Geomechanics and Engineering, *Vol. 34, No. 6 (2023) 697-726*
https://doi.org/10.12989/gae.2023.34.6.697

Predicting rock brittleness indices from simple laboratory test results using some machine learning methods

Davood Fereidooni*[1] and Zohre Karimi[2a]

[1]*School of Earth Sciences, Damghan University, Damghan, Iran*
[2]*School of Engineering, Damghan University, Damghan, Iran*

(*Received February 8, 2023, Revised August 17, 2023, Accepted August 23, 2023*)

Abstract. Brittleness as an important property of rock plays a crucial role both in the failure process of intact rock and rock mass response to excavation in engineering geological and geotechnical projects. Generally, rock brittleness indices are calculated from the mechanical properties of rocks such as uniaxial compressive strength, tensile strength and modulus of elasticity. These properties are generally determined from complicated, expensive and time-consuming tests in laboratory. For this reason, in the present research, an attempt has been made to predict the rock brittleness indices from simple, inexpensive, and quick laboratory test results namely dry unit weight, porosity, slake-durability index, P-wave velocity, Schmidt rebound hardness, and point load strength index using multiple linear regression, exponential regression, support vector machine (SVM) with various kernels, generating fuzzy inference system, and regression tree ensemble (RTE) with boosting framework. So, this could be considered as an innovation for the present research. For this purpose, the number of 39 rock samples including five igneous, twenty-six sedimentary, and eight metamorphic were collected from different regions of Iran. Mineralogical, physical and mechanical properties as well as five well known rock brittleness indices (i.e., B_1, B_2, B_3, B_4, and B_5) were measured for the selected rock samples before application of the above-mentioned machine learning techniques. The performance of the developed models was evaluated based on several statistical metrics such as mean square error, relative absolute error, root relative absolute error, determination coefficients, variance account for, mean absolute percentage error and standard deviation of the error. The comparison of the obtained results revealed that among the studied methods, SVM is the most suitable one for predicting B_1, B_2 and B_5, while RTE predicts B_3 and B_4 better than other methods.

Keywords: boosting regression trees; fuzzy inference; rock brittleness; rock sample; simple tests; support vector machine

1. Introduction

In the domains of rock mechanics and engineering geology, brittleness is a term generally used to identify the mechanical behavior and failure characteristics of rocks under loading and unloading conditions. This parameter has been utilized as a crucial property of rock for assessing rock burst and stability of underground projects, fatigue damage, penetrability, cuttability, drillability and sawability of rocks. Therefore, evaluation of the brittleness is a major challenge and plays an essential role in the above-mentioned fields.

To date, several different expressions have been proposed to quantify the brittleness of rocks by many researchers who have considered the brittleness as a combination of various rock properties, rather than any one specific property (Tarasov and Potvin 2013, Yang *et al*. 2020, Ghadernejad *et al.* 2020, Ye *et al.* 2020). However, the definitions of the brittleness are still ambiguous (Zhang *et al.* 2021), and there is no consensus as to which definition is the most suitable and reliable (Tarasov and Potvin 2013). At present, different brittleness definitions are presented by different researchers for this property of rocks based on their purposes. For instance, Morley (1944) and Hetenyi (1966) defined brittleness as the lack of ductility, Howell (1960) introduced it as a property of rock material rupture or fracture with small or no plastic flow, Obert and Duvall (1967) defined it as material failure by fracture at or only slightly beyond the yield stress, Ramsay (1967) defined it as the destruction of internal cohesion, Andreev (1995) defined it as the ability for a rock to deform and fail with a low degree of inelastic behavior, and Tarasov and Potvin (2013) stated that it is a self-sustaining failure process. Although there is no consensus on the definition of the brittleness, it is commonly believed that brittle rocks are characterized by the following features: (1) easily fractured with the formation of more cracks and (2) little plastic deformation under compression (Ye *et al*. 2020).

Jahandideh and Jafarpour (2016) stated that the influencing factors of rock brittleness are divided into internal factors and external factors. The intrinsic factors mainly include composition materials, texture, and porosity of the rock. The external factors comprise the ambient pressure and temperature. Considering the affecting factors, the most widely used quantification methods for evaluating rock brittleness are categorized by Ye *et al.* (2020) into

*Corresponding author, Ph.D.
E-mail: d.fereidooni@du.ac.ir
[a]Ph.D.
E-mail: z.karimi@du.ac.ir

ISSN: 2005-307X (Print), 2092-6219 (Online)

three general groups: (1) brittleness indices based on rock elastic parameters, (2) brittleness indices attained from rock mineral compositions, and (3) brittleness indices derived from rock stress-strain curves.

Many researchers have obtained some brittleness indices from different rock properties and investigated the relation between the parameters and other rock properties. In this regard, some brittleness indices were obtained from stress-strain curves (Baron 1962, Coates and Parsons 1966, Aubertin and Gill 1988, Aubertin *et al*. 1994, Ribacchi, 2000, Hajiabdolmajid and Kaiser 2003). Evans and Pomeroy (1966) theoretically showed that the impact energy of a cutter pick is inversely proportional to brittleness. Singh (1986) indicated that cuttability, penetrability, and the Protodyakonov strength index of coal strongly depend on the brittleness of coal. Kahraman (2002) statistically investigated the relationships between three different brittleness definitions for both drillability and borability using the raw data obtained from the experimental works of different researchers. Ghobadi and Naseri (2016) predicted rock brittleness from geomechanical properties of Hamekasi Limestone (Iran) using regression and artificial neural networks analyses and found that the geomechanical properties have various reliability for predicting the rock brittleness. Wood (2020) configured the optimized data-matching, correlation-free, machine-learning technique, and the transparent-open-box algorithm to predict mineralogical brittle index from a suite of well-log data derived from two published Lower Barnett Shale (Texas) wells that include core-derived mineralogical analysis. Ye *et al*. (2020) indicated that brittleness index is a key parameter used to identify the desirable fracturing intervals of shale gas reservoirs which its correlation with fracability is still controversial. Meng *et al*. (2020) in a comprehensive review article, stated that a number of mechanical responses and failure processes of rocks, are inherently influenced by rock brittleness to different extents. They mentioned that over the past 50 years, dozens of rock brittleness indices have been developed. Despite the widespread use of brittleness indices in different fields of rock engineering and engineering geology, both definition and measurement method of rock brittleness have been significantly differed and not yet been standardized, and most of the indices were proposed to cater for specific fields of application. Also, a comprehensive list of rock brittleness indices introduced by different researchers is presented in Zhang *et al*. (2021).

A simple rock brittleness index is the ratio of compressive strength to tensile strength ($B_1=\sigma_c/\sigma_t$) (see Eq. (1)). This index has been used and discussed in the literatures. Another rock brittleness index is B_2 which is obtained from three Utah coals (see Eq. (2)). Singh (1987) showed that a direct relationship exists between in-situ specific energy and B_2. Goktan (1991) stated that B_2 might not be a representative measure of rock cutting specific energy consumption.

Altindag (2000, 2002, 2003) found significant correlations between his proposed brittleness index (B_3) (see Eq. (3)) and the penetration rate of percussive drills, the drillability index in rotary drilling, and the specific energy in rock cutting. Kahraman and Altindag (2004) correlated fracture toughness values with different brittleness values using the raw data obtained from the experimental works of two researchers. They indicated that B_3 can be used as a predictive rock property for the estimation of the fracture toughness value. Kahraman *et al.* (2003a, b) found a strong correlation between Los Angeles abrasion loss and B_3 for 26 different rocks. Gunaydin et al. (2004) found a very strong correlation between hourly production and B_3 and they emphasized that B_3 is the most reliable index among the brittleness indexes adopted in their study. Yarali (2007) found a power relation with correlation coefficient of 0.86 between drilling rate index (DRI) and B_3 for 14 different rocks. Yilmaz *et al*. (2008) stated that the grain size seems to predominantly influence their relative brittleness index values in granitic rocks. Yarali and Soyer (2011) proposed another brittleness index namely B_4 (see Eq. (4)) which is calculated from uniaxial compressive and tensile strengths of rocks. was by Nejati and Moosavi (2017) proposed a brittleness index (B_5) obtained from elasticity modulus, uniaxial compressive strength and tensile strength of rocks (see Eq. (5)). This index has no physical meaning, and its positive relation with the fracture toughness (Meng *et al*. 2020).

