despite the limitations associated with index tests, when coupled with experienced judgement, they provide initial estimates of rock properties required at the feasibility and preliminary design stage. The point load index has long been used as the most suitable intermediary for the σ_c. It is evident from published research that the equations exhibit a wide range varying from linear to quadratic and power laws, but the issue is that there is no agreement between authors on a specific conversion factor [31]. The strength parameters of rock depend on specimen size, sample geometry, loading rate. I_{s50} is size dependent, it is customary to correct I_{s50} to a standard diameter of 50 mm [112], because point load testing can be conducted over a wide range of diameters. Prakoso [113] did not observe an effect of sample diameter on the coefficient of variation (COV) of rock strengths (σ_c, σ_{bt}, and $I_{s(50)}$), therefore sample size correction is not adopted for σ_c parameter or σ_{bt} in the compiled data base ROCK/10/4025.
3. Stiffness: Young modulus (E), the database contains average values of E as reported in the various case studies. ISRM suggested methods could be determined using tangent, secant, and average modulus. Ching et al. [96], report that the difference between (tangent modulus at 50% (E_{t50}) and average modulus

for the linear portion of the stress strain curve (E_{av}) is not significant compared to the transformation uncertainty. Thus, average values of Young's modulus are recorded in the database.

4. Dynamic property: P-wave velocity (V_p). Determination of dynamic properties of rock using ultrasonic pulse or sound tests, also known as P-wave velocity. It is regarded as a non-destructive test as it emits low-amplitude waves whose stress is below that of the yield stress of rock [64]. Ferentinou and Fakir [31] mention that P-wave velocity is an important parameter that should be measured for reliable σ_c prediction. The P-wave velocity of rock is affected by lithology, formation porosity, pore pressure, matrix, and temperature, rock mass weathering and alteration zones, bedding planes, and joint properties [114].

2.1. Hoek–Brown constant m_i estimation

The constant m_i is a fundamental parameter essential for the Hoek–Brown (HB) failure criterion developed for estimating the strength of rock mass properties. Hoek–Brown constant parameter m_i is estimated from a series of triaxial compression tests [115]. The constant m_i depends on the frictional characteristics of the component minerals in the intact rock, and it has a significant influence on rock strength, depends on, grain size, and cementation of rocks, [116]. Based on the general pattern of the correlations between m_i and rock types, Hoek and Brown [117] proposed guidelines for determining m_i values for different rock types that can be used for preliminary design when triaxial tests are not available. An updated version of these guidelines was proposed by Hoek [118] based on a more complete and detailed lithologic classification of rocks, and the range of m_i values depended on the accuracy of the geologic description of rock types. These guidelines were adopted in the current study, for the estimation of m_i. There are alternative methods suggested in the literature for the estimation of m_i in the absence of laboratory triaxial tests, which can be useful at the early stages of various design applications. An alternative way to estimate m_i values in the absence of triaxial tests is the R index, [117,119–121]. R index, is the ratio of σ_c to tensile strength σ_{bt}. Research by Read and Richards [122] investigated the relation between m_i and R values that were calculated from direct and Brazilian tensile tests which indicated that the use of direct tensile tests does not improve the ability to predict m_i values compared with Brazilian tests. R index was not adopted in the current study due to the non-completeness of the data base in terms of tensile and uniaxial and compressive strength. Another method proposed by Cai [120], Peng et al. [123], for the prediction of m_i is directly from the σ_c of the intact rock in which the m_i values depend on the ratio of crack-initiation stress obtained using acoustic emission techniques to the peak strength. Crack-initiation stress of rock samples data were not included in the studies that were used for the compilation of the ROCK/10/4025 database. As a result, the model was not used.

Shen and Karakus [124], proposed a simplified method for m_i values estimation directly from σ_c values for specific rock types in the absence of triaxial test data. They considered, correlations for five common rock types (sandstone, limestone, marble, granite, coal), and proposed a simplified method Eq. (1) together with the rock-specific relations that can estimate m_i (or normalized m_{in}) values using only σ_c and rock types. They used 112 groups of data for five common rock types in an existing database together with laboratory tests.

The proposed regression equation is

$$m_{in} = a\sigma_{ci}^{b} \tag{1}$$

where m_{in} is a normalized m_i/σ_c where a and b, are constants, and their values depend on rock types, Shen and Karakus (2014). The results revealed that there is a close agreement between estimated and experimental rock strength values. The reliability of the proposed method was evaluated and compared with existing methods ([125] guidelines and R index) that are commonly used for estimating m_i values when triaxial test data are not accessible. According to their results, the simplified method could be used reliably in the HB criterion to estimate intact rock strength with small discrepancies between estimated and experimental strengths. This simplified method was used in this study to develop machine learning (ML) models that can predict σ_c and was included in the database only for the common type of rocks (sandstone, limestone, marble, granite and coal).

The basic statistics for the ten parameters in the ROCK/9/4069 database are listed in Table 3. The statistics are the mean value, coefficient of variation (COV), minimum value (min), and maximum value (max). It should be noticed that the mean and COV are not for a specific site, but for the entire ROCK/10/4025 database that covers numerous sites. They should not be used for design, which requires statistics at the site level. The number of data points is further subdivided into the number of igneous, sedimentary, and metamorphic data points.

2.2. Basic statistics

The number of cases is shown in the third column in the format of "number of tests, (number of igneous + sedimentary and metamorphic). The COV and range are large in the global database and are comparable to the ROCK/9/4069 [96]. In ROCK/10/4025 the range of n, σ_{bt} and σ_c is higher, which is well illustrated in the series of correlations in Fig. 1. The COV value is broadly in agreement in the two databases. Tables 4,5,6 show the site level statistics for the igneous, sedimentary, and metamorphic rock intact rock properties. The site level statistics are broadly consistent with those in ROCK/9/4069. COV values and range for the specific sedimentary igneous and metamorphic groups of rocks is narrower than at global level, yet the site-specific range is larger in ROCK/10/4025 than in ROCK/9/4069. It was found that unit weight) generally exhibit low variability while porosity exhibit high variability across the three types of rocks considered for both data bases. There is high variability in strength properties across the three rock types, with upper COV values greater than 30%. Young's modulus is found to be highly variable, with high COV values across the three types of rock, with the lowest value at metamorphic rocks. In general, deformation and strength properties are more variable rock properties than index properties. Also, properties of sedimentary rocks tend to have higher COV values in most of the properties considered. Indicating that possibly sedimentary rocks exhibit greater variability than other types of rocks.

Fig. 1 presents the correlations between the index properties of the behaviour of intact rock properties as measured for both ROCK/9/2069 and ROCK/10/4025. The general trends are in agreement, and ROCK/10/4025 seems to be broader in terms of range. We could also suggest agreeing with [96] that there is not strong evidence indicating that the transformation relationship exhibited by the data points depend on the rock classes (i.e sedimentary, igneous or metamorphic).

3. Generic prediction models for the prediction of σ_c

3.1. Data preparation

Several, machine learning techniques, including artificial neural networks (ANN), adaptive neuro-fuzzy inference system (ANFIS), and classification learner algorithms (SVM, KNN, EBGM) were utilized, to predict the value of σ_c in this study. The use of these soft computational techniques enabled the comparison of different machine learning models in σ_c estimation from rock index properties, included in the ROCK/10/4025. The data samples were normalized with respect to maximum and minimum values to adapt to the interval [0, 1]. Normalization is conducted to ensure that all variables receive equal attention, soften the training procedure, and improve the accuracy of the results according to the linear mapping function [20,31]. The data were separated into training and testing data sets before the model was built. The training data set is used to create the model, and the testing dataset is used to validate the model.

Table 3
Summary statistics of the 10 parameters in the ROCK/10/4025 database at global level.

Nr	Parameter	Nr of data	Mean	COV	Min	Max
1	γ (kN/m^3)	1918 (1288)	23.85 (24.6)	0.16 (0.12)	4.41 (15.0)	35.41 (34.7)
2	n(%)	1371 (1371)	7.55 (11.0)	1.26 (1.02)	0.06 (0.01)	37.8 (55.0)
3	R_L	2157 (812)	42.78 (41.4)	0.27 (0.29)	10 (8.1)	72 (81.6)
4	σ_{bt} (MPa)	582 (854)	8.69 (8.2)	0.73 (0.61)	0.48 (0.07)	49.24 (34.4)
5	$I_{s(50)}$ (MPa)	2348 (1303)	4.05 (4.2)	0.80 (0.69)	0.05 (0.05)	30.55 (17.4)
6	BPI (MPa)	380	8.64	0.84	1.89	38.98
7	m_i	3994	17.25	0.54	6.00	32.00
8	σ_c (MPa)	3997 (3226)	78.81 (72.3)	0.73 (0.77)	0.27(0.68)	560.31 (379.0)
9	E (GPa)	953 (1495)	68.95 (26.5)	1.24 (0.99)	0.03(0.03)	183.30 (116.3)
10	V_p (km/s)	2366 (1858)	4.30 (4.0)	0.36 (0.39)	0.375 (0.44)	9.814 (8.0)

Table 4
Summary statistics of ROCK/10/4025 for igneous intact rocks.

Nr	Parameter	Nr of data	Mean	COV	Min	Max
1	γ (kN/m^3)	709 (17)	22.7 (24.55)	0.17 (0.03)	7.95 (16.98)	29.43 (30.07)
2	n(%)	645 (9)	8.63 (7.11)	1.33 (0.45)	0.01 (0.15)	41.80 (41.80)
3	R_L	841(14)	44.90 (45.76)	0.25 (0.12)	16.80 (16.80)	87.40 (65.76)
4	σ_{bt} (MPa)	312(10)	5.93 (11.59)	0.50 (0.23)	0.90 (1.9)	34.40 (16.75)
5	$I_{s(50)}$ (MPa)	946 (11)	4.43 (5.22)	0.76 (0.41)	0.05 (1.85)	12.53 (10.78)
6	BPI (MPa)	63	14.54	0.77	1.30	38.98
7	m_i	1417	27.82	0.20	8.00	33.00
8	σ_c (MPa)	1402 (33)	90.77 (112.42)	0.66 (0.33)	2.09 (17.63)	459.17 (246)
9	E (GPa)	380 (22)	35.30 (34.26)	1.05 (0.30)	0.19 (0.96)	183.3 (7.98)
10	V_p (km/s)	753 (22)	4.62 (4.53)	0.32 (0.13)	0.85 (2.88)	9.81 (7.55)

Table 5
Summary statistics of ROCK/10/4025 for sedimentary intact rocks.

Nr	Parameter	Nr of data	Mean	COV	Min	Max
1	γ (kN/m^3)	1028 (42)	23.88 (24.04)	0.16 (0.05)	4.46 (18.17)	26.56 (29.04)
2	n(%)	647 (43)	7.28 (13.51)	1.01 (0.33)	0.06 (0.43)	34.79 (38.40)
3	R_L	1123 (19)	40.67 (33.31)	0.29 (0.16)	11.80 (20.40)	70.70 (57.35)
4	σ_{bt} (MPa)	252 (21)	7.03 (7.03)	1.05 (0.28)	0.48 (1.25)	49.24 (14.85)
5	$I_{s(50)}$ (MPa)	1180 (34)	3.43 (3.47)	0.84 (0.30)	0.02 (0.17)	30.55 (9.69)
6	BPI (MPa)	301	7.17	0.57	1.75	20.79
7	m_i	2207	10.81	0.33	6.00	32.00
8	σ_c (MPa)	2201 (112)	68.46 (61.39)	0.78 (0.32)	0.27 (1.92)	560.31(159.13)
9	E (GPa)	555 (55)	15.91 (25.88)	1.28 (0.40)	0.03 (0.14)	99.9 (62.47)
10	V_p (km/s)	1339 (56)	4.11 (3.41)	0.39 (0.13)	0.38 (0.81)	7.84 (6.27)

Table 6
Summary statistics of ROCK/10/4025 for metamorphic intact rocks.

Nr	Parameter	Nr of data	Mean	COV	Min	Max
1	γ (kN/m^3)	181 (10)	26.38 (27.18)	0.14 (0.02)	23.64 (25.70)	30.31 (30.69)
2	n(%)	79 (7)	0.85 (2.25)	1.96 (0.54)	0.10 (0.03)	9.22 (6.67)
3	R_L	193 (3)	45.94 (49.78)	0.23 (0.09)	10.38 (39.9)	72.00 (51.60)
4	σ_{bt} (MPa)	18 (8)	8.76 (10.89)	0.59 (0.27)	2.41 (3.45)	19.70 (18.12)
5	$I_{s(50)}$ (MPa)	197 (12)	6.01 (4.04)	0.64 (0.30)	1.03 (1.03)	23.10 (6.92)
6	BPI (MPa)	16	13.08	0.46	4.61	27.05
7	m_i	370	15.73	0.48	9.00	33.00
8	σ_c (MPa)	369 (21)	97.83 (79.62)	0.60 (0.28)	8.00 (3)	288.20 (200.43)
9	E (GPa)	18 (7)	38.88 (54.59)	0.37 (0.21)	12.80 (1)	64.80 (102.33)
10	V_p (km/s)	259 (16)	4.58 (4.32)	0.32 (0.10)	1.47 (1.48)	7.94 (6.27)

3.2. Combination summary of generic data sets

Four combinations of parameters included in the ROCK/10/4025 were selected to develop and train predictive models. The selection of the four combinations of input parameters was based on four criteria. Predominantly, on the completeness of the database, incorporation of a physical property, strength, preferably through indexes resulting from non-destructive test methods, inclusion of lithology through m_i in the training. The authors selected one physical property as essential to be represented in the ML models, the intention was to investigate whether ML trained with a physical property as an input parameter, would have a better performance as opposed to the ML models that would not include a physical property as an input parameter. The second criterion was to comprise a non-destructive, easy, and cost-effective method like V_p and R_L which are present in all combinations. Point load index has long been regarded as the best intermediary for the σ_c, and therefore was comprised as a training parameter in all combinations. The constant m_i is incorporated as input parameter in all models as the purpose of the study is to propose a method that would allow to include qualitative properties such as lithology in the training of an ML model towards σ_c prediction for multiple lithologies.