As can be understand from above-mentioned literature, there are many rock brittleness indices obtained or predicted from different hard, expensive, and time-consuming rock properties. The successful of machine learning and soft computing methods in estimating some rock properties in recent years is the authors' motivation for exploiting these algorithms for predicting rock brittleness simple laboratory test. Cao *et al.* (2021) predicted Young's modulus and unconfined compressive strength of rock by combining extreme gradient boosting machine with the firefly algorithm. The feed forward neural network with three various optimization methods is utilized to predict Rock mass deformation modulus (Hasanipanah *et al*. 2022). The combination of three machine learning method including feed forward neural network, k-nearest neighbors and random forest in the stacking ensemble framework is exploited to predict rock deformations (Koopialipoor *et al.* 2022). An improved version of adaptive neuro-fuzzy inference system (ANFIS) model by artificial bee colony algorithm has been developed to predict the tunnel boring machine performance (Parsajoo 2021). Vipulanandan models have been used in various studies for different applications (e.g., Mahmood 2020, Mohammed 2019). In other study, the correlation model proposed by Vipulanandan has been employed to explicate the attributes of soils that have been subjected to alterations involving cement and polymer (Vipulanandan *et al.* 2018). Also, Vipulanandan *et al.* (2021) have developed a novel Vipulanandan failure model aimed at predicting the limit of maximum shear strength of rocks.

Certain researchers have established the potential of machine learning techniques to effectively predict rock brittleness indices. The combination of firefly and ANN algorithm is employed to predict some brittleness indices (Koopialipoor 2019). The support vector machine methods with various kernels are also explored for predicting brittleness indices from Schmidt hammer rebound number,

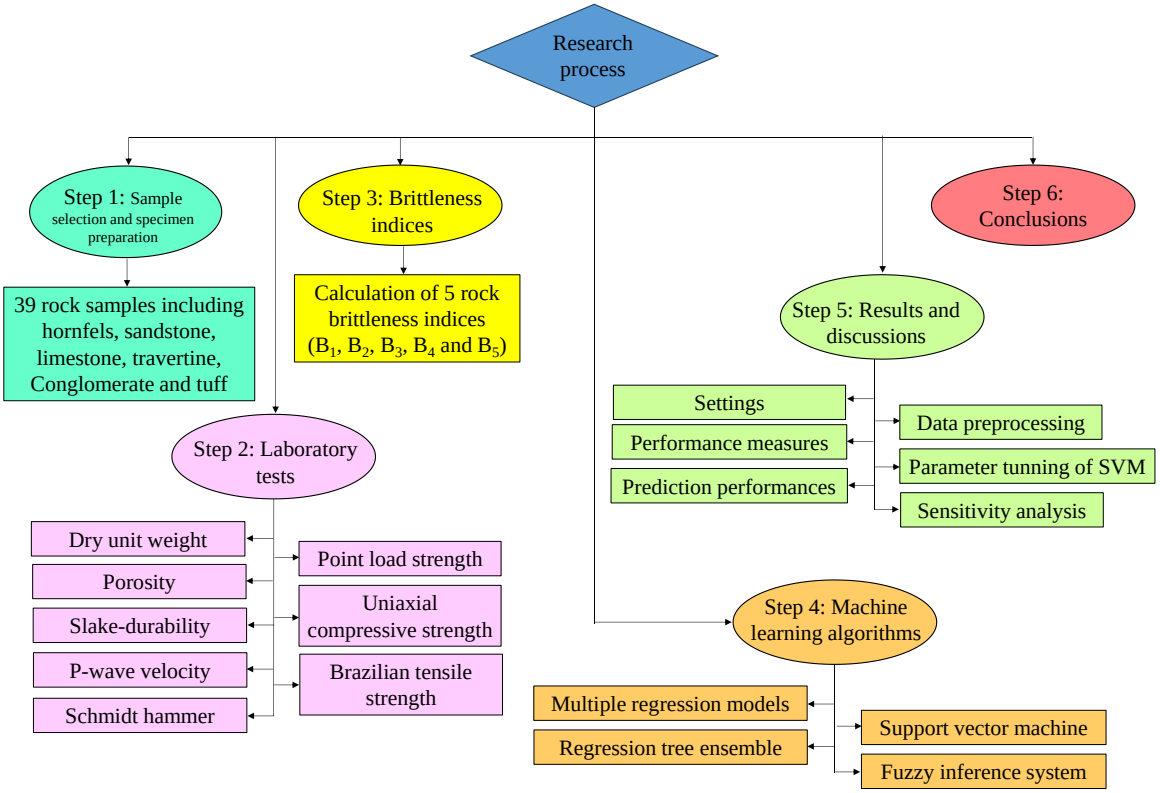


Fig. 1 Research process flowchart

p-wave velocity, point load strength index, and density (Jahed Armaghani *et al*. 2020). A differential evolution algorithm is employed to develop several linear and nonlinear models for predicting the brittleness indices (Yagiz *et al.* 2020).

Several machine learning algorithms including SVM, Chi-square automatic interaction detector, K-nearest neighbors, random forest (RF), and artificial neural network (ANN) were compared to predict the rock sample's brittleness index (Sun *et al.* 2020). Linear genetic programming and regression-based feature selection model has been also utilized for predicting brittleness indices (Jamei *et al*. 2022).

In the present research, an attempt has been made to predict five common rock brittleness indices namely B_1, B_2, B_3, B_4, and B_5 from simple laboratory test results such as unit weight, effective porosity, slake-durability index, P-wave velocity, Schmidt rebound hardness, and point load strength index using regression models, Support Vector Machine (SVM), fuzzy inference system (FIS) and regression tree ensemble (RTE). Therefore, the main aim of this research is to assess the accuracy of simple, inexpensive, and quick laboratory test results for predicting rock brittleness indices. This could be considered as an innovation for the present research. The regression models are selected as basic models for evaluating other models. FIS has been selected due to its ability to deal with uncertainty. SVM has achieved successful results in predicting various rock features and it is commonly resists to overfit and can provide high generalization performance. It can handle nonlinear relations between input and output by utilizing suitable kernels. RTE is an ensemble tree-based models and it is chosen for simplicity in the comparison of other machine learning models, and having relatively lower computational costs. Since each tree is built on a subset of inputs, it performs feature selection implicitly. So, it is very effective when the inputs are not of equal importance. In this study, various models that can potentially be suitable for predicting brittleness indices have been evaluated. This variety of models includes single and aggregate models, fuzzy and non-fuzzy models, linear and non-linear models. SVM's Hyper-parameter tunning is done in k-fold cross validation settings.

2. Materials and methods

2.1 Sample selection and specimen preparation

The number of 39 rock samples including five igneous (tuff and andesite tuff), twenty-six sedimentary (conglomerate, sandstone, limestone, and travertine), and eight metamorphic (hornfels) were collected from different regions of Iran. They are various in mineral content, color, luster, surface texture, and other apparent features. One or two blocks for each rock sample were collected from natural outcrops in different parts of Iran for the laboratory testing. The block samples were inspected for macroscopic defects to provide test specimens free from fractures, partings or alteration zones. Fig. 1 shows the flowchart of the research process along with its six performing steps.

Table 1 Types and location of the selected rock samples from different provinces of Iran

Sample No.	Rock type	Location	Province	Sample No.	Rock type	Location	Province
1	Hornfels	South west of Hamedan	Hamedan	21	Conglomerate	North west of Damghan	Semnan
2	Hornfels	South west of Hamedan	Hamedan	22	Conglomerate	North west of Damghan	Semnan
3	Hornfels	South of Hamedan	Hamedan	23	Conglomerate	North west of Damghan	Semnan
4	Hornfels	South east of Hamedan	Hamedan	24	Tuff	South east of Soltanieh	Zanjan
5	Hornfels	South west of Hamedan	Hamedan	25	Tuff	South east of Soltanieh	Zanjan
6	Hornfels	South west of Hamedan	Hamedan	26	Sandstone	Sojas	Zanjan
7	Hornfels	South west of Hamedan	Hamedan	27	Sandstone	Sojas	Zanjan
8	Hornfels	South west of Hamedan	Hamedan	28	Andesitic tuff	North east of Semnan	Semnan
9	Sandstone	North west of Damghan	Semnan	29	Andesitic tuff	North east of Semnan	Semnan
10	Sandstone	North west of Damghan	Semnan	30	Andesitic tuff	North east of Semnan	Semnan
11	Sandstone	North west of Damghan	Semnan	31	Limestone	South west of Damghan	Semnan
12	Dolomitic limestone	North west of Damghan	Semnan	32	Limestone	South west of Damghan	Semnan
13	Dolomitic limestone	North west of Damghan	Semnan	33	Limestone	South west of Damghan	Semnan
14	Dolomitic limestone	North west of Damghan	Semnan	34	Limestone	North west of Damghan	Semnan
15	Travertine	North west of Damghan	Semnan	35	Limestone	North west of Damghan	Semnan
16	Travertine	North west of Damghan	Semnan	36	Tuff	North west of Damghan	Semnan
17	Travertine	North west of Damghan	Semnan	37	Tuff	North west of Damghan	Semnan
18	Travertine	North west of Damghan	Semnan	38	Sandstone	North west of Damghan	Semnan
19	Travertine	North west of Damghan	Semnan	39	Sandstone	North west of Damghan	Semnan
20	Travertine	North west of Damghan	Semnan				

Table 1 shows types and location of the selected rock samples

2.2 Laboratory tests

For performing laboratory tests, required rock cores were provided by a coring machine. The diameters of the obtained rock core were 54 mm (NX). A total of 1170 rock specimens were prepared for all destructive and non-destructive tests. A comprehensive test plan was prepared for laboratory experiments (Fig. 2). One or two thin sections for each rock sample were conducted to investigate petrographic and mineralogical characteristics. Five laboratory cylindrical specimens with length to diameter ratio of 2-3 were used from each rock samples for determining physical properties including dry unit weight and porosity based on ISRM (2007).

To perform the slake-durability test, rock lumps (10 pieces of about 40–60 g for each rock sample) were prepared and rotated for 10 minutes in a test drum made of a standard sieve mesh so that slaking products would be finer than 2 mm and pass through the drum. The drum was half immersed in water at 20°C. The slake-durability index (Id) corresponding to each cycle was calculated as the percentage ratio of final to initial dry weights of rock lumps in the drum after the drying and wetting cycles. The slake-durability test was carried out in two cycles, and the results are presented in Table 2. Based on Gamble (1971) and ASTM (1990), the tested rocks were classified as durable and very durable.

The ultrasonic wave velocity was determined in the laboratory based on ISRM (2007) and ASTM (1996). This technique is often used to determine and characterize the dynamic properties of rocks. Since this method is nondestructive and relatively easy to apply, it is being increasingly used in geological and geotechnical engineering. This property depends on various parameters such as elastic properties, mineral content and orientation,

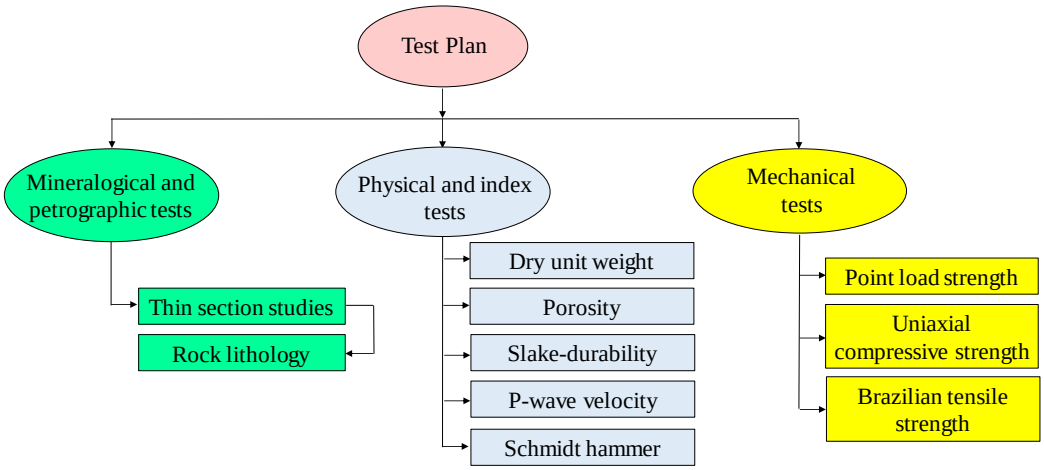


Fig. 2 Test plan for laboratory tests

Table 2 Average values of the physical, index, and mechanical properties of the rocks

Sample No.	γ_d (kN/m^3)	ne (%)	SDI (%)	v_p (m/s)	H_S	PLI (MPa)	UCS (MPa)	TS (MPa)	E (GPa)
1	27.96	2.00	99.22	5529	40.1	4.13	99.22	3.97	79.36
2	27.96	0.92	99.32	5065	49.1	6.88	125.21	5.61	65.70
3	27.57	0.97	99.40	5225	45.4	5.38	129.22	8.17	69.29
4	27.27	2.78	98.04	3911	40.5	4.00	95.93	4.84	36.17
5	27.47	0.56	98.58	4078	58.4	8.20	123.14	6.26	41.69
6	27.08	0.42	98.27	3873	59.7	11.08	140.83	8.18	37.62
7	26.49	0.49	97.41	3123	58.9	10.60	136.20	7.13	25.57
8	26.29	0.54	97.18	3015	53.7	7.08	118.01	6.11	23.38
9	24.82	7.58	97.25	3854	29.0	8.85	46.29	6.67	34.53
10	24.72	8.41	97.26	3517	26.0	6.54	32.46	5.38	28.88
11	24.13	10.32	97.04	3093	24.0	5.80	16.09	3.81	22.07
12	26.19	5.80	98.28	5171	36.0	10.32	50.13	7.93	64.04
13	25.51	7.83	98.30	4672	28.0	8.20	25.19	4.76	51.31
14	26.49	4.41	98.28	5284	36.0	11.86	51.51	9.22	67.51
15	23.05	2.80	98.97	5984	30.0	9.15	31.96	7.74	74.70
16	22.17	5.60	97.85	4198	20.0	6.28	19.37	5.08	36.33
17	22.86	4.82	98.53	5402	25.0	7.00	25.34	6.85	60.80
18	22.56	5.04	98.55	4789	29.0	8.61	28.35	7.18	47.60
19	23.45	4.71	98.68	4707	26.0	7.54	25.89	6.19	47.85
20	20.99	6.69	97.90	3427	20.0	6.40	18.52	5.16	23.34
21	25.21	5.44	97.35	3735	27.0	8.86	26.15	4.66	33.04
22	24.53	8.22	95.45	2931	18.0	5.16	14.97	2.97	20.24
23	25.21	6.98	96.21	3789	28.0	7.67	23.88	3.88	33.96
24	22.27	11.68	97.99	3446	40.1	3.15	73.07	9.06	25.02
25	22.27	11.42	97.62	3286	37.2	2.38	56.42	5.02	22.85
26	22.37	12.07	95.62	2096	28.6	2.21	40.55	2.96	9.73
27	22.46	10.69	96.91	2392	31.4	3.02	58.41	4.12	12.59
28	23.45	9.49	98.73	3658	36.9	3.61	57.22	4.27	32.16
29	23.25	9.76	99.02	3602	46.8	5.05	88.63	5.73	41.65
30	23.15	11.07	98.85	3521	42.5	3.98	71.88	4.46	36.72
31	26.19	1.08	98.08	5825	38.6	3.63	72.28	5.28	80.57
32	26.29	0.87	99.31	6030	30.2	2.53	38.13	3.50	86.46
33	26.29	0.98	99.19	6082	43.5	5.25	100.22	6.12	87.91
34	23.74	1.81	98.99	4913	45.2	5.75	104.00	7.01	73.00
35	24.23	1.76	99.34	4366	46.4	6.07	111.00	8.26	74.50
36	23.64	12.80	99.77	4061	41.9	4.95	86.40	5.82	68.40
37	23.84	13.10	99.57	4706	41.1	4.49	83.00	5.95	67.80
38	26.00	3.00	99.40	4061	45.8	5.62	103.50	6.79	39.98
39	26.00	4.00	99.40	4706	48.5	5.68	98.20	5.95	53.03

Note: in this table γ_d, n_e, SDI, v_p, H_S, PLI, TS, UCS, and E are dry unit weight, effective porosity, slake-durability index, P-wave velocity, Schmidt rebound hardness, point load index, uniaxial compressive strength, tensile strength, and elasticity modulus, respectively