1. The completeness of the data base: The intention of the authors was to compile a large representative database with typical ranges of rock properties that could be used as an approximation of rock property variability, when rock property data are not available or are very limited. The aim is to use ML algorithms and train with many representative educational examples which serve to build prior knowledge in the ML models. The authors performed a thorough check in the database to examine which combinations would allow for the larger data sets and concluded

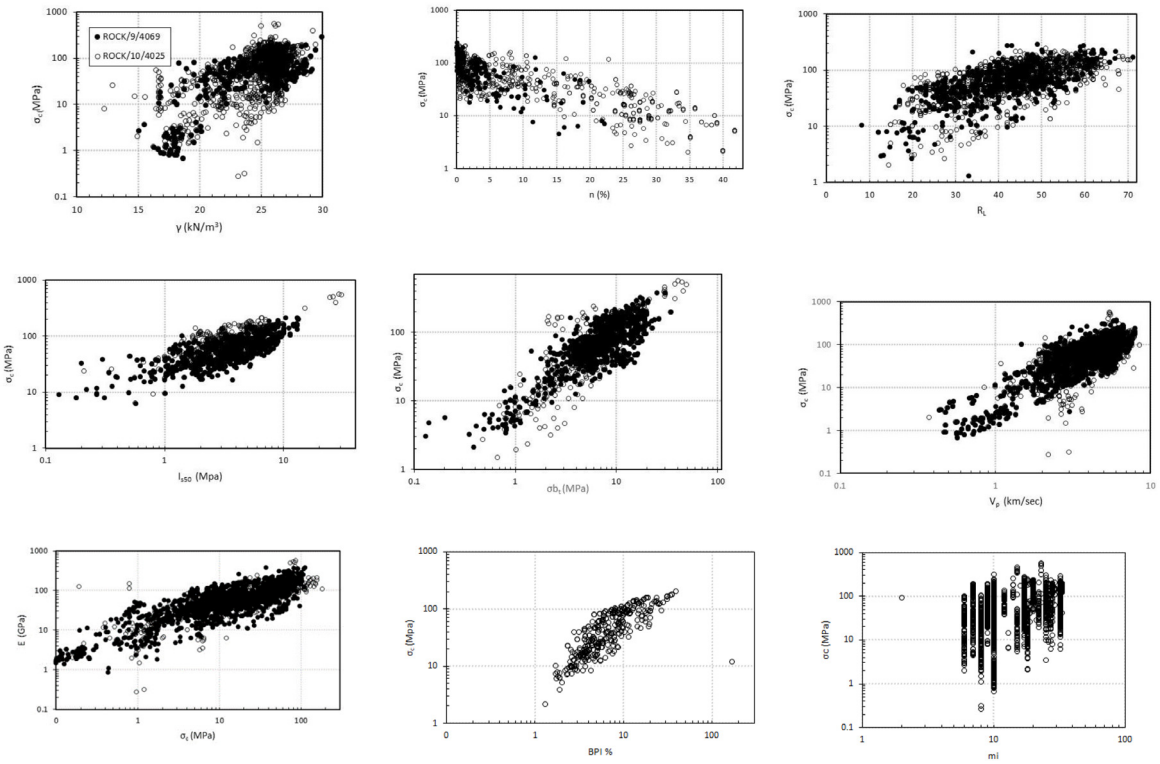

Fig. 1. Correlations behaviours among intact rocks in database ROCK/10/4025 and ROCK/9/4069.

Table 7
Selected parameter combinations for soft computing modelling.

Combination	No of data	Training set	Validation/ Test set	σ_c class range
R_L-$I_{s(50)}$- σ_{bt}	227 (118i +92s +17 m)	159	34	2–6
R_L-$I_{s(50)}$- σ_{bt} -m_i	227	159	34	2–6
R_L-$I_{s(50)}$- σ_{bt} -m_{iab}	97	68	14	2–6
V_p- $I_{s(50)}$-m_i	1454	1018	218	0–6
V_p- $I_{s(50)}$-m_i	1454 (564i + 752s + 123 m + mx)	1193	218	0–6
V_p- $I_{s(50)}$-m_{iab}	741	519	111	0–6
n-V_p- R_L - $I_{s(50)}$	575	402	85	1–5
n-V_p- R_L - $I_{s(50)}$-m_i	575 (282i + 251s + 41 m)	402	86	1–5
n-V_p- R_L - $I_{s(50)}$-m_{iab}	380	265	57	1–5
V_p- R_L - $I_{s(50)}$	1192	834	179	1–6
V_p- R_L - $I_{s(50)}$- m_i	1192 (511i+ 574s + 102m+ 5mx)	834	179	1–6
V_p- R_L - $I_{s(50)}$- m_{iab}	658	460	99	1–6

into the selected presented in Table 7. This strategy, of sample and parameter selection, allowed us to investigate the importance of sample data in comparison to the selected parameters in the generalization capacity of the proposed ML models.

2. Inclusion of the physical properties: Porosity n was selected representing the physical property of the intact rock sample. Unit weight γ has the lowest COV value of 0.12% and low variability and therefore was not selected. Porosity n, on the other hand is found on the upper bound of COV values greater than 60% which are considered with high variability, and therefore would allow for adequate representative educational examples included in ML models.
3. Inclusion of non-destructive test methods: The P-wave velocity has been successful as a non-destructive test for the prediction of mechanical properties of rocks easy and cost effective, the COV value is 0.30 which shows average variability. The relationship between σ_c and V_p has been investigated by many of researchers and a high correlation is identified. The Schmidt hammer is a hand-held portable device which is commonly used to assess the strength of rocks and concrete. It is a non-destructive cost-effective method used to assess the mechanical properties and therefore R_L index is included as an input parameter although the COV value is close to 30% which is considered low to medium.
4. Inclusion of strength index: Point load index I_{s50} has long been regarded as the best intermediary for the σ_c. It is relatively easy to conduct and economical, and thus widely applied both in the field and laboratory. It is evident from literature that the equations published exhibit a wide range, varying from linear to quadratic, and power laws. COV value is 0.80 which shows high variability and therefore adequate representation of a variability in the model. Ching et al. [96] report that Is_{50} is the most effective parameter into predicting the σ_c. The Brazilian tensile strength has been widely used as an indirect test to measure tensile strength (σ_{bt}). It has also been employed to produce estimates of σ_c strength as these two parameters are commonly required and determined in most geotechnical projects. As σ_{bt} can be easily determined from the Brazilian tensile strength, it is useful to find strong conversion factors between the two parameters. Brazilian tensile strength test is also associated to the constant m_i.

The selected combinations are the following four R_L-$Is_{(50)}$-σ_{bt}-mi, V_p-R_L-$Is_{(50)}$-mi, n-V_p-R_L-$Is_{(50)}$-m_i, V_p-$Is_{(50)}$-mi. Combinations of physical and index strength parameters of rock are developed from the data available in the database. Table 7 presents the different combinations and number of samples used for testing and validating.

The models were trained with and without the inclusion of m_i constant as an input parameter to verify the importance of m_i constant in the convergence and prediction capability of ML models. To further compare the effect of m_i estimation method on the generalization performance of ML models, four additional models were developed, m_i estimation was based on Shen and Karakus (2014), for the common type of rocks (sandstone, limestone, marble, granite and coal), in the data base, summarized in Table 7. The training data sets in this case were smaller.

R_L-$I_{s(50)}$-σ_{bt}-m_i: This combination includes a total of 227 rock samples. Many of the samples are from Turkey, and some are from South Africa. The combination contains 118 igneous, 92 sedimentary, and 17 metamorphic rock type samples.

V_p-R_L-$I_{s(50)}$-m_i: This combination includes 1192 intact rock samples. The set contains 511 igneous rocks, 102 metamorphic, 574 sedimentary and 5 unclassified. The samples are from Malaysia, Croatia, Iran, Turkey, India, and China.

n-V_p-R_L-$I_{s(50)}$-m_i: This combination includes 575 intact rock samples. The set contains 282 igneous rocks, 41 metamorphic and 251 sedimentary intact rock samples. The samples are from Malaysia, Croatia, Iran, Turkey, India, and China.

V_p-$I_{s(50)}$-m_i: This combination includes a total of 1454 intact rock samples. The set contains 564 igneous, 752 sedimentary, 123 metamorphic, and 15 non classified. This combination contains samples from France, Iran, Greece, Malaysia, Turkey, the United Kingdom, Denmark, India, Croatia, and China.

3.3. Machine learning methods

During the last two decades ML have been successfully applied to solve various problems in geotechnical engineering applications. In a survey of 444 papers by ISSMGE TC304/309 in 2021, both supervised and unsupervised algorithms are used with ANN, Support vector machines, nearest neighbour classifiers and Bayesian networks, being the most popular amongst them employed to solve problems such as site characterization, geomaterial behaviour modelling, foundations, retaining structures, slope stability, landslides, tunnels and underground openings, liquefaction assessment, etc, [126].

ML methods have also proven to be very effective in solving nonlinear and complicated problems in geotechnical engineering, such as the settlement of shallow foundations on cohesionless soils [127], thermo-hydromechanical behaviour of hydrate reservoirs [128], non-stationary and non-Gaussian geotechnical properties [129], soil constitutive modelling [130] suction distribution in shallow soil layers [131] and soils' air entry value [132].

Some of the advantages of ML methods that guided the authors to develop a generic compressional strength prediction model for multiple lithologies is that they conserve the complexity of the systems they model because they have complex structures themselves, they recognize different sets of data within a whole data set, they do not require pre-existing knowledge or experience, they do not require a statistical pre-existing model in order to train data and they give reasonable results even when data are inaccurate and incomplete which is typical for geotechnical data. Phoon et al. [99] presented a successful mnemonic, MUSIC-X (Multivariate, Uncertain and Unique, Sparse, Incomplete, and potentially Corrupted with "X" denoting the spatial/temporal dimension) to highlight seven common ugly attributes in real site geotechnical data.

In this study compressional strength σ_c prediction using a generic database was approached as a function approximation problem and as a classification problem. For the function approximation problem, Back-Propagation Artificial Neural Networks (BP-ANN) and Artificial Neuro-Fuzzy Inference Systems (ANFIS), were employed whereas to solve the classification problem binary and multiclass classification algorithms, such as decision trees, nearest neighbour classifiers (KNN), support vector machines, (SVM), ensemble classifiers (EBGM) and neural networks classifiers algorithms were selected based on their accuracy.

Artificial Neural Networks (ANNs) are supervised machine learning techniques that evaluate the relationships between a known set of observations (thus the training data) to predict unknown data sets. ANN consists of multiple interconnected processors, called neurons which are inspired by biological neurons. The neurons are logically arranged in two or more layers and interact with each other via multiple weighted connections. The links that connect neurons carry a numerical weight that is associated with the neuron. These weights express the strength, thus the importance of input neurons. A neural network learns through the constant/regular adjustment of these weights. Training of a network is modifying the network input/output behaviour to align with the external stimulus. An ANN learns through the process of reducing and minimizing the difference between the actual output and the desired output through adjusting the weights/synapses of the network. A widely used type of activation function is the continuous one designed to respond to the magnitude of the received excitation. The sigmoid function is one of the continuous transfer functions typically used in modelling the activity of neurons. (Eq. (2))

$$f(x) = \frac{1}{e^{-ax}} \tag{2}$$

where f is the amount of activation, x is the net excitation and "a" is the slope function. These functions are also known as squashing functions since their output is limited in a finite range of values.

A simple BP-ANN consists of three layers: the input layer, the hidden layer, and the output layer, as illustrated in Fig. 2. BP-ANN is usually layered, with each layer fully connected to the layers below and above. The first layer is the input layer, the only layer in the network that can receive external input. The second layer is the hidden layer in which the processing units are interconnected to layers below and above. The third layer is the output layer. Each unit of the second hidden layer is interconnected to the units in the output layer. Units are not connected to other units in the same layer. Each interconnection has associative connection strength, depicted as weight in Fig. 2. When an output array is presented the error vector, which represents the difference between the desired value and the actual output, is calculated. Using a "wide range" and "representative data sets", allows for a better representation of the input space and therefore, better generalization capability. A network is said to generalize when it appropriately classifies vectors not included in the training set. The generalization ability is measured by the accuracy of these classifications. It is important for the number of training input vectors to be greater than the number of degrees of freedom (the number of variable weights) of the network. The basic mathematical concepts of the backpropagation algorithm are found in literature [133].

The main hyperparameters of a multi-layer perceptron (BP-ANN) are the learning rule, the number of neurons in the hidden layer, and the transfer function. An optimum ANN architecture selection is an essential step in building a model that is ideal for prediction purposes.

For this study, a trial and error, approach is adopted to select the number of neurons in the hidden layer starting from a small number and increasing gradually according to the received accuracy index for the 12 developed ANNs. The size of the BP-ANN model was verified following Baum and Haussler [134], who suggest the bounds on appropriate sample vs. network size, for a feedforward network of linear threshold, with N nodes and W weights. The tangent sigmoid nonlinear transfer function is used between the input and hidden layer and the *purelin* transfer function between the hidden and output layer.