Table 3 Values of the five indices for the rocks

Sample No.	B_1	B_2	B_3 (MPa)^2	B_4	B_5
1	24.99	0.92	196.95	73.91	11.29
2	22.32	0.91	351.21	112.09	13.05
3	15.82	0.88	527.86	150.30	18.26
4	19.82	0.90	232.15	83.19	8.99
5	19.67	0.90	385.43	119.85	11.39
6	17.22	0.89	575.99	160.05	13.15
7	19.10	0.90	485.55	141.52	9.69
8	19.31	0.90	360.52	114.22	8.38
9	6.94	0.75	154.38	62.02	13.40
10	6.03	0.72	87.32	41.15	11.00
11	4.22	0.62	30.65	19.36	8.32
12	6.32	0.73	198.77	74.40	20.88
13	5.29	0.68	59.95	31.39	14.03
14	5.59	0.70	237.46	84.56	24.20
15	4.13	0.61	123.69	52.87	24.32
16	3.81	0.58	49.20	27.22	13.13
17	3.70	0.57	86.79	40.97	20.75
18	3.95	0.60	101.78	45.95	18.61
19	4.18	0.61	80.13	38.68	16.79
20	3.59	0.56	47.78	26.66	10.72
21	5.61	0.70	60.93	31.76	10.93
22	5.04	0.67	22.23	15.37	6.55
23	6.15	0.72	46.33	26.07	9.68
24	8.07	0.78	331.01	107.40	13.36
25	11.24	0.84	141.61	58.28	8.20
26	13.70	0.86	60.01	31.41	3.65
27	14.18	0.87	120.32	51.83	5.09
28	13.40	0.86	122.16	52.40	8.50
29	15.47	0.88	253.92	88.74	11.32
30	16.12	0.88	160.29	63.72	8.99
31	13.69	0.86	190.82	72.24	15.44
32	10.89	0.83	66.73	33.90	12.96
33	16.38	0.88	306.67	101.66	17.07
34	14.84	0.87	364.52	115.13	17.26
35	13.44	0.86	458.43	135.79	19.75
36	14.85	0.87	251.42	88.11	14.85
37	13.95	0.87	246.93	86.97	15.19
38	15.24	0.88	351.38	112.13	12.38
39	16.50	0.89	292.15	98.17	12.93

density, porosity, presence of cracksand microfractures, and the degree of weathering (Goodman 1989). In this study, the average ultrasonic wave velocity was determined from five tests for each sample by a Pundit Lab Ultrasonic Pulse Velocity Tester (Proceq).

The Schmidt hammer test was performed according to ISRM (2007) and ASTM (2001a). In the current research, the Schmidt hammer was used on rock blocks with dimensions of about 30 × 40 × 50 cm Based on the results (Table 5).

The point load test method was standardized by ISRM (2007) and ASTM (2001b). In this test, cylindrical, prismatic, or irregular rock specimens that are loaded between two conical platens (of stipulated geometry and hardness) fail when they develop one or more extensional planes along the line of loading. The applied force (P) and the distance (D_e) between the platens at failure (or equivalent core diameter) were measured, and the point load index (I_s) was calculated, using the following equation (ISRM 2007, ASTM 2001b)

$$I_s = \frac{P}{D_e^2} \quad (1)$$

The point load index for a core diameter equal to 50 mm (PLI) was calculated from following expression (ISRM 2007, ASTM 2001c)

$$PLI = I_s \left(\frac{D_e}{50}\right)^2 \quad (2)$$

In this research, 10 tests were undertaken for each rock sample. The number of 390 cylindrical / lump specimens with length to diameter between 0.5 and one were prepared for point load strength tests.

Fig. 3 Application of various laboratory tests on the studied rocks

The uniaxial compressive strength tests were performed on trimmed core samples having a length-to-diameter ratio of 2.0 to 2.5 for determining uniaxial compressive strength (UCS) and elasticity modulus (E). The stress rate was applied within the limits of 1.0 MPa/s. In the current research, five prepared cores with a 2:3 length to diameter ratio were tested for each sample to determine UCS and E.

The tests were carried out according to ISRM (2007) and ASTM (1995) suggested method. The number of 195 cylindrical specimens with length to diameter between two and three were prepared for uniaxial compressive strength tests.

The Brazilian tensile strength (BTS) tests were conducted on core samples having a thickness-to-diameter ratio of 0.5 for determining tensile strength (TS) of the rocks. A loading rate of 200 N/s was applied. The test was repeated ten times for each rock type and the results were averaged. The tests were performed in accordance with ISRM (2007) suggested method. The BTS was determined using the following equation

$$BTS = \frac{2P}{\pi Dt} \tag{3}$$

where P is the maximum load recorded in the experiment, and D and t are diameter and thickness of the specimen, respectively. In this research, 10 tests were undertaken for each rock sample. The number of 390 disc-shaped specimens with length to diameter between 0.5 and 0.75 were prepared for Brazilian tensile strength tests. Application of the above-mentioned laboratory tests on the studied rocks are illustrated in Fig. 3. The average values of the results and their distribution curves and histograms are presented in Table 2 and Fig. 4, respectively. The number of bins in the histogram is computed according to Ishikawa formula (Ishikawa 1986).

2.3 Brittleness indices

In this study, the used brittleness concepts from the compressive and tensile strengths are as follows

$$B_1 = \frac{\sigma_c}{\sigma_t}\text{; (Hucka and Das 1974)} \tag{4}$$

$$B_2 = \frac{\sigma_c - \sigma_t}{\sigma_c + \sigma_t}\text{; (Hucka and Das 1974)} \tag{5}$$

$$B_3 = \frac{\sigma_c \times \sigma_t}{2}\text{; (Altindag 2010)} \tag{6}$$

$$B_4 = (\sigma_c \times \sigma_t)^{0.72}\text{; (Yarali and Soyer 2011)} \tag{7}$$

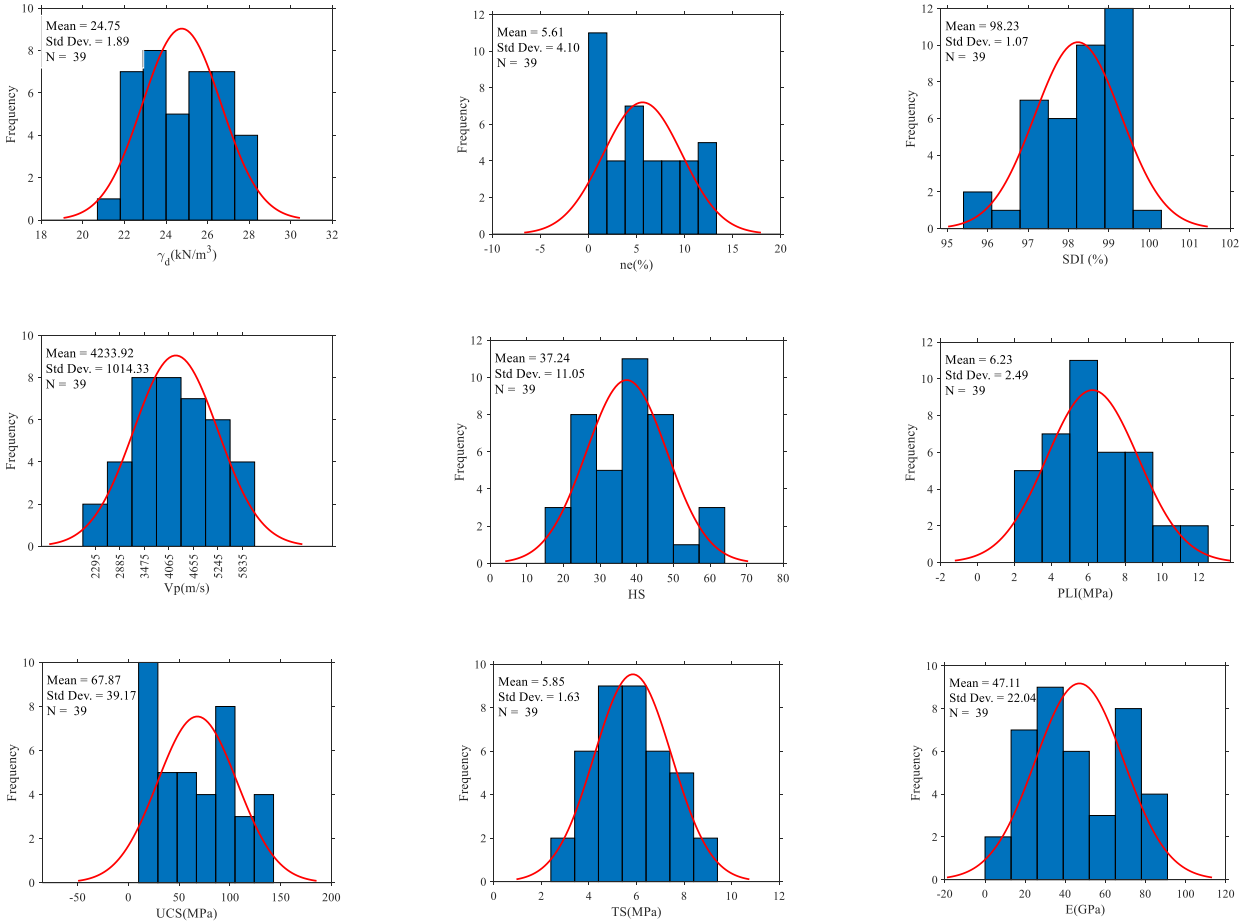


Fig. 4 Distribution curves and histograms of the rock properties in the database

$$B_5 = \frac{\sigma_t^{0.84} \times E^{0.51}}{\sigma_c^{0.21}}; \quad \text{(Nejati and Moosavi 2017)}, \qquad (8)$$

where B_1, B_2, B_3, B_4 and B_5 denote brittleness, σ_c is uniaxial compressive strength, σ_t is tensile strength of rock which is obtained from Brazilian test, and E is the elastic modulus of rock. The calculated values of the indices for the rocks are presented in Table 3.

2.4 Machine learning algorithms

In this study, the capability and applicability of regression analysis, SVM, RTE, and FIS are examined for predicting the five above-mentioned brittleness indices. SVM has a suitable generalization power when small data is available and can model nonlinear relations between input and output variables. RTE is an ensemble model in the comparison to other methods in this paper that learn only one model and it is a successful model in various applications. The FIS considers uncertainty and imperfectness in data by considering fuzzy sets instead of crisp sets. SVM, RTE, and FIS make no assumptions about the distribution of the data, as opposed to regression models that are constrained by the assumption of a distribution on the data. The details of these models are explained in this section.

a. Multiple Regression models

Multiple regression analysis is a statistical technique that investigates the relationships between multiple independent variables and one dependent variable (Tiryaki 2008). The relationship between inputs (independent variables) and outputs (dependent variables) is determined in the form of a predefined function. Based on the type of function, regression analysis is either linear or nonlinear. In the multiple linear regression, the relationship between output and inputs is determined as below linear function

$$y = b_0 + \sum_{i=1}^{D} b_i X_i \qquad (9)$$

where X_i is ith input and D is the dimension of input data (the number of independent variables).

In multiple nonlinear regression, the predefined function has a nonlinear form such as power, logarithmic and exponential forms. In this research, the exponential nonlinear equation is investigated as (Eq. (10)).

$$y = b_0 + b_{D+1} e^{\sum_{i=1}^{D} b_i X_i} \qquad (10)$$

In this study, the five mentioned brittleness indices are dependent variables, γ_d, n_e, Id, v_p, H_S, and PLI are

Table 4 Kernel's formula

Kernel	Description	Equation
RBF or Gaussian	Gaussian or Radial Basis Function (RBF) kernel	$G(x_i, x_j) = \exp(-\gamma \left\| x_i - x_j \right\|^2)$
Linear	Linear kernel	$G(x_i, x_j) = x_i' x_k$
Polynomial	Polynomial kernel, q specifies the order	$G(x_i, x_j) = (1 + \gamma x_i' x_j)^q$

Note: γ is the parameter of RBF and Polynomial kernels

independent variables, and D is equal to 5. Both multiple linear regression (REG_LIN) analysis and multiple nonlinear regression analysis with exponential functions (REG_EXP) have been carried out.

b. Support vector machine

The support vector machine is a supervised machine learning method that employs the structural risk minimization principle and statistical learning theory. It transforms input data into high dimensional (maybe infinite) feature space by utilizing kernel functions to learn a linear regression in the feature space. It is a popular machine learning method for regression and classification; in this paper, the focus is on regression.

Let we have m training samples $\{x_i, y_i\}_{i=1}^m$, where $x_i = < x_i^1, \dots x_i^D > \in \mathbb{R}^D$ and $y_i \in \mathbb{R}$ and the aim of SVM is to learn a nonlinear function $f(x)$ that is defined as (Eq. (11)).

$$f(x) = < \omega, \phi(x) > + b \tag{11}$$

where $<.,.>$ indicate the dot product, $\omega \in \mathbb{R}^{d_k}$ is the weight vector in primal weight space, and $\phi(.): \mathbb{R}^D \to \mathbb{R}^{d_k}$ is a function serving as a mapping from the input space to high dimensional feature space, where linear regression is performed in the feature space; b is the bias term, and d_k is the dimension of feature space that is implicitly defined (Smola and Schölkopf 2004). It does not need to explicitly define ϕ and the feature space, because the kernel function is used to generate dot product directly.

SVM attempts to find function $f(x)$ that violates from the true output by a value no greater than ε for each training point, and at the same time it is as flat as possible. The first condition is defined in ε-insensitive loss function formally as (Eq. (12))

$$|y - f(x,\omega)|_\varepsilon = \begin{cases} 0, & if \ \ |y - f(x,\omega)| \le \varepsilon \\ |y - f(x,\omega)| - \varepsilon, & otherwise \end{cases} \tag{12}$$

The ε-insensitive loss function incurs no penalty to errors within ε distance of observed value.

The optimization problem is defined as two terms; the first term vanishes the violation of the flatness of f and the second term penalizes the ε-insensitive loss function.

$$\min \tfrac{1}{2}\|\omega\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*)$$

$$\text{such that } \begin{cases} \forall i, \ 1 \le i \le m: & y_i - < \omega, \phi(x_i) > - b \le \varepsilon + \xi_i \\ \forall i, \ 1 \le i \le m: & < \omega, \phi(x_i) > + b - y_i \le \varepsilon + \xi_i^* \\ \forall i, \ 1 \le i \le m: & \xi_i, \xi_i^* \ge 0 \end{cases} \tag{13}$$

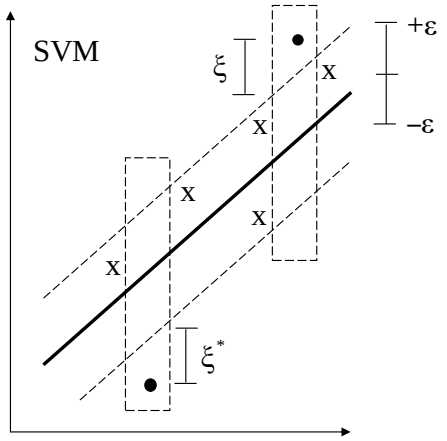


Fig. 5 Schematic of linear SVM for regression

where ξ and ξ^* are slack variables and ε is the approximation accuracy that can be violated utilizing slack variables. $C > 0$ is a constant that determines the trade-off between flatness and loss function. The schematic of linear SVM for regression is shown in Fig. 5. The optimization problem can be solved using sequential minimal optimization. For more details about this method, please refer to Chen *et al.* (2006)

Different types of kernels may be used to transform data to high dimensional space; linear kernel, polynomial and radial basis function (RBF) are important examples. The kernel's formula for the mentioned kernels is given in Table 4.

c. Regression tree ensemble

Regression trees are nonparametric methods that achieve suitable results. They attempt to learn a tree by minimizing the sum of node impurities in child nodes by considering various measures for impurity. RTE combine results from several learners to maximize the generalization accuracy using a weighted average of results obtained from applying some trees. Ensemble learning methods have been widely used due to their remarkable generalized performance and previous research has verified that an ensemble is often more accurate than any single base learners in the ensemble (Zhou 2019). Two main frameworks for combining several learners are bagging and boosting. This paper deals with boosting for regression problems, where base predictors are CART (Classification and Regression Tree) model (Loh 2011). Boosting framework combines the results of several predictors and has great promise. It has been utilized successfully in many

domains (Shin 2015). It is an iterative procedure that learns a regression tree at each step. As shown in Fig. 6, it attempts to boost the performance of any given models by combing them to achieve better generalization. It has two essential specifications; The first is about the samples used as the training set at each step. At iteration k, the weight of incorrectly predicted cases from a given phase is increased during the next phase. The second is about the aggregation part of the algorithm, where the final predictor considers the resampling distribution by weighting the predicted outputs of all individual trees. Boosting regression tree approximates the function $f(x)$ as (Eq. (14))

$$f(x) = \sum_{t=1}^{T} \alpha_t h_t(x) \tag{14}$$

where $h_t(x)$ is the t th estimated function by t th base learner, α_t is the weight of tth learner, and T is the number of base learners. As mentioned earlier, the base learners in this study are CART regression trees, which probably is the most popular regression tree in the literature.

d. Fuzzy inference system

A FIS is an artificial intelligence modeling technique that utilizes the available high-level prior knowledge to build a set of rules using fuzzy logic to detect, predict or classify data. The fuzzy approaches provide a natural way to resolve the problems in uncertain environment employing fuzzy inference engines.