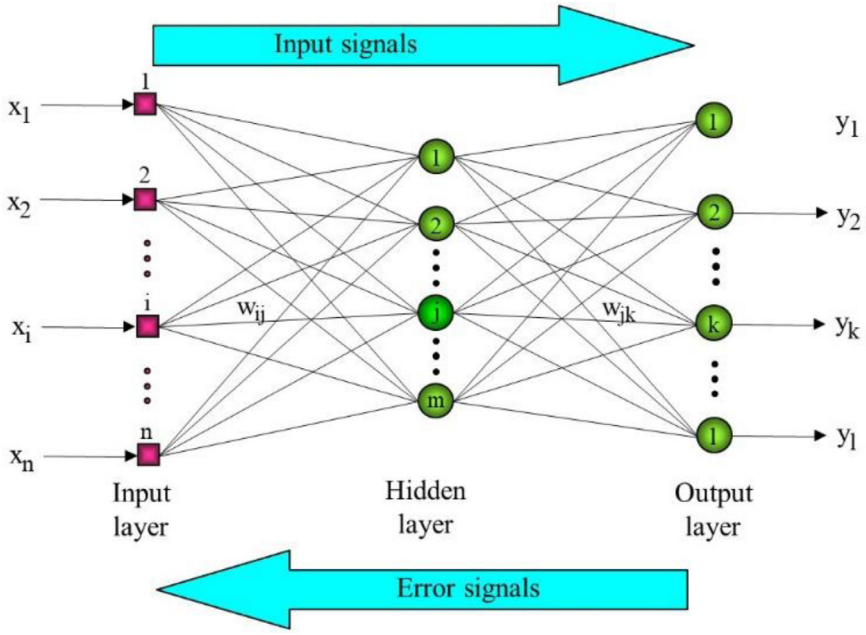

Fig. 2. Typical BP-ANN architecture of a generalized multilayer perceptron. Simulation is applied to the inputs of the first layer, and signals propagate through the middle (hidden) layer to the output layer. Each link between the neurons has a unique weighting value.

3.4. Adaptive Neuro-Fuzzy Inference System (ANFIS)

The artificial neural networks learning ability is combined with the fuzzy logic's decision-making mechanism in ANFIS. It is a hybrid soft computing technique that combines the best of ANN and fuzzy inference system (FIS). Fuzzy inference is the process of mapping a given set of inputs to an output using fuzzy logic [135]. ANFIS eliminates the primary problem in FIS defining membership function (MF) parameters and obtaining the fuzzy if–then rules through the effective use of ANN learning capability for automated fuzzy rule generation and parameter optimization, Maiti and Tiwari [136]. This method's innovation is derived from the fact that it does not require expert knowledge for the assignment of parameters of the FIS but utilizes neural network algorithms to adjust parameters. This method uses representative sets of input and output to generate a FIS whose parameters are adjusted using backpropagation or a combination with the least-squares method. A FIS develops "if…then…" fuzzy rules and determines MFs to map the input and output data of the system. ANFIS automatically transforms a training set into a set of fuzzy rules, thus reducing the dependence on expert knowledge for building intelligent models. The linguistic variables are subdivided into a specified number of sets during the initialization step of ANFIS.

The structure of an ANFIS model is similar to that of a multilayer perceptron neural network, Fig. 3. In general, a neuro-fuzzy system has five layers, thus with one input, one output, and hidden layers representing the membership functions and fuzzy rules. The Takagi and Sugeno [137] type FIS is employed in this research as it uses a systematic proposition to generate fuzzy rules from the given input/output sets.

Details on the mathematical formulation of ANFIS can be found in [138]. A trial and error approach is taken to select whether grid partition or subtractive clustering is used to generate the FIS structure. In the current study Subtractive clustering first introduced by Chiu [139], was applied. To build an ANFIS model for this combination of parameters, a trial and error approach was applied to select the best model features.

To obtain an optimum Adaptive Neuro-Fuzzy Inference System, 34 models are developed. All the input membership functions are of Gaussian type, and the subtractive clustering method was applied for the proposition of fuzzy rules and to generate the desired training architecture. In the FIS, five descriptive linguistic labels (i.e., very low, low, moderate, high, and very high) were assigned to the input membership functions. The hybrid rule was employed in the learning procedure, and training was conducted until the error measure of the output was tolerant. Testing and checking data pairs were used to characterize the model. The range of influence (ROI) and squash factor (SF) is varied until the optimum Artificial Neuro-Fuzzy Inference System is obtained to develop the models. The ROI and SF influence the number of fuzzy rules and membership functions, thus affect the ANFIS network structure. The structure details of the models is presented in Table 8.

A total number of five rules show the best performance in predicting σ_c. An example of one of the "IF–THEN" rules is illustrated in linguistic terms is as follows:

1. "IF (σ_c is low) and ($I_{s(50)}$ is low) and (σ_{bt} is low) and (m_i is low) THEN (is cluster4) (1)".
2. "IF (R_L is very_high) and ($I_{s(50)}$ is very_high) and (σ_{bt} is very_high) and (m_i is very_high) THEN (σ_c is cluster5) (1)".

3.5. Classification models

Classification is a supervised machine learning method that allows for the prediction of discrete responses or cluster generation according to common features. The models are trained to classify data into categories. For the specific problem, the data outputs are first turned into categorical data. The σ_c value is translated into seven classes from 0 to 6 as per Hoek and Brown (2007)). The values of σ_c in the Rock-96/10/4025 range from 0.27 MPa<UCS<560.31 MPa, from extremely weak to extremely strong rock, Table 9, according to field estimates of uniaxial compressive strength [117]

As can per frequency distribution, 25% of the samples have σ_c ranging from 50 MPa–250 MPa, corresponding to strong and very strong. A 20% of samples have σ_c in the range 5 MPa–50 MPa, which are weak and medium-strong rocks. whereas 10% samples have σ_c of 0 MPa–5 MPa and 8% σ_c of greater than 250 MPa.

The ML models simulate multiclass classification problems since there are seven desired output categories for all datasets to be classified. Determining which classification algorithm to use is often very

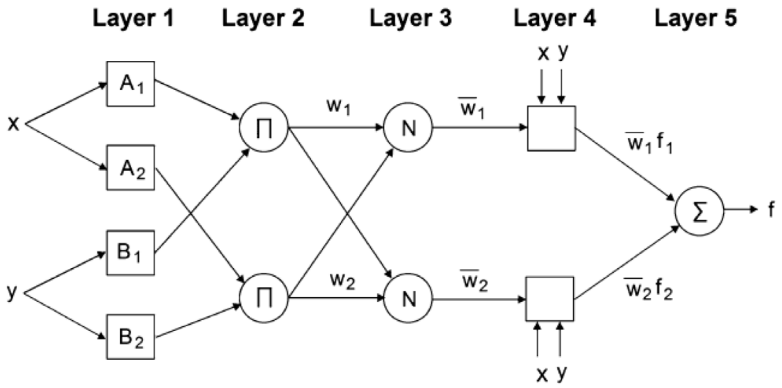

Fig. 3. A typical structure of first order Takagi and Sugeno model.

Table 8
Structure details for the ANFIS models developed.

	R_L-$I_{s(50)}$-σ_{bt}-m_i	V_p-R_L-$I_{s(50)}$$m_i$	n-V_p-R_L- $Is_{(50)}$-m_i	V_p-$I_{s(50)}$-m_i
Number of nodes	57	47	56	22
Number of linear parameters	25	20	4	8
Number of nonlinear parameters	40	32	40	12
Total number of parameters	65	52	44	20
Number of training data pairs	182	954	460	1193
Number of checking data pairs	23	119	57	149
Number of testing data pairs	22	119	57	149
Number of fuzzy rules	5	4	4	2
Input MF:	Gaussian	Gaussian	Gaussian	Gaussian
Output MF	Linear	Linear	Linear	Linear

Table 9
Field estimates of uniaxial compressive strength (Brown, 1981).

UCS range (MPa)	Class	Grade
0.25–1	0	Extremely weak
1–5	1	Very Weak
5–25	2	Weak
25–50	3	Medium Strong
50–100	4	Strong
100–250	5	Very Strong
250+	6	Extremely Strong

challenging. These ML algorithms were used in other classification problems and have proved to achieve high performance. A typical work-flow followed when building a classification model can include Selection of classification method, training the classifier, measure the accuracy of the classifier, simplify the model. The used classifiers that we experimented with are: Naïve Bayes Classifiers, Support Vector Machines, Nearest Neighbour Classifiers, Ensemble Classifiers which were compared to single Neural Network Classifiers.

Naïve Bayes Classifiers (NB) are used for classification where the observations are differentiated using specified features. This model is a probabilistic classifier based on strong independence assumptions between features. In this study kernel naïve bayes variant is applied.

Support Vector Machine (SVM), Cortes and Vapnik [140] is a type of supervised learning ML technique. The supervised learning algorithm selects the hyper-plane or the decision boundary defined by the solution vector w to determine the maximum margins between the training data samples and the test data. The variants of SVM used in the current research are medium Gaussian SVM, Linear SVM, Cubic SVM, Quadratic SVM.

Nearest Neighbour classifiers (KNN), the K-Nearest neighbour model is successfully used in previous studies in solving non-linear problems. KNN is used to assign a class label using the smallest Euclidean distance between the target point and the training point in the feature space. The variants of KNN used in the current training are Weighted KNN, Medium KNN, Cubic KNN, and Fine KNN.

Ensemble classifiers is a ML approach in which numerous models (called "weak learners") are trained to tackle the same problem and then combined to achieve superior results [141]. Usually, these base models do not perform well individually, either because they contain too much bias or too much variation to be robust. Ensemble techniques are used to try to reduce then bias and variance of weak learners by stacking many of them to generate a strong learner. Base models and a meta-learner (or a second-stage model) that uses base-model predictions are used to design a stacking ensemble model. Variants of ensemble models used in the classification process are Bagged Trees and subspace KNN.

4. Statistic performance evaluation – Model performance indicators

In order to check the overall performance of the developed predictive ML models, some statistical performance indices were calculated for each model according to the Eqs. (3)–(4) given below. Root mean square error (RMSE) which evaluates the residual between desired and output data, and R^2 which evaluates the linear relation between desired and output data.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(d_t - y_t)^2} \tag{3}$$

$$R^2 = \left[\frac{\sum_{t=1}^{n}\left(d_t - d_{mean}\right)\left(y_t - y_{mean}\right)}{\sqrt{\sum_{t=1}^{n}\left(d_t - d_{mean}\right)^2}\sqrt{\sum_{t=1}^{n}(y_t - y_{mean})^2}}\right]^2 \tag{4}$$

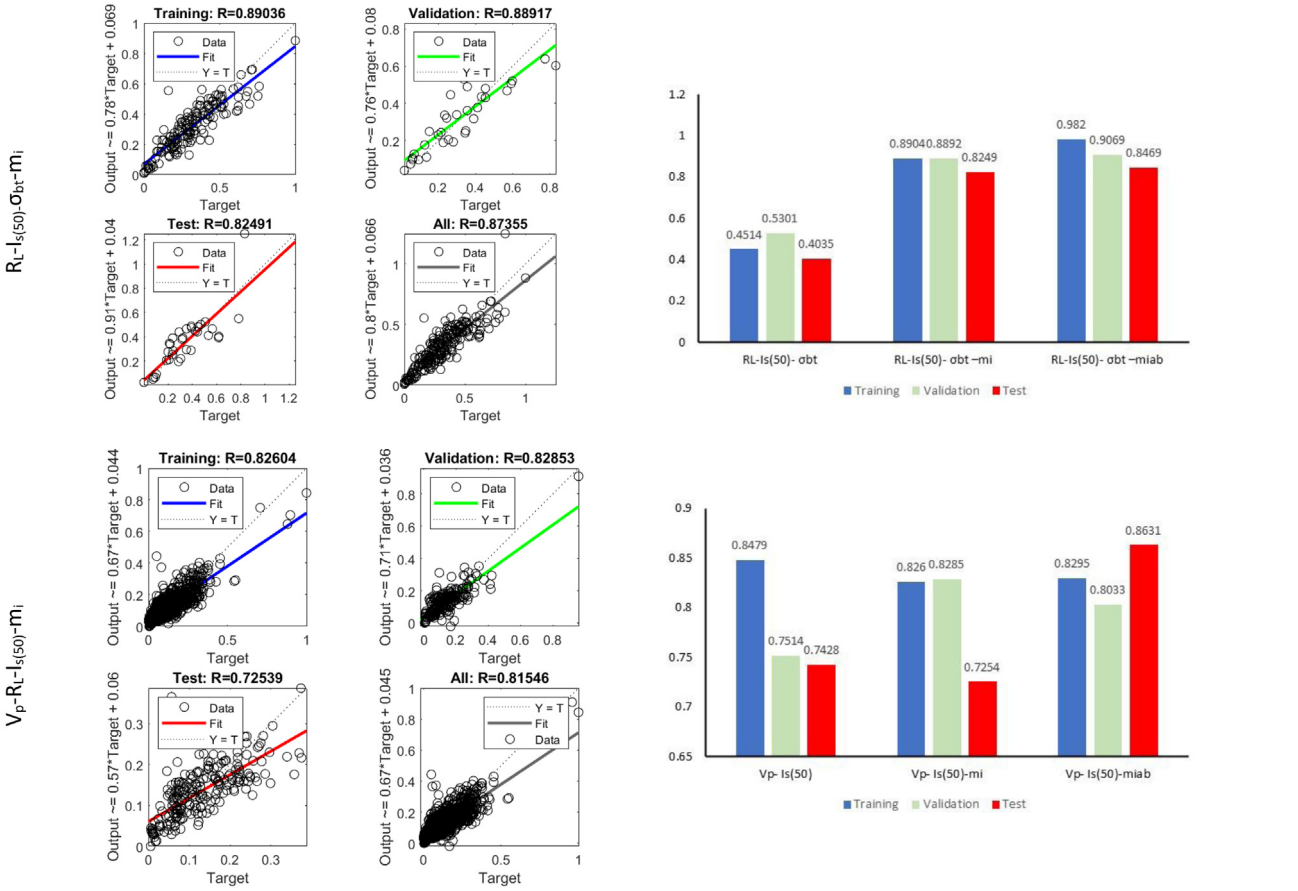

Fig. 4. Left: Performance of the ANN algorithm, parity plots showing the ANN predictions against the corresponding measured values for training, validation, target and all data. Right: Performance of all models measured through R^2 index performance indicators in relation to m_i impact (without m_i), with m_i estimated through [118] guidelines, mi $_{ab}$ estimated through (Shen and Karakus, 2014).

where n is the number of training or testing samples, d_t and y_t are the measured and predicted values, respectively.