A FIS is an artificial intelligence modeling technique that utilizes the available high-level prior knowledge to build a set of rules using fuzzy logic to detect, predict or classify data. The fuzzy approaches provide a natural way to resolve the problems in uncertain environment employing fuzzy inference engines.

Two main types of FIS are Mamdani (Mamdani and Assilian 1975) and Sugeno (Sugeno 1985) FISs. The Sugeno FISs accurately model highly nonlinear systems (Jang and Mizutani 1997) and are considered in this paper. For Sugeno FISs, the form of the ith rule is as follows:

$$R_i: if\ X_1\ is\ A_{i1} and\ X_2\ is\ A_{i2}\ and\ \dots X_D\ is\ A_{iD}\ \ Then\ y_i = f_i(X_1, X_2, \dots, X_D)$$

Where $X_1,\dots, X_D$ are input attributes and $A_{i1},\dots, A_{iD}$ are the fuzzy sets. The output form of a Sugeno FIS is typically a function of the input vector of X. This function can be any type of function that presents the output within the fuzzy space stated by the rule antecedent; but it is polynomial in the typical form.

In this paper, the clustering algorithm of fuzzy c-means (FCM) (Li *et al*. 2012, Bezdek, 2013) is utilized to extract rules from a set of training data for the FIS. FCM allows each data point to belong to multiple clustering with varying degree of membership. It attempts to minimize the optimization problem in (Eq.(15)).

$$J_M = \sum_{i=1}^{m}\sum_{j=1}^{C} u_{ij}^{M} \left\|x_i - c_j\right\|^2 \tag{15}$$

where x_i is ith data and c_j is the center of the jth cluster, M indicates any real number that is greater than one and u_{ij} is the membership degree of x_i within the j th cluster. $\|*\|$ indicates any norm that states the similarity between the cluster center and data and C is the number of clusters. The minimization of (Eq. (15)) is carried out by iteratively updating membership degree, u_{ij}, and cluster center, c_j, respectively, by (Eqs. (16) and (17))

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_k\right\|}\right)^{\frac{2}{M-1}}} \tag{16}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_k\right\|}\right)^{\frac{2}{M-1}}} \tag{17}$$

The updating process is stopped when the condition in (Eq. (18)) is satisfied.

$$\max_{ij}\left\{\left|u_{ij}^{(k+1)} - u_{ij}^{(k)}\right|\right\} < \epsilon \tag{18}$$

where ϵ indicates a stopping criterion and has a value between 0 and 1 and k stands for the iteration steps. In this study, the standard function in Matlab, GENFIS, is utilized to produce a FIS by applying FCM clustering to both inputs and output datasets.

3. Results and discussions

3.1 Settings

k-fold cross-validation method is exploited to assess the generalization performance of models. It is a promising procedure of evaluating machine learning methods (Ramezan *et al*. 2019). In the first step of k-fold cross-validation, the instances are randomly divided the dataset into k parts with approximately equal size. Then, one part is utilized as test data, while the remaining parts are employed as training data. This procedure is repeated k times, each time using a different partition as the test data. Therefore, each sample is used as test data exactly once. The 3-fold cross-validation is illustrated in Fig. 7

In this study, all experiments have been carried out on 3-fold cross-validation for a fair comparison. The parameters for applied models are set as below:

Multiple Regression: For nonlinear regression, the initial value of both b_0 and b_{D+1} is set to $\max(y)$, and the initial value of $b_d (1 \le d \le D)$ is set to $\frac{\max(\mathrm{X_d})}{10}$

RTE: The number of trees is in {150, 200, 250}, learning rate is in {0.1, 0.25, 0.5, 0.75, 1}, maximum number of decision splits (or branch nodes) per tree is set to $\lfloor log2(m-1) \rfloor$.

GENFIS: Input membership function for each fuzzy cluster is set to be Gaussian membership function.

The selection of parameters is an important step for predicting with SVM. In this research, various combinations of parameters are investigated and the results are detailed in section 3.4.

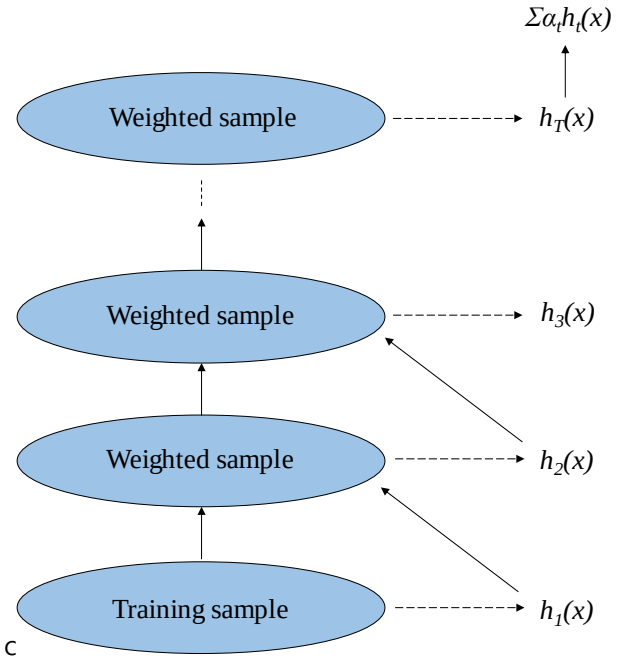


Fig. 6 Schematic of boosting procedure (Adapted from Shin 2015)

	Fold 1	Fold 2	Fold 3
Case 1			
Case 2			
Case 3			

Fig. 7 Graphical representation of k-fold cross-validation

3.2 Data preprocessing

Before applying machine learning methods, some preprocessing steps should be carried out to achieve accurate results. At the first step, the input data are normalized to avoid the model from being affected by the variables that vary over different ranges. For every input of x, min-max normalization is applied using Equation 19 which scales the range of x in [0,1].

$$z_i^d = \frac{x_i^d - x_{\min}^d}{x_{max}^d - x_{min}^d}, (1 \le i \le n; 1 \le d \le D) \quad (19)$$

where x_i^d is d th dimension of x_i, x_{min}^d and x_{max}^d are minimum and maximum value of dth dimension of inputs, respectively and z_i^d is the corresponding scaled value. At the second step, outliers should be detected and removed. Outliers are data that have different behaviors from other data. For each feature, an outlier is defined as a value that is exceeded the interval between median plus or minus three times of the median absolute deviations (MAD), where MAD for dth dimension is computed as (Eq. (20))

$$MAD = median(|x_i^d - median(x^d)|) \quad (20)$$

We compute MAD for all input features, and the results indicated that there are no outliers in the data. Thus, there is no need to eliminate any value in the dataset.

3.3 Performance measures

The performance of applied methods is assessed in terms of mean square error (MSE), relative absolute error (RAE), root relative absolute error (RRSE), determination coefficients (R^2) , variance account for (VAF), mean absolute percentage error (MAPE), and standard deviation of the error (ErrStd). They calculated as follows

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{N} \quad (21)$$

$$RAE = \frac{\sum_{i=1}^{n}|y - \hat{y}_i|}{\sum_{i=1}^{n}|y_i - \bar{y}|} \quad (22)$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_I)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (23)$$

$$VAF = (1 - \frac{var(y - \hat{y})}{var(y)}) \quad (24)$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (25)$$

$$ErrStd = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}} \quad (26)$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \quad (27)$$

Table 6 Results of 3-fold cross-validation and the best parameter settings with RBF kernel