The performance of all the 12 ANN models is presented through Fig. 4a and 4b and in Table 8.

In classification algorithms accuracy is the statistical indicator used to quantify the performance of the four compression strength class classifiers. It is the total number of classes correctly predicted. The mathematical expression of accuracy is given in Eq. (5):

$$Accuracy = \frac{TP + TN}{observations} \tag{5}$$

where TP is the true positive and TN is the true negative.

5. Results

5.1. ANN model validation

5.1.1. R_L-$I_{s(50)}$-σ_{bt}-m_i

The network topology adopted in this model is 4:5:1, corresponding to four predictors input neurons, five neurons in the hidden layer, and one output layer node. The training parameter used was 1000 epochs. The MSE for training is 0.0037 at epoch 52. The regression plots Fig. 4a show that the model is a good fit as for all data the R values are above 0.87, meaning the ANN model accurately represents more than 87% of the data.

5.1.2. V_p-R_L-$I_{s(50)}$-m_i

The network topology adopted in this model is 4:4:1, corresponding to four predictors input neurons, four neurons in the hidden layer, and one output layer node. The training process stopped when the minimum gradient was reached (8.22e−08) at epoch 74. The regression plot shows that the model is a good fit as for all data the R values are 0.81, meaning the ANN model accurately represents more than 81% of the data, as shown in Fig. 4a.

5.1.3. n-V_p-R_L-$I_{s(50)}$-m_i

The network topology adopted in this model is 5:4:1, corresponding to four predictors input neurons, four neurons in the hidden layer, and one output layer node. The training process stopped when the maximum epoch number was reached at 10,000. The regression plot shows that the model is a good fit as all the R values are above 0.88, meaning the ANN model accurately represents more than 88% of the data, Fig. 4a.

5.1.4. V_p- $I_{s(50)}$-m_i

For this combination of parameters, the network topology adopted in this model is 3:5:1. The training process stopped when the minimum gradient was reached (9.80e−08) at epoch 338. The MSE for training is 0.0027 at epoch 29. The regression plots Fig. 4a shows that the model is a good fit as for all data the R values are 0.90, meaning the ANN model accurately represents more than 90% of the data.

In order to investigate the importance of m_i constant as a training parameter, and therefore verify if models that included m_i constant as input parameter outperform the models that do not include m_i as an input parameter 8 additional models were developed.

The models that were trained with the inclusion of m_i constant in the input parameters seem to have systematically higher performance than the models that did not include m_i parameter as an input parameter as summarized in Fig. 4b (R^2), and Table 10, for all the four

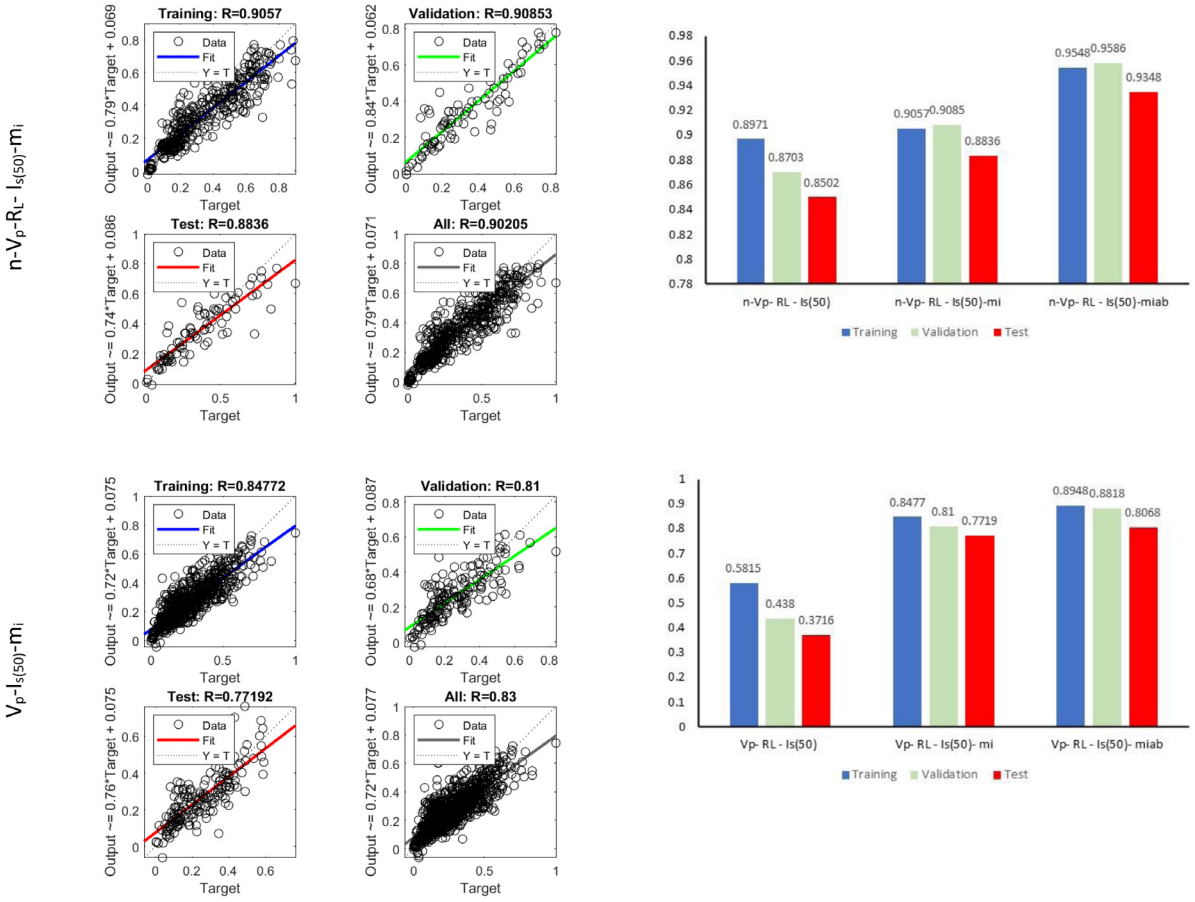

Fig. 4. (*continued*).

selected combinations. This applies to training validation and test data, with the exception of V_p-$I_{s(50)}$-m_i combination, where no significant improvement is recorded. To further compare the effect of m_i estimation method on the generalization, m_{inab} was estimated according to simplified method proposed by Shen and Karakus (2014), for the common type of rocks (sandstone, limestone, marble, granite and coal), in the data sets, and therefore four more models were developed. These models presented a clearly higher performance four all the selected combinations in above 85% for broadly all the models.

R_L-$I_{s(50)-}\sigma_{bt}$-m_i predictive model and n-V_p-R_L- $I_{s(50)-}m_i$ based on the performance indicators seem to have higher performance, than V_pR_L-$I_{s(50)}$-m_i.

5.1.5. Ranking of the training parameters

A powerful feature of neural networks is their ability to perform parametric analysis through manipulation of the connection weights. The developed ANN model can also provide the parameter relative importance by partitioning the hidden output neuron connection weights into components connected with each input neuron [142–144]. This parametric analysis evaluates which parameters are more important in the prediction of σ_c. Fig. 5 shows the parametric analysis and relative importance of the input variables for the neural network models. The results indicate that the most important parameters that affect the compression strength prediction are the P-wave velocity (V_p), and Schmidt hammer rebound index, (R_L). According to the same analysis the point load index $I_{s(50)}$, and constant mi are equally important.

The MRMR algorithm [145] which finds an optimal set of features that is mutually and maximally dissimilar and can represent the response variable effectively was also applied to assess the importance of the parameters in the classification problem. The algorithm minimizes the redundancy of a feature set and maximizes the relevance of a feature set to the response variable. MRMR algorithm was found to broadly produce similar ranking with partitioning of connection weights method as illustrated in Fig. 5.

5.2. ANFIS model validation

Fig. 5 presents a complete constructed fuzzy model for R_L-$I_{s(50)}$-σ_{bt}-m_i combination. The yellow plots show the antecedent part, which are the input membership functions representing the IF-part of the rule. The consequent part is shown by the blue plots and defines the membership functions are referenced as the THEN part of the rule. All the plots are filled since these correspond to the characterization of 'none' for the variable for that rule. The aggregate weighted decision for the FIS is plotted on the final plot of the last column. The bold red line on this plot represents the defuzzified output.

Training is a process of adjusting the pre-set membership functions and fuzzy rules to model the training dataset. The epoch number is set to 1000, and the error tolerance is set to 0.01 to allow the network to train until the error tolerance of the training is to acceptable limits and there is no overfitting. The graph in Fig. 6 presents the plot of a trained data set, for n-V_p-R_L-$I_{s(50)}$-m_i combination the blue circles represent the training data while the red asterisks the output data after training. The plot shows that the training pattern is comparable to the desired output pattern.

The testing data set is used to validate FIS structure. After the model is trained, a set of data points which were not used for training are used for data validation. The trained model is used for simulation of the testing datasets. Table 11 presents the testing error which is an indication of the ability of the model to predict σ_c. According to the obtained results R_L-$I_{s(50)}$-σ_{bt}-m_i predictive model seems to yield the higher performance of 0.089502, followed by n-V_p-R_L-$I_{s(50)}$-m_i with 0.076215.

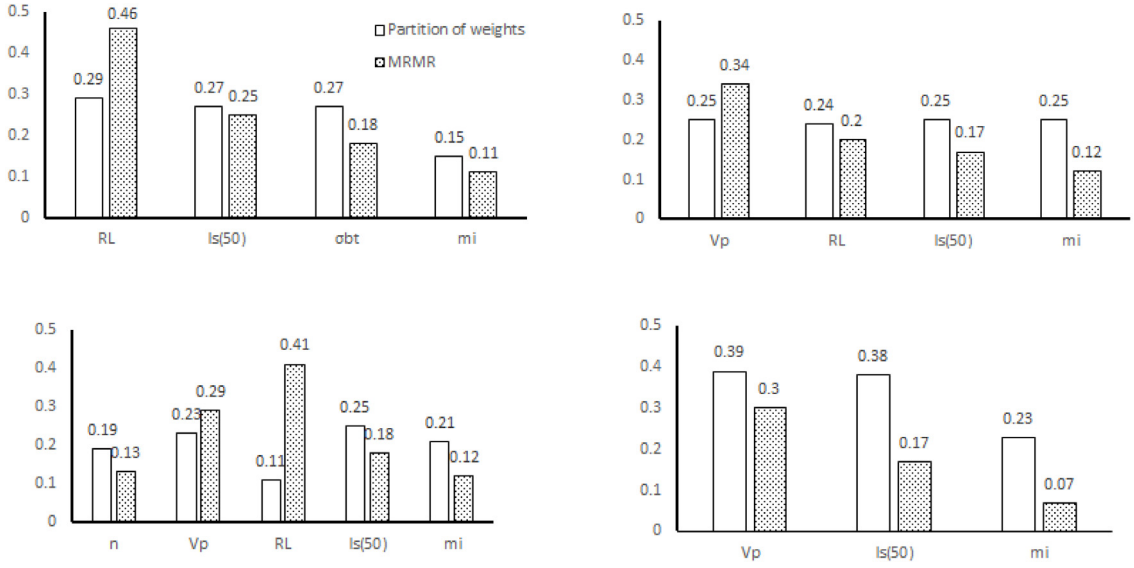

Fig. 5. Parametric analysis for the importance of input variables to the prediction of compressive strength σ_c.

Table 10
Performance indicators for the predictive ANN models.

Combination	No of data	RMSE	R^2	Combination	No of data	RMSE	R^2
R_L-$I_{s(50)}$- σ_{bt}				n-V_p- R_L - $Is_{(50)}$			
Training	159	0.0708	0.4567	Training	402	0.0086	0.8971
Validation	34	0.0558	0.4002	Validation	86	0.0106	0.8703
Test	34	0.0741	0.3088	Test	86	0.0152	0.8502
R_L-$I_{s(50)}$- σ_{bt} -m_i				n-V_p- R_L - $Is_{(50)}$–m_i			
Training	159	0.006	0.9138	Training	402	0.0088	0.8979
Validation	34	0.0078	0.8315	Validation	86	0.0116	0.8444
Test	34	0.0096	0.8926	Test	86	0.0083	0.9198
R_L-$I_{s(50)}$- σ_{bt} -m_{iab}				n-V_p- R_L - $Is_{(50)}$-m_{iab}			
Training	68	0.0091	0.8718	Training	265	0.0052	0.9548
Validation	14	0.0159	0.86	Validation	57	0.0055	0.9586
Test	14	0.0162	0.7743	Test	57	0.0077	0.9348
V_p- $I_{s(50)}$				V_p- R_L - $I_{s(50)}$			
Training	1018	0.0032	0.8276	Training	834	0.0828	0.5815
Validation	218	0.0028	0.7725	Validation	179	0.0992	0.438
Test	218	0.0033	0.8005	Test	179	0.1157	0.3716
V_p- $I_{s(50)}$-m_i				V_p- R_L - $I_{s(50)}$- m_i			
Training	1018	0.0025	0.8703	Training	834	0.0072	0.8525
Validation	218	0.0024	0.8142	Validation	179	0.0088	0.7752
Test	218	0.0024	0.8087	Test	179	0.0069	0.8586
V_p- $I_{s(50)}$-m_{iab}				V_p- R_L - $I_{s(50)}$- m_{iab}	658		
Training	519	0.0113	0.846	Training	460	0.0081	0.8948
Validation	111	0.012	0.8484	Validation	99	0.0098	0.8818
Test	111	0.0095	0.8293	Test	99	0.015	0.8068

5.3. Classification models

The 28 models were arranged in order of accuracy of validation. The predictive networks are then made to simulate the test data sets so as to validate the models prediction capability. The model with the highest accuracy on testing data classification is selected as the best algorithm for this combination of parameters. The overall classification performance results for the tested classification algorithms are presented in Fig. 7 and Table 12. Linear SVM shows the higher performance (80%) based on the testing data set accuracy for the model including R_L-$I_{s(50)}$-σ_{bt}-m_i as predictors. Cubic SVM is the most accurate (81.60%) classifier for the model including n-V_p-R_L-$I_{s(50)}$-m_i as predictors. Bagged trees present the highest performance 73.50% both for the model including V_p-R_L-$I_{s(50)}$-m_i as predictors, and for the model including V_p-$I_{s(50)}$-m_i as predictors with an accuracy of 69.10%. Overall the models with the highest performance are n-V_p-R_L-$I_{s(50)}$-m_i and R_L-$I_{s(50)}$-σ_{bt}-m_i.

Fig. 8 illustrates the confusion matrix which is a measure of accuracy of the classification models. This kind of validation allows for further insight into the successful classification at class level. The diagonal elements indicate the number of correctly classified instances of a given class. The off-diagonal elements are instances of misclassified observations. The figure shows the true positives rates and false positive rates for the classified data. For example, Linear SVM, correctly classified 80% of test data, originally belonging to class 1 (Very weak) and misclassified 20% of them to class 2 (Weak). Looking at all four confusion matrices resulting from the four combinations, we notice that the classifiers have a problem to classify accurately Class 3 (Medium Strong) that is σ_c in the range 25–50 MPa and Class 2 (Weak) σ_c in the range of 5–25 MPa only for the combination (n-V_p-R_L-$I_{s(50)}$$m_i$), which is perhaps justified by the limited sample data falling in these categories.

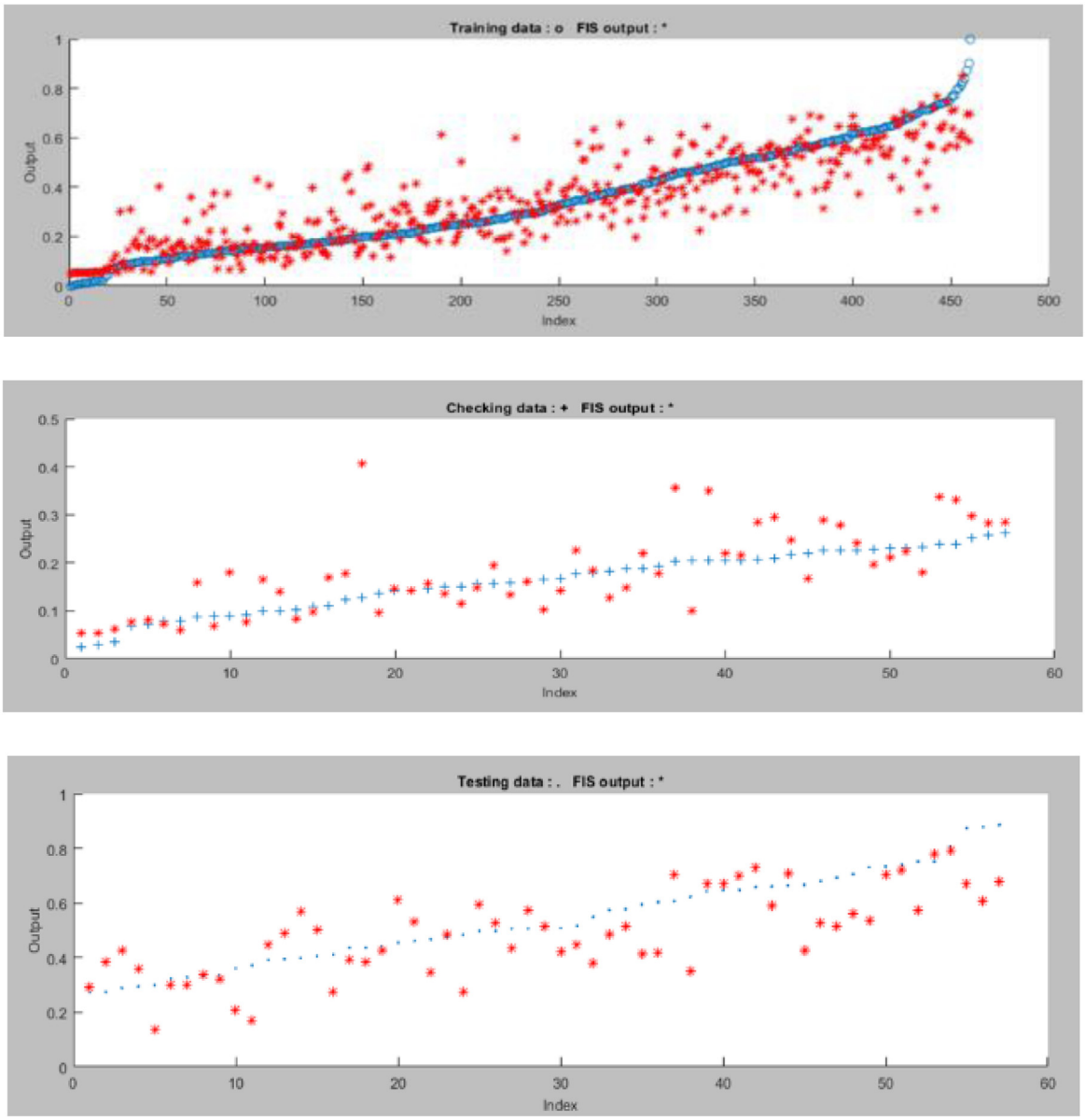

Fig. 6. ANFIS results for training, checking and testing data for n-V_p-R_L- $I_{s(50)}$-m_i combination.

6. Discussion

In this paper, a global intact rock database is compiled (ROCK/10/4025) and used to train generic predictive models for σ_c prediction for multiple lithologies. The database includes samples from igneous, sedimentary and metamorphic intact rocks with a wide range of characteristics from a wide range of geographical regions.

According to the findings of this paper the lithology seems to be an important factor contributing to the estimation of σ_c. Based on the study of this data base and the associated transformation models presented in the cases that included these results, there seems to be agreement with the findings of [96] that suggest that there is no strong evidence that the transformation relationships among intact rock properties exhibited by ROCK/10/4025 data points depend on rock classes (igneous, sedimentary, and metamorphic) or the degree of weathering and metamorphism.

Analysis of partitioning of connection weights and analysis minimum redundancy maximum relevance algorithm, indicate that R_L, and V_p parameters seem to be the most significant in terms of estimating σ_c, while m_i parameter is equally important as $I_{s(50)}$.

Computational intelligence algorithms were employed to predict σ_c of intact rock, and classify rock samples to a reliable grade of strength from extremely weak to extremely strong. Regression and classification machine learning techniques including artificial neural networks (ANN), artificial neuro-fuzzy inference systems (ANFIS), support vector machines (SVM) and ensemble bagged tress models (EBTM) were tested. Multiple statistical evaluation criteria were used to assess the performance of the machine learning algorithms on prediction results.

For the performance evaluation of the prediction models, σ_c values were estimated from different constructed combinations of physical and mechanical rock properties by the machine learning algorithms. A comparison of the performance of the results for all suggested models are summarized in Table 13. The R^2 value of 0.85–0.96 shows a good prediction performance of the ANFIS model for all combinations. ANN shows an equally satisfactory performance with RMSE values of 0.80–0.96, while the classifiers, seem to be less adequate with 0.70–0.82.

A qualitative analysis of the results suggests that certain combinations of parameters resulted in predictive models of higher performance.

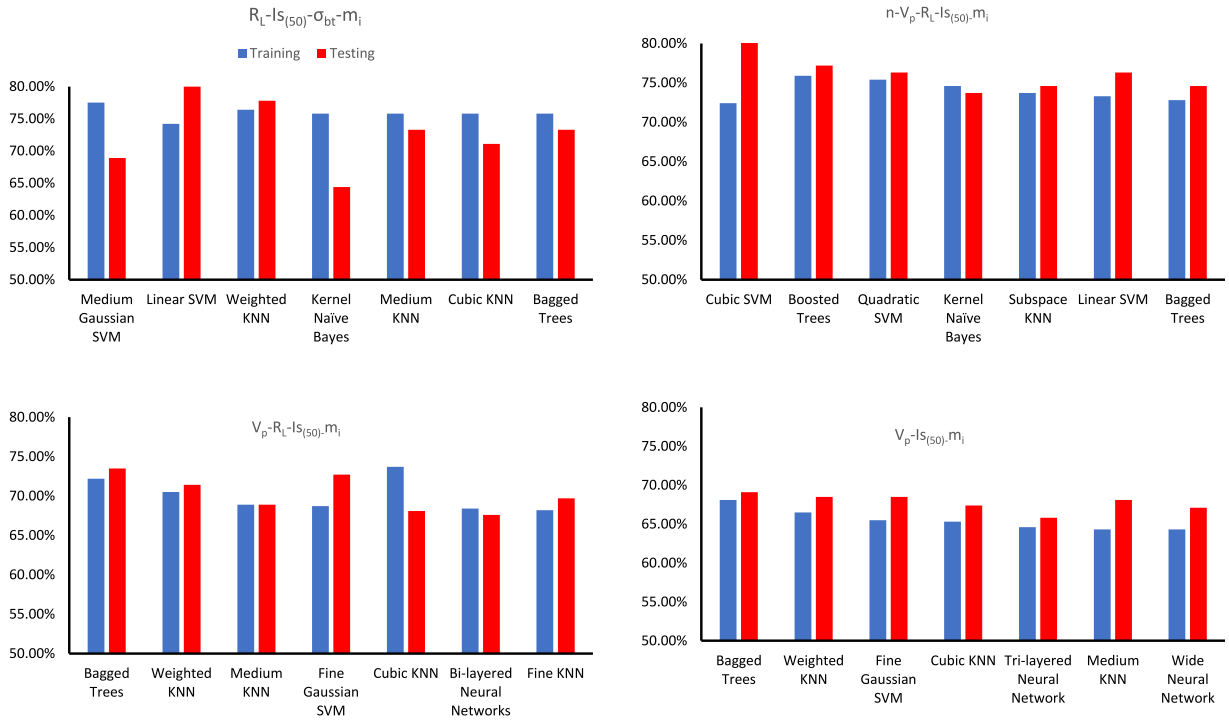

Fig. 7. Comparing the training and testing performance of for the applied classifiers, for the four selected combinations of parameters.

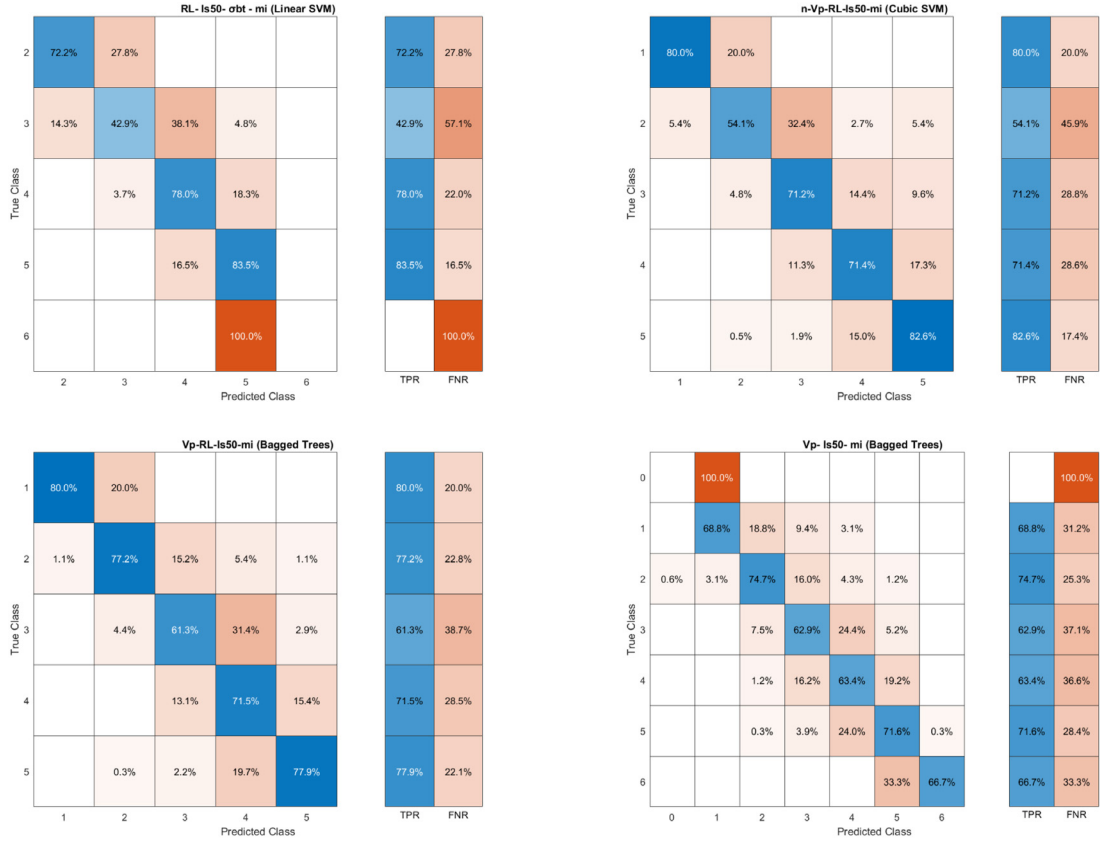

Fig. 8. Confusion matrix, of linear SVM, Cubic SVM, and Bagged Trees.

Table 11
ANFIS training results.