Brittleness index	Fold	Best ε	Best C	Best γ	$MSE_{opt.}$	$R^2_{opt.}$
B_1	1st	0.0100	27.8701	2.3286	1.6284	0.9359
	2nd	0.0100	26.4558	1.1799	2.2090	0.9299
	3rd	0.0100	25.0416	2.3286	16.4080	0.7696
		Average of the best results			6.7485	0.8785
B_2	1st	0.0100	30.6985	10.3695	0.0020	0.8648
	2nd	0.0100	17.9706	3.4773	0.0016	0.8874
	3rd	0.0100	23.6274	3.4773	0.0014	0.8994
		Average of the best results			0.0017	0.8839
B_3	1st	0.0100	30.6985	1.1799	34188.8880	0.8414
	2nd	0.0100	30.6985	1.1799	4802.7089	0.8347
	3rd	0.0100	30.6985	1.1799	6020.8465	0.5845
		Average of the best results			15004.1478	0.7536
B_4	1st	0.4100	13.7279	1.1799	1579.1988	0.8836
	2nd	0.9600	30.6985	1.1799	124.6361	0.8496
	3rd	0.4600	30.6985	1.1799	226.7464	0.7386
		Average of the best results			643.5271	0.8240
B_5	1st	0.9100	6.6569	1.1799	5.9977	0.7015
	2nd	0.0100	30.6985	1.1799	3.5793	0.9741
	3rd	0.0100	30.6985	1.1799	3.8984	0.6933
		Average of the best results			4.4918	0.7896

Table 5 The range and step size of SVM parameters

Parameter	Range	Step size
C	$[2^0\ 2^5]$	$2^{0.5}$
γ	$[2^{-5}\ 2^5]$	$2^{0.2}$
ϵ	[0.01 1]	0.05

where y indicates a vector of measured values, $\hat{y}$ denotes a vector of estimated values, $\bar{y}$ is the mean of y, var shows the variance and n is the number of data points. R^2 evaluates the linear relation between observed and predicted data. In the best case, the predicted values exactly match the observed values, which results in VAF= 1, MAPE=0, and R^2=1.

3.4 Parameter tunning of SVM

Parameter tunning plays a crucial part in ensuring and improving the performance of SVM classifier. The parameters C and ϵ are provided as input to the SVM and affect the optimization process; The parameter C specifies the trade-off between training error and VC (Vapnik–Chervonenkis) dimension and the parameter ϵ specifies the insensitivity zone in ϵ-insensitive loss function. Applying Gaussian and polynomial kernel functions, γ is another important parameter that controls the sensitivity of the kernel function. The problem of SVM parameter tunning is still of much interest. In order to find out the optimum values of the parameters, we investigate wide range of combinations of parameters C, ϵ and γ, as shown in Table 5.

SVM with the specified parameter settings is performed and finally, the best parameter based on the following fitness function is picked

$$Fitness = Max(R^2 + \frac{1}{MSE}) \quad (27)$$

The results of 3-fold cross-validation and the best parameter settings for SVM with RBF kernel are presented in Table 6. For more clarity, the results of cross-validation 3D surface plots of MSE versus $\log_2 C$ and $\log_2 \gamma$ are illustrated in Fig. 8. As it is seen, the best fitness for B_1 is obtained when C and γ are approximately are in the range (25, 28) and (1.1, 2.4), respectively.

The parameter tunning is also investigated for SVM with Polynomial kernel. The best obtained parameter settings and the results of 3-fold cross-validation are presented in Table 7. The values of MSE versus $\log_2 C$ and $\log_2 \gamma$ in three folds are plotted in Fig. 9. The value of degree of polynomial kernel, q, in our experiments is set to 3.

For linear kernel, the best parameter values are reported in Table 8. As shown the results of a wide range of experiments, the acceptable values of MSE and R^2 are obtained in different settings with various kernels.

3.5 Prediction performances

We have conducted some experiments to assess the capability of the various models in forecasting brittleness indices.

The prediction performance results of various models for predicting brittleness indices are presented in Table 9.

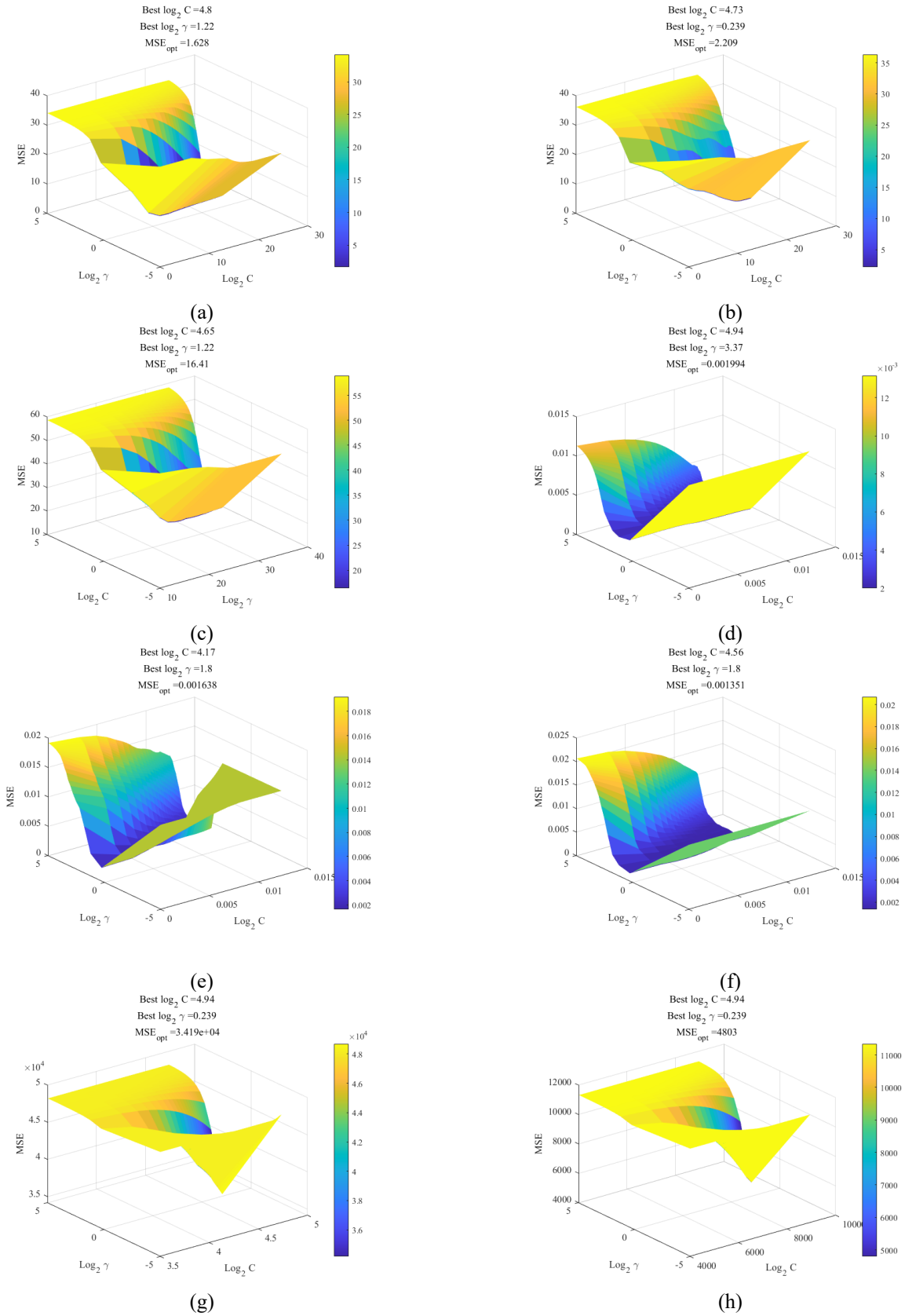


Fig. 8 3D view of MSE of predicting Brittleness indices by SVM using RBF kernel versus $\log_2 C$ and $\log_2 \gamma$. (a), (b) and (c), 1st, 2nd fold and 3rd folds of predicting B_1, respectively. (d), (e) and (f), 1st, 2nd fold and 3rd folds of predicting B_2, respectively. (g), (h) and (i), 1st, 2nd fold and 3rd folds of predicting B_3, respectively. (j), (k) and (l), 1st, 2nd fold and 3rd folds of predicting B_4, respectively. (m), (n) and (o), 1st, 2nd fold and 3rd folds of predicting B_5, respectively