Nr	Model	ROI	SF	Testing data error	Training data error	Network
	R_L-$I_{s(50)}$-σ_{bt}-m_i					
1	i	0.5	1.2	0.076082	0.059052	Good
2	ii	0.5	1.3	0.075723	0.058672	Good
3	iii	0.5	1.4	0.072524	0.058966	Good
4	iv	0.5	1.5	0.057036	0.069115	Good
5	v	0.6	1.25	0.081682	0.062799	Good
6	vi	0.7	1.25	0.089339	0.063073	Good
7	vii	0.8	1.25	0.084677	0.062057	Good
8	viii	0.9	1.25	0.089502	0.08222	Good
	V_p-R_L-$I_{s(50)}$-m_i					
9	i	0.5	1.2	0.063801	0.079464	Poor
10	ii	0.5	1.3	0.067232	0.08016	Poor
11	iii	0.5	1.4	0.064704	0.081719	Good
12	iv	0.5	1.5	0.065504	0.082955	Good
13	v	0.5	1.6	0.06763	0.087915	Good
14	vi	0.6	1.25	0.064931	0.083956	Poor
15	vii	0.7	1.25	0.066352	0.089601	Good
16	viii	0.8	1.25	0.066577	0.089459	Good
17	ix	0.9	1.25	0.066517	0.089438	Good
	n-V_p-R_L-$I_{s(50)}$-m_i					
18	i	0.5	1.2	0.076215	0.08106	Good
19	ii	0.5	1.3	0.076471	0.07863	Good
20	iii	0.5	1.4	0.080227	0.083866	Good
21	iv	0.5	1.5	0.082032	0.086613	Good
22	v	0.6	1.25	0.071893	0.081752	Good
23	vi	0.7	1.25	0.074205	0.088143	Good
24	vii	0.8	1.25	0.07362	0.088447	Good
25	viii	0.9	1.25	0.069552	0.085715	Good
	V_p-$I_{s(50)}$-m_i					
27	i	0.5	1.2	0.072323	0.055964	Good
28	ii	0.5	1.3	0.072323	0.055964	Good
29	iii	0.5	1.4	0.071863	0.054258	Poor
30	iv	0.5	1.5	0.071956	0.056368	Good
31	v	0.5	1.6	0.070917	0.060226	Poor
32	vi	0.6	1.25	0.066594	0.055454	Poor
33	vii	0.7	1.25	0.068336	0.054695	Poor
34	viii	0.8	1.25	0.071812	0.055403	Good

R_L-$I_{s(50)}$-σ_{bt} – m_i combination has the highest performance, and includes index properties and strength properties (non-destructive R_L index-Point load index I_{s50}, Brazilian indirect tensional strength σ_{bt}, and constant m_i). The model systematically outperformed, both in function approximation, with ANN, ANFIS algorithms, and in classification problem with Linear SVM. The specific combination has the smallest data set 227 no of sets. The number of data satisfies the lower bound of number of samples vs. net size needed such that valid generalization can be expected.

Next high performer model/combination includes n-V_p-R_L-$I_{s(50)}$-m_i (index property n porosity, and R_L Schmidt rebound index, dynamic property V_p, and constant m_i). This model is systematically ranked high in terms of performance capability when ANN, ANFIS or Cubic SVM algorithms are used. This combination includes 575 number of sets. It does not seem however that inclusion of physical properties in training affect the performance of the ML models.

The other two combinations V_p-R_L-$I_{s(50)}$-mi and V_p-$I_{s(50)}m_i$ in descending order, show a lower prediction capability according to the statistic metrics, for ANN, ANFIS, KNN, ensemble classifiers and Bagged Trees. It is interesting though to underline that these classification models performed in a more balanced level, similar prediction accuracy for all the classes showing less false misclassifications. It is also of note that the number of sample data was higher almost double or triple than the latter models. This perhaps suggests that the performance capability is not necessarily associated to the sample of data, but to the quality of data, and the inclusion of specific parameters in the predictive models. This links to what Phoon and Zhang [126] suggest that a deep appreciation of the geotechnical context is critical to the development of novel ML methods that can lead to 'data-centric geotechnics' as a distinctive field that can transform practice.

7. Conclusions

In this paper, a global intact rock database (ROCK/10/4025) is developed and compared with ROCK/9/4069 database. This global database contains igneous, sedimentary, and metamorphic intact rocks which cover a wide range of lithologies and selected from a wide range of geographical locales. The database includes rock index properties, strength stiffness and dynamic properties. For the compiled data sets, Hoek–Brown constant mi, was estimated, using Hoek and Brown proposed guidelines for determining mi values for different rock types that can be used for preliminary design when triaxial tests are not available.

The main objective of this paper is to produce generic model that can estimate a reliable value for σ_c, for multiple lithologies. The proposed models used a function approximation approach and a classification approach. The results of the analysis indicate that the function approximation approach yields more reliable results with ANFIS algorithm to appear marginally superior. The classification approach suggests that support vector machines, nearest neighbour and ensemble classifiers performed better. The details of performance index indicators suggest the highest 90% performance for ANN model and 96% performance for ANFIS. The classifiers' accuracy was 82% for the best performer Cubic SVM.

The selected parameters and sample data suggest that for the selected combinations, and training data sets, the model that includes data on a physical parameter n, P-wave velocity, R_L, $I_{s(50)}$ and m_i,

Table 12
Performance of classification models.

R_L-$I_{s(50)}$-σ_{bt}-m_i				
No	Model	Model type	Training accuracy	Testing accuracy
1	SVM	Medium Gaussian SVM	77.50%	68.90%
2	SVM	Linear SVM	74.20%	80.00%
3	KNN	Weighted KNN	76.40%	77.80%
4	Naïve Bayes	Kernel Naïve Bayes	75.80%	64.40%
5	KNN	Medium KNN	75.80%	73.30%
6	KNN	Cubic KNN	75.80%	71.10%
7	Ensemble	Bagged Trees	75.80%	73.30%
n-V_p-R_L-$I_{s(50)}$-m_i				
1	SVM	Cubic SVM	72.40%	81.60%
2	Ensemble	Boosted Trees	75.90%	77.20%
3	SVM	Quadratic SVM	75.40%	76.30%
4	Naïve Bayes	Kernel Naïve Bayes	74.60%	73.70%
5	Ensemble	Subspace KNN	73.70%	74.60%
6	SVM	Linear SVM	73.30%	76.30%
7	Ensemble	Bagged Trees	72.80%	74.60%
V_p-R_L-$I_{s(50)}$-m_i				
1	Ensemble	Bagged Trees	72.20%	73.50%
2	KNN	Weighted KNN	70.50%	71.40%
3	KNN	Medium KNN	68.90%	68.90%
4	SVM	Fine Gaussian SVM	68.70%	72.70%
5	KNN	Cubic KNN	73.70%	68.10%
6	Neural Network	Bi-layered Neural Networks	68.40%	67.60%
7	KNN	Fine KNN	68.20%	69.70%
V_p-$I_{s(50)}$-m_i				
1	Ensemble	Bagged Trees	68.10%	69.10%
2	KNN	Weighted KNN	66.50%	68.50%
3	SVM	Fine Gaussian SVM	65.50%	68.50%
4	KNN	Cubic KNN	65.30%	67.40%
5	Neural Network	Tri-layered Neural Network	64.60%	65.80%
6	KNN	Medium KNN	64.30%	68.10%
7	Neural Network	Wide Neural Network	64.30%	67.10%

Table 13
Performance of predictive models for the four selected combinations.

Method	R^2	RMSE
R_L-$I_{s(50)}$-σ_{bt}-m_i		
ANN	0.89	16.72
ANFIS	0.96	10.65
L.SVM	0.81	0.39
V_p-$I_{s(50)}$-m_i		
ANN	0.84	19.08
ANFIS	0.85	22.89
Ensemble	0.81	0.48
n-V_p-R_L-$I_{s(50)}$-m_i		
ANN	0.90	23.33
ANFIS	0.91	19.09
C.SVM	0.76	0.50
V_p-R_L-$I_{s(50)}$-m_i		
ANN	0.82	23.33
ANFIS	0.91	19.09
Ensemble	0.77	0.50

has a higher generalization capability although trained with a smaller sample.

The ranking of the parameters suggests that R_L, and V_p are important and m_i and $I_{s(50)}$ are mutually important.

The models developed in this study are applicable to a plethora of tock types (more than forty) lithologies were included in the data base. These models could prove to be powerful tools that allow for a reliable estimation of the σ_c for any such data set including the ten parameters found in the data base. These models are not classified as decision or design support tools but provide a method that allows for a generic model, calibrated by a global broad rock database that allow for an estimate of σ_c value or classification of σ_c.

These models may be extended in various ways. One may be to refine and/or densify it for a given rock group by inclusion of other site specific samples. Another extension could be to form more comprehensive sets that include specific rock group similarities. This could allow possible ways to analyse the strength data of different classes or groups of, igneous rock material assuming strength parameters that are similar (supposing they are all igneous) but not identical (in that they are different igneous materials). Nonetheless, this kind of analysis will require further campaigns of data gathering from the literature, and further research is required to investigate the feasible extent of such expansions.

Details of the global database can be found in supplementary material.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank the reviewers for their constructive comments that significantly improved the quality of the paper.

Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.probengmech.2022.103400.

References

[1] K.K. Phoon, F.H. Kulhawy, Evaluation of geotechnical property variability, Can. Geotech. J. 36 (4) (1999) 625–639, http://dx.doi.org/10.1139/t99-039.

[2] Y. Abdi, A.T. Garavand, R.Z. Sahamieh, Prediction of strength parameters of sedimentary rocks using artificial neural networks and regression analysis, Arab. J. Geosci. 11 (19) (2018).

[3] L.O. Afolagboye, A.O. Talabi, C.A. Oyelami, The use of index tests to determine the mechanical properties of crushed aggregates from precambrian basement complex rocks, Ado-Ekiti, SW Nigeria, J. Afr. Earth Sci. 129 (2017) 659–667.

[4] G. Aggistalis, A. Alivizatos, D. Stamoulis, G. Stournaras, Correlating uniaxial compressive strength with Schmidt hardness, point load index, Young's modulus, and mineralogy of gabbros and basalts (northern Greece), Bulletin - Inter. Assoc. Eng. Geol. 54 (1996) 3–11.

[5] M. Akram, M.Z.A. Bakar, Correlation between uniaxial compressive strength and point load index for salt-range rocks, Pak. J. Engg. & Appl. Sci. 1 (2007).

[6] M.M. Aliyu, J. Shang, W. Murphy, J.A. Lawrence, R. Collier, F. Kong, Z. Zhao, Assessing the uniaxial compressive strength of extremely hard cryptocrystalline flint, Int. J. Rock Mech. Min. Sci. 113 (2019) 310–321.

[7] D. Jahed Armaghani, E. Tonnizam Mohamad, E. Momeni, M.S. Narayanasamy, M.F. Mohd Amin, An adaptive neuro-fuzzy inference system for predicting unconfined compressive strength and Young's modulus: a study on Main Range granite, Bull. Eng. Geol. Environ. 74 (4) (2015) 1301–1319.

[8] D.J. Armaghani, E. Tonnizam Mohamad, E. Momeni, M. Monjezi, M. Sundaram Narayanasamy, Prediction of the strength and elasticity modulus of granite through an expert artificial neural network, Arab. J. Geosci. 9 (1) (2016) 1–16, art. (48).

[9] A. Aydin, A. Basu, The Schmidt hammer in rock material characterization, Eng. Geol. 81 (1) (2005) 1–14.

[10] A. Azimian, Application of statistical methods for predicting uniaxial compressive strength of limestone rocks using nondestructive tests, Acta Geotech. 12 (2) (2017) 321–333.

[11] A. Basu, M. Kamran, Point load test on schistose rocks and its applicability in predicting uniaxial compressive strength, Int. J. Rock Mech. Min. Sci. 47 (5) (2010) 823–828.

[12] A. Basu, T.B. Celestino, A.A. Bortolucci, Evaluation of rock mechanical behaviors under uniaxial compression with reference to assessed weathering grades, Rock Mech. Rock Eng. 42 (1) (2009) 73–93.

[13] F.G. Bell, P. Lindsay, The petrographic and geomechanical properties of some sandstones from the Newspaper Member of the Natal Group near Durban, South Africa, Eng. Geol. 53 (1) (1999) 57–81.

[14] Z.T. Bieniawski, The point-load test in geotechnical practice, Eng. Geol. 9 (1) (1975) 1–11.

[15] N. Bilgin, H. Copur, C. Balci, Use of Schmidt Hammer with special reference to strength reduction factor related to cleat presence in a coal mine, Int. J. Rock Mech. Min. Sci. 84 (2016) 25–33.

[16] Z. Briševac, D. Špoljarić, V. Gulam, Estimation of uniaxial compressive strength based on regression tree models, Rudarsko Geolosko Naftni Zbornik 29 (1) (2014) 39–47.

[17] G. Bruno, G. Vessia, L. Bobbo, Statistical method for assessing the uniaxial compressive strength of carbonate rock by schmidt hammer tests performed on core samples, Rock Mech. Rock Eng. 46 (1) (2013) 199–206.

[18] J.S. Cargill, A. Shakoor, Evaluation of empirical methods for measuring the uniaxial compressive strength of rock, Int. J. Rock Mech. Min. Sci. Geomech. Abstr. 27 (6) (1990) 495–503, http://dx.doi.org/10.1016/0148-9062(90)91001-N.

[19] S.B. Çelik, Prediction of uniaxial compressive strength of carbonate rocks from non-destructive tests using multivariate regression and LS-SVM methods, Arab. J. Geosci. 12 (6) (2019) 193.

[20] N. Ceryan, U. Okkan, A. Kesimal, Prediction of unconfined compressive strength of carbonate rocks using artificial neural networks, Environ. Earth Sci. 68 (3) (2013) 807–819.

[21] A. Cheshomi, E. Mousavi, E. Ahmadi-Sheshde, Evaluation of single particle loading test to estimate the uniaxial compressive strength of sandstone, J. Pet. Sci. Eng. 135 (2015) 421–428.

[22] İ. Çobanoğlu, S.B. Çelik, Estimation of uniaxial compressive strength from point load strength, Schmidt hardness and P-wave velocity, Bull. Eng. Geol. Environ. 67 (2008) 491–498, http://dx.doi.org/10.1007/s10064-008-0158-x.

[23] S. Dehghan, G. Sattari, S. Chehreh Chelgani, M.A. Aliabadi, Prediction of uniaxial compressive strength and modulus of elasticity for Travertine samples using regression and artificial neural networks, Min. Sci. Technol. 20 (1) (2010) 41–46.

[24] S. Demirdag, N. Sengun, I. Ugur, R. Altindag, Estimating the uniaxial compressive strength of rocks with Schmidt rebound hardness by considering the sample size, Arab. J. Geosci. 11 (17) (2018) art. no. 502.

[25] K. Diamantis, E. Gartzos, G. Migiros, Study on uniaxial compressive strength, point load strength index, dynamic and physical properties of serpentinites from Central Greece: Test results and empirical relations, Eng. Geol. 108 (3–4) (2009) 199–207.

[26] I. Dinçer, A. Acar, I. Çobanoğlu, Y. Uras, Correlation between Schmidt hardness, uniaxial compressive strength and Young's modulus for andesites, basalts and tuffs, Bull. Eng. Geol. Environ. 63 (2) (2004) 141–148.

[27] I. Dinçer, A. Acar, S. Ural, Estimation of strength and deformation properties of Quaternary caliche deposits, Bull. Eng. Geol. Environ. 67 (3) (2008) 353–366.

[28] H. Ersoy, S. Acar, Influences of petrographic and textural properties on the strength of very strong granitic rocks, Environ. Earth Sci. 75 (22) (2016) art. no. 1461.

[29] M. Fakir, M. Ferentinou, S. Misra, An investigation into the rock properties influencing the strength in some Granitoid rocks of KwaZulu-Natal, South Africa, Geotech. Geol. Eng. 35 (3) (2017) 1119–1140.

[30] M. Fener, S. Kahraman, A. Bilgil, O. Gunaydin, A comparative evaluation of indirect methods to estimate the compressive strength of rocks, Rock Mech. Rock Eng. 38 (4) (2005) 329–343.

[31] M. Ferentinou, M. Fakir, An ANN approach for the prediction of uniaxial compressive strength, of some sedimentary and igneous rocks in eastern KwaZulu-Natal, Procedia Eng. 191 (2017) 1117–1125.

[32] E. Ghasemi, H. Kalhori, R. Bagherpour, S. Yagiz, Model tree approach for predicting uniaxial compressive strength and Young's modulus of carbonate rocks, Bull. Eng. Geol. Environ. 77 (1) (2018) 331–343.

[33] M. Gomez-Heras, D. Benavente, C. Pla, J. Martinez-Martinez, R. Fort, V. Brotons, Ultrasonic pulse velocity as a way of improving uniaxial compressive strength estimations from Leeb hardness measurements, Constr. Build. Mater. 261 (2020) 119996.

[34] J. Gonz'lez, M. Saldaña, J. Arzúa, Analytical model for predicting the UCS from P-wave velocity, density, and porosity on saturated limestone, Appl. Sci. 9 (23) (2019) 5265.

[35] A. Guney, R. Altindag, Evaluation of the Relationships between Schmidt Hardness Rebound Number and Other (Engineering) Properties of Rocks, in: The 19th International Mining Congress and Fair of Turkey, IMCET 2005, İzmir, Turkey.

[36] R. Hebib, D. Belhai, B. Alloul, Estimation of uniaxial compressive strength of North Algeria sedimentary rocks using density, porosity, and Schmidt hardness, Arab. J. Geosci. 10 (17) (2017) 383.

[37] M. Heidari, G.R. Khanlari, M. Torabi-Kaveh, Reply to discussion by Li and Wong on "predicting the uniaxial compressive and tensile strengths of gypsum rock by point load testing" by Heidari et al., rock mechanics and rock engineering (2012) 45:265-273, Rock Mech. Rock Eng. 45 (6) (2012) 1131–1135.

[38] M. Heidari, H. Mohseni, S.H. Jalali, Prediction of uniaxial compressive strength of some sedimentary rocks by fuzzy and regression models, Geotech. Geol. Eng. 36 (1) (2018) 401–412.

[39] I. Ince, M. Fener, A prediction model for uniaxial compressive strength of deteriorated pyroclastic rocks due to freeze-thaw cycle, J. Afr. Earth Sci. 120 (2016) 134–140.

[40] İ. İnce, A. Bozdağ, M. Fener, S. Kahraman, Estimation of uniaxial compressive strength of pyroclastic rocks (Cappadocia, Turkey) by gene expression programming, Arab. J. Geosci. 12 (24) (2019) 756.

[41] A. Jamshidi, R. Yazarloo, S. Gheiji, Comparative evaluation of Schmidt hammer test procedures for prediction of rock strength, Int. J. Mining and Geo-Eng. (2018).

[42] S. Kahraman, O. Gunaydin, The effect of rock classes on the relation between uniaxial compressive strength and point load index, Bull. Eng. Geol. Environ. 68 (3) (2009) 345–353.

[43] S. Kahraman, Evaluation of simple methods for assessing the uniaxial compressive strength of rock, Int. J. Rock Mech. Min. Sci. 38 (7) (2001) 981–994.

[44] S. Kahraman, C. Balci, S. Yazici, N. Bilgin, Prediction of the penetration rate of rotary blast hole drills using a new drillability index, Int. J. Rock Mech. Min. Sci. 37 (5) (2000) 729–743, 926.

[45] S. Kahraman, The determination of uniaxial compressive strength from point load strength for pyroclastic rocks, Eng. Geol. 170 (2014) 33–42.

[46] S. Kahraman, M. Fener, C.O. Kilic, A preliminary study on the conversion factor used in the prediction of the UCS from the BPI for pyroclastic rocks, Bull. Eng. Geol. Environ. 75 (2) (2016) 771–780.

[47] S. Kahraman, A.N. Canpolat, M. Fener, The influence of microwave treatment on the compressive and tensile strength of igneous rocks, Int. J. Rock Mech. Min. Sci. 129 (2020) 104303.

[48] A. Kainthola, P.K. Singh, D. Verma, R. Singh, K. Sarkar, T.N. Singh, Prediction of strength parameters of himalayan rocks: A statistical and ANFIS approach, Geotech. Geol. Eng. 33 (5) (2015) 1255–1278.

[49] M. Kamani, R. Ajalloeian, Evaluation of the mechanical degradation of carbonate aggregate by rock strength tests, J. Rock Mech. Geotech. Eng. 11 (1) (2019) 121–134.

[50] M. Karakus, M. Kumral, O. Kilic, Predicting elastic properties of intact rocks from index tests using multiple regression modelling, Int. J. Rock Mech. Min. Sci. 42 (2) (2005) 323–330.

[51] M. Karakus, Function identification for the intrinsic strength and elastic properties of granitic rocks via genetic programming (GP), Comput. Geosci. 37 (9) (2011) 1318–1323.

[52] K. Karaman, A. Kesimal, H. Ersoy, A comparative assessment of indirect methods for estimating the uniaxial compressive and tensile strength of rocks, Arab. J. Geosci. 8 (4) (2015) 2393–2403.

[53] K. Karaman, A. Kesimal, A comparative study of Schmidt hammer test methods for estimating the uniaxial compressive strength of rocks, Bull. Eng. Geol. Environ. 74 (2) (2015) 507–520.
[54] M. Kasim, A. Shakoor, An investigation of the relationship between uniaxial compressive strength and degradation for selected rock types, Eng. Geol. 44 (1–4) (1996) 213–227.
[55] M. Khandelwal, T.N. Singh, Correlating static properties of coal measures rocks with P-wave velocity, Int. J. Coal Geol. 79 (1–2) (2009) 55–60.
[56] A. Kılıç, A. Teymen, Determination of mechanical properties of rocks using simple methods, Bull. Eng. Geol. Environ. 67 (2) (2008) 237–244.
[57] M. Korkanç, B. Solak, Estimation of engineering properties of selected tuffs by using grain/matrix ratio, J. Afr. Earth Sci. 120 (2016) 160–172.
[58] N. Mahdiabadi, G. Khanlari, Prediction of uniaxial compressive strength and modulus of elasticity in calcareous mudstones using neural networks, fuzzy systems, and regression analysis, Period. Polytech. Civil Eng. 63 (1) (2019) 104–114.
[59] N. Madhubabu, P.K. Singh, A. Kainthola, B. Mahanta, A. Tripathy, T.N. Singh, Prediction of compressive strength and elastic modulus of carbonate rocks, Measurement: J. Int. Meas.Confed. 88 (2016) 202–213.
[60] A. Mahmoodzadeh, M. Mohammadi, H. Hashim Ibrahim, S. Nariman Abdulhamid, S. Ghafoor Salim, H. Farid Hama Ali, M. Kamal Majeed, Artificial intelligence forecasting models of uniaxial compressive strength, Transp. Geotech. 27 (2021) 100499.
[61] M. Ludovico-Marques, C. Chastre, G. Vasconcelos, Modelling the compressive mechanical behaviour of granite and sandstone historical building stones, Constr. Build. Mater. 28 (1) (2012) 372–381.
[62] F.F. Martins, A. Begonha, M. Amália Sequeira Braga, Prediction of the mechanical behavior of the Oporto granite using Data Mining techniques, Expert Syst. Appl. 39 (10) (2012) 8778–8783.
[63] S. Mehrabi Mazidi, M. Haftani, B. Bohloli, A. Cheshomi, Measurement of uniaxial compressive strength of rocks using reconstructed cores from rock cuttings, J. Pet. Sci. Eng. 86–87 (2012) 39–43.
[64] D.A. Mishra, A. Basu, Estimation of uniaxial compressive strength of rock materials by index tests using regression analysis and fuzzy inference system, Eng. Geol. 160 (2013) 54–68.
[65] E.T. Mohamad, D. Jahed Armaghani, E. Momeni, S.V. Alavi Nezhad Khalil Abad, Prediction of the unconfined compressive strength of soft rocks: a PSO-based ANN approach, Bull. Eng. Geol. Environ. 74 (3) (2015) 745–757.
[66] E. Momeni, D. Jahed Armaghani, M. Hajihassani, M.F. Mohd Amin, Prediction of uniaxial compressive strength of rock samples using hybrid particle swarm optimization-based artificial neural networks, Measurement: J. Int. Meas. Confed. 60 (2015) 50–63.
[67] H.A. Nefeslioglu, Evaluation of geo-mechanical properties of very weak and weak rock materials by using non-destructive techniques: Ultrasonic pulse velocity measurements and reflectance spectroscopy, Eng. Geol. 160 (2013) 8–20.
[68] I.-T. Ng, K.-V. Yuen, C.-H. Lau, Predictive model for uniaxial compressive strength for Grade III granitic rocks from Macao, Eng. Geol. 199 (2015) 28–37.
[69] V. Palchik, Y.H. Hatzor, The influence of porosity on tensile and compressive strength of porous chalks, Rock Mech. Rock Eng. 37 (4) (2004) 331–341.
[70] C.I. Sachpazis, Correlating Schmidt hardness with compressive strength and young's modulus of carbonate rocks, Bull. Int. Assoc. Eng. Geol. 42 (1) (1990) 75–83.
[71] M. Saldaña, J. González, I. Pérez-Rey, M. Jeldres, N. Toro, Applying statistical analysis and machine learning for modeling the UCS from P-wave velocity, density and porosity on dry travertine, Appl. Sci. 10 (13) (2020) 4565.
[72] S. Salehin, E. Hadavandi, S.C. Chelgani, Exploring relationships between mechanical properties of marl core samples by a coupling of mutual information and predictive ensemble model, Model. Earth Syst. Environ. 6 (1) (2020) 575–583.
[73] K. Sarkar, A. Tiwary, T.N. Singh, Estimation of strength parameters of rock using artificial neural networks, Bull. Eng. Geol. Environ. 69 (4) (2010) 599–606.
[74] N. Sengun, R. Altindag, S. Demirdag, H. Yavuz, P-wave velocity and Schmidt rebound hardness value of rocks under uniaxial compressional loading, Int. J. Rock Mech. Min. Sci. 48 (4) (2011) 693–696.
[75] F.I. Shalabi, E.J. Cording, O.H. Al-Hattamleh, Estimation of rock engineering properties using hardness tests, Eng. Geol. 90 (3–4) (2007) 138–147.
[76] P.K. Sharma, T.N. Singh, A correlation between P-wave velocity, impact strength index, slake durability index and uniaxial compressive strength, Bull. Eng. Geol. Environ. 67 (1) (2008) 17–22.
[77] A. Sharo, B. Al-Shorman, Correlation between unconfined compression strength and point load index for selected rocks from Jordan, in: Int. Civil Engineering and Architecture Conference 2019, Trabzon, Turkey.
[78] T.N. Singh, A. Kainthola, A. Venkatesh, Correlation between point load index and uniaxial compressive strength for different rock types, Rock Mech. Rock Eng. 45 (2) (2012) 259–264.
[79] S. Sulukcu, R. Ulusay, Evaluation of the block punch index test with particular reference to the size effect, failure mechanism and its effectiveness in predicting rock strength, Int. J. Rock Mech. Min. Sci. 38 (8) (2001) 1091–1111.
[80] R.S. Tandon, V. Gupta, Estimation of strength characteristics of different himalayan rocks from Schmidt hammer rebound, point load index, and compressional wave velocity, Bull. Eng. Geol. Environ. 74 (2) (2015) 521–533.
[81] A. Teymen, Prediction of basic mechanical properties of tuffs using physical and index tests, J. Min. Sci. 54 (5) (2018) 721–733.
[82] A. Teymen, E.C. Mengüç, Comparative evaluation of different statistical tools for the prediction of uniaxial compressive strength of rocks, Int. J. Mining Sci. Technol. 30 (6) (2020) 785–797.
[83] B. Tiryaki, Predicting intact rock strength for mechanical excavation using multivariate statistics, artificial neural networks, and regression trees, Eng. Geol. 99 (1–2) (2008) 51–60.
[84] S.R. Torabi M. Ataei M. Javanshir, Application of Schmidt rebound number for estimating rock strength under specific geological conditions, J. Min. Environ. (2019).
[85] D. Tumac, S. Hojjati, Predicting performance of impact hammers from rock quality designation and compressive strength properties in various rock masses, Tunn. Underground Space Technol. 59 (2016) 38–47.
[86] R. Ulusay, K. Türeli, M.H. Ider, Prediction of engineering properties of a selected litharenite sandstone from its petrographic characteristics using correlation and multivariate statistical techniques, Eng. Geol. 38 (1–2) (1994) 135–157, http://dx.doi.org/10.1016/0013-7952(94)90029-9.
[87] E. Vasanelli, A. Calia, D. Colangiuli, F. Micelli, M.A. Aiello, Assessing the reliability of non-destructive and moderately invasive techniques for the evaluation of uniaxial compressive strength of stone masonry units, Constr. Build. Mater. 124 (2016) 575–581.
[88] M. Wang, W. Wan, A new empirical formula for evaluating uniaxial compressive strength using the Schmidt hammer test, Int. J. Rock Mech. Min. Sci. 123 (2019) 104094.
[89] L. Wen, Z. Luo, S. Yang, Y. Qin, W. Wang, Correlation of geo-mechanics parameters with uniaxial compressive strength and P-wave velocity on dolomitic limestone using a statistical method, Geotech. Geol. Eng. 37 (2) (2019) 1079–1094.
[90] S. Yagiz, Utilizing rock mass properties for predicting TBM performance in hard rock condition, Tunn. Underground Space Technol. 23 (3) (2008) 326–339.
[91] O. Yarali, E. Soyer, Assessment of relationships between drilling rate index and mechanical properties of rocks, Tunnell. Underground Space Technol. 33 (2013) 46–53.
[92] H. Yavuz, R. Altindag, S. Sarac, I. Ugur, N. Sengun, Estimating the index properties of deteriorated carbonate rocks due to freeze-thaw and thermal shock weathering, Int. J. Rock Mech. Min. Sci. 43 (5) (2006) 767–775.
[93] H. Yavuz, I. Ugur, S. Demirdag, Abrasion resistance of carbonate rocks used in dimension stone industry and correlations between abrasion and rock properties, Int. J. Rock Mech. Min. Sci. 45 (2) (2008) 260–267.
[94] D. Jahed Armaghani, E. Tonnizam Mohamad, M. Hajihassani, S. Yagiz, H. Motaghedi, Application of several non-linear prediction tools for estimating uniaxial compressive strength of granitic rocks and comparison of their performances, Eng. Comput. 32 (2) (2016) 189–206.
[95] M. Heidari, G.R. Khanlari, M.T. Kaveh, S. Kargarian, Predicting the uniaxial compressive and tensile strengths of gypsum rock by point load testing, Rock Mech. Rock Eng. 45 (2) (2012) 265–273.
[96] J. Ching, K.H. Li, K.K. Phoon, M.C. Weng, Generic transformation models for some intact rock properties, Can. Geotech. J. 55 (12) (2018) 1702–1741, http://dx.doi.org/10.1139/cgj-2017-0537.
[97] BS ENV 1997-1, Eurocode 7—Geotechnical design, part 1: General rules, 1994, Brussels: European Committee for Standardization.
[98] K.K. Phoon, The story of statistics in geotechnical engineering, Georisk: Assess. Manag. Risk Eng. Syst. Geohazards 14 (1) (2020) 3–25, http://dx.doi.org/10.1080/17499518.2019.1700423.
[99] K.K. Phoon, J. Ching, T. Shuku, Challenges in data-driven site characterization, Georisk: Assess. Manag. Risk Eng. Syst. Geohazards 9 (2021) (2021) http://dx.doi.org/10.1080/17499518.2021.1896005, [Published online ahead of print].
[100] T. Rahman, K. Sarkar, Lithological control on the estimation of uniaxial compressive strength by the P-wave velocity using supervised and unsupervised learning, Rock Mech. Rock Eng. 54 (2021) 3175–3191, http://dx.doi.org/10.1007/s00603-021-02445-8.
[101] J. Ching, K.K. Phoon, Y-H. Ho, Weng. M.C., Quasi-site-specific prediction for deformation modulus of rock mass, Can. Geotech. J. 58 (7) (2020) 936–951, http://dx.doi.org/10.1139/cgj-2020-0168.
[102] J. Ching, K.K. Phoon, Modeling parameters of structured clays as a multivariate normal distribution, Can. Geotech. J. 49 (5) (2012) 522–545, http://dx.doi.org/10.1139/t2012-015.
[103] J. Ching, K.K. Phoon, Multivariate distribution for undrained shear strengths under various test procedures, Can. Geotech. J. 50 (9) (2013) 907–923, http://dx.doi.org/10.1139/cgj-2013-0002.
[104] J. Ching, K.K. Phoon, Correlations among some clay parameters – the multivariate distribution, Can. Geotech. J. 51 (6) (2014) 686–704, http://dx.doi.org/10.1139/cgj-2013-0353.
[105] J. Ching, K.K. Phoon, Transformations and correlations among some parameters of clays – the global database, Can. Geotech. J. 51 (6) (2014b) 663–685, http://dx.doi.org/10.1139/cgj-2013-0262.

[106] J. Ching, G.H. Lin, J.R. Chen, K.-K. Phoon, Transformation models for effective friction angle and relative density calibrated based on generic database of coarse-grained soils, Can. Geotech. J. 54 (4) (2017) 481–501, http://dx.doi.org/10.1139/cgj-2016-0318.
[107] F.H. Kulhawy, C.H. Trautmann, T.D. O'Rourke, The soil-rock boundary: What is it and where is it? in: Detection of and Construction At the Soil Rock Interface (GSP 38), ASCE, New York, 1991, pp. 1–15.
[108] E. Kolaiti, Z. Papadopoulos, Evaluation of schmidt rebound hammer testing: A critical approach, Bull. Int. Assoc. Eng. Geol. 48 (1) (1993) 69–76.
[109] C. Kurtulus, F. Sertçelik, I. Sertçelik, Estimation of unconfined uniaxial compressive strength using schmidt hardness and ultrasonic pulse velocity, Tehn. Vjes. 25 (5) (2018) 1569–1574.
[110] ISRM, The complete ISRM suggested methods for rock characterization, testing and monitoring: 1974–2006, in: R. Ulusay, J.A. Hudson (Eds.), Suggested Methods Prepared By the Commission on Testing Methods, International Society for Rock Mechanics, ISRM Turkish National Group, Ankara, 2007, p. 628.
[111] P. Grasso, S. Xu, A. Mahtab, Problems and promises of index testing of rocks, Rock Mech. (1992) 879–888, Balkema.
[112] J.A. Franklin, Suggested method for determining point load strength, Int. J. Rock Mech. Min. Sci. Geomech. Abstr. (ISSN: 0148-9062) 22 (2) (1985) 51–60, http://dx.doi.org/10.1016/0148-9062(85)92327-7.
[113] W.A. Prakoso, Reliability-based design of foundations on rock masses for transmission line and similar structures, Doctoral dissertation, Cornell University, 2002.
[114] F. Birch, The velocity of compressional waves in rocks to 10 kilobars, J. Geophys. Res. 65 (4) (1960).
[115] E. Hoek, Brown E.T., Practical estimates of rock mass strength, Int. J. Rock Mech. Min. Sci. 34 (8) (1997) 1165–1186.
[116] E. Hoek, P. Marinos, Predicting tunnel squeezing problems in weak heterogeneous rock masses, 2000, http://www.rockscience.com/hoek/references/H2000d.pdf (Oct. 7, 2022).
[117] E. Hoek, E.T. Brown, Underground Excavations in Rock, Institution of Mining and Metallurgy, London, 1980.
[118] E. Hoek, Rock mass properties. Chapter 11, Practical Rock Eng. (2007) http://www.rocscience.com/hoek/corner/11_Rock_mass_properties.pdf (Sep. 7, 2022).
[119] K.J. Douglas, The shear strength of rock masses, (Ph.D. thesis), School of Civil and Environmental Engineering, Univ. of New South Wales, Sydney, Australia, 2002.
[120] M. Cai, Practical estimates of tensile strength and Hoek–Brown parameter mi of brittle rocks, Rock Mech. Rock Eng. 43 (2) (2010) 167–184.
[121] J. Zuo, J. Shen, The Hoek-Brown Constant mi, in: The Hoek-Brown Failure Criterion—from Theory To Application, Springer, Singapore, 2020, http://dx.doi.org/10.1007/978-981-15-1769-3_5.
[122] S.A.L. Read, L. Richards, A comparative study of mi, the hoek–brown constant for intact rock material, in: Proceedings of the 12th International Society for Rock Mechanics Congress, ISRM, Lisbon, 2011.
[123] J. Peng, G. Rong, M. Cai, X. Wang, C. Zhou, An empirical failure criterion for intact rocks, Rock Mech. Rock Eng. 47 (2) (2013) 347–356.
[124] Shen, Karakus, Simplified method for estimating the Hoek-Brown constant for intact rocks, J. Geotech. Geoenviron. Eng. 140 (6) (2014) http://dx.doi.org/10.1061/(ASCE)GT.1943-5606.0001116.
[125] E Hoek, Practical Rock Eng. (2007) http://www.rocscience.com. (Oct 30., 2022).
[126] K.K. Phoon, W. Zhang, Future of machine learning in geotechnics, Georisk: Assess. Manag. Risk Eng. Syst. Geohazards (2022) http://dx.doi.org/10.1080/17499518.2022.2087884.
[127] M. Rezania, A.A. Javadi, A new genetic programming model for predicting the settlement of shallow foundations, Can. Geotech. J. 44 (12) (2007) 1462–1473.
[128] M. Zhou, M. Shadabfar, H. Huang, Y.F. Leung, S. Uchida, Meta-modelling of coupled thermo-hydro-mechanical behaviour of hydrate reservoir, Comput. Geotech. 128 (2020) 103848, http://dx.doi.org/10.1016/j.compgeo.2020.103848.
[129] C. Shi, Y.u. Wang, Non-parametric machine learning methods for interpolation of spatially varying non-stationary and non-Gaussian geotechnical properties, Geosci. Front. 12 (1) (2021) 339–350.
[130] M. Rezania, G. Ma, Stress–strain modelling of soils in drained and undrained conditions using a multi-model intelligent approach, in: International Conference on Information Technology in Geo-Engineering, vol. 41, Springer, 2019, pp. 9–428.
[131] Z.-L. Cheng, W.-H. Zhou, A. Garg, Genetic programming model for estimating soil suction in shallow soil layers in the vicinity of a tree, Eng. Geol. 268 (2020) 105506, http://dx.doi.org/10.1016/j.enggeo.2020.105506.
[132] M.S. Es-haghi, M. Rezania, M. Bagheri, Machine learning-based estimation of soil's true air-entry value from GSD curves, Gondwana Res. (ISSN: 1342-937X) (2022) http://dx.doi.org/10.1016/j.gr.2022.06.012, in press.
[133] R. Hush Don, B.G. Horne, What's new since Lippman? IEEE Signal Process. Mag. (1999) 9–39.
[134] E. Baum, D. Haussler, What size net gives valid generalisation? Neural Comput. 1 (1) (1989) 151–160, http://dx.doi.org/10.1162/neco.1989.1.1.151.
[135] H.J. Oh, B. Pradhan, Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area, Comput. Geosci. 37 (2011) 1264–1276.
[136] S. Maiti, R.K. Tiwari, A comparative study of artificial neural networks, Bayesian neural networks and adaptive neuro-fuzzy inference system in groundwater level prediction, Environ. Earth Sci. 71 (2014) 3147–3160, http://dx.doi.org/10.1007/s12665-013-2702-7.
[137] T. Takagi, M. Sugeno, Derivation of fuzzy control rules from human operator's control actions. in: Proceedings of the IFAC symposium on fuzzy information, knowledge representation anddecision analysis, 1983, pp 55–60.
[138] J.S.R. Jang, ANFIS: adaptive-network-based fuzzy inference system, IEEE Trans. Syst. Man Cybern. 23 (1993) 665–685.
[139] S.L. Chiu, An efficient method for extracting fuzzy classification rules from high dimensional data, Adv. Comput. Intell. 1 (1997) 1–7.
[140] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297, http://dx.doi.org/10.1007/BF00994018.
[141] X. Dong, Z. Yu, W. Cao, Y. Shi, Q. Ma, A survey on ensemble learning, Front. Comput. Sci. 14 (2020) 241–258.
[142] G.D. Garson, Interpreting neural-networks connection weights, AI Expert 6 (1991) 47–51.
[143] A.T.C. Goh, Seismic liquefaction potential assessed by neural networks, J. Geotech Eng. 120 (9) (1994) 1467–1480.
[144] M.G. Sakellariou, M.D. Ferentinou, A study of slope stability prediction using neural networks, Geotech. Geol. Eng. 23 (2005) 419–445, http://dx.doi.org/10.1007/s10706-004-8680-5.
[145] C. Ding, H. Peng, Minimum redundancy feature selection from microarray gene expression data, J. Bioinfor. Comput. Biol. 3 (2) (2005) 185–205.