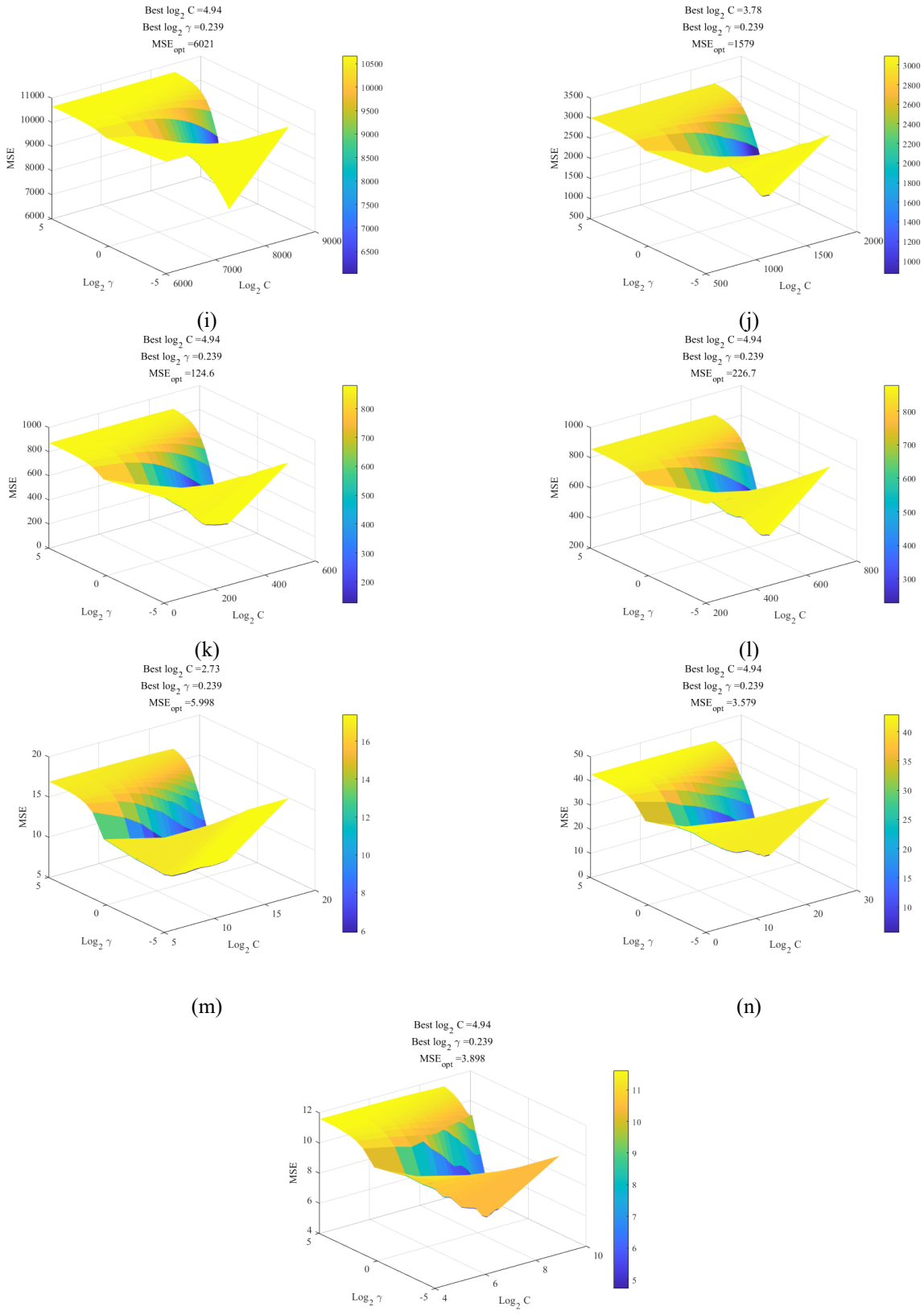


Fig. 8 Continued-

The Best and second-best results in each measure are highlighted in boldface and underlined, respectively. The results signify the effectiveness of the SVM_LIN model over other models in estimating B_1 in the terms of MSE,

Table 7 Results of 3-fold cross-validation and the best parameter settings of SVM with Polynomial kernel

Brittleness index	Fold	Best ε	Best C	Best γ	$MSE_{opt.}$	$R^2_{opt.}$
B_1	1st	0.0100	29.2843	3.4773	1.6090	0.9378
	2nd	0.0100	30.6985	2.3286	2.1433	0.9409
	3rd	0.7600	1.0000	1.1799	15.4708	0.7950
		Average of the best results			6.4077	0.8912
B_2	1st	0.0100	25.0416	11.5182	0.0020	0.8653
	2nd	0.0100	27.8701	5.7747	0.0016	0.8924
	3rd	0.0100	13.7279	2.3286	0.0013	0.9095
		Average of the best results			0.0016	0.8891
B_3	1st	0.4100	3.8284	1.1799	25668.6510	0.8566
	2nd	0.8100	30.6985	1.1799	1784.5492	0.9014
	3rd	0.9100	30.6985	1.1799	3801.3616	0.6919
		Average of the best results			10418.1873	0.8166
B_4	1st	0.9100	1.0000	1.1799	1540.8353	0.8817
	2nd	0.8100	12.3137	1.1799	129.4964	0.9161
	3rd	0.3100	29.2843	1.1799	239.6169	0.7451
		Average of the best results			636.6495	0.8476
B_5	1st	0.5600	15.1421	2.3286	5.4313	0.6940
	2nd	0.9600	30.6985	1.1799	2.1637	0.9676
	3rd	0.3100	6.6569	1.1799	3.6822	0.6959
		Average of the best results			3.7591	0.7859

Table 8 Results of 3-fold cross-validation and the best parameter settings with linear kernel

Brittleness index	Fold	Best ε	Best C	$MSE_{opt.}$	$R^2_{opt.}$
B_1	1st	1.5449	0.9386	1.5449	0.9386
	2nd	1.5085	0.9526	1.5085	0.9526
	3rd	16.0854	0.7733	16.0854	0.7733
		Average of the best results		6.3796	0.8882
B_2	1st	0.0100	1.0000	0.0021	0.8503
	2nd	0.0100	3.8284	0.0015	0.9008
	3rd	0.0100	5.2426	0.0014	0.8934
		Average of the best results		0.0017	0.8815
B_3	1st	0.0100	30.6985	34743.7396	0.8363
	2nd	0.0100	30.6985	4929.6114	0.7717
	3rd	0.0600	30.6985	6419.1351	0.5616
		Average of the best results		15364.1620	0.7232
B_4	1st	0.0100	16.5563	1455.5771	0.8783
	2nd	0.5600	30.6985	155.2238	0.8181
	3rd	0.9600	30.6985	253.1863	0.7126
		Average of the best results		621.3291	0.8030
B_5	1st	0.6600	15.1421	5.3652	0.6980
	2nd	0.7100	29.2843	5.9055	0.9494
	3rd	0.3600	30.6985	3.8359	0.6372
		Average of the best results		5.0355	0.7615

RAE, RRSE, and MAPE. Also, it is observed that SVM_POL have better VAF and R2 values as compared to other models. According to Table 9, the statistical performance criteria confirm the superiority of SVM_POL over the other models in forecasting B_2. RTE achieves the best results in the comparison of other models in predicting both B_3 and B_4. Moreover, Table 9 indicates the superiority of SVM_POL and SVM_RBF in most evaluation criteria

Table 9 Prediction performance measures of various models for estimating B_1, B_2, B_3, B_4, and B_5

Brittleness index	Measure/ Model	REG_LIN	REG_EXP	SVM_LIN	SVM_POL	SVM_RBF	RTE	GENFIS
B_1	MSE	7.0408	7.9494	**6.3796**	6.4077	6.7485	7.8114	10.0130
	RAE	0.3498	0.3681	**0.2830**	0.2912	0.3032	0.3774	0.4192
	RRSE	0.4389	0.4664	**0.4178**	0.4187	0.4297	0.4623	0.5234
	VAF	0.8080	0.7827	0.8288	**0.8323**	0.8185	0.7893	0.7274
	R^2(%)	80.8890	78.8366	83.6692	**84.3330**	82.8313	78.9844	73.8145
	ErrStd	1.3734	1.4088	**1.2353**	1.2530	0.3218	1.4264	1.5035
	MAPE(%)	19.1395	17.7034	**14.0385**	14.6775	15.3632	19.9841	26.135
B_2	MSE	0.0018	0.0018	0.0017	**0.0016**	0.0017	0.0020	0.0022
	RAE	0.3201	0.3222	0.2981	**0.2899**	0.2927	0.3260	0.3396
	RRSE	0.3693	0.3710	0.3541	**0.3466**	0.3521	0.3825	0.4041
	VAF	0.8639	0.8626	0.8747	**0.8806**	0.8763	0.8538	0.8389
	R^2(%)	86.3948	86.2629	87.4723	**88.0597**	87.6557	85.3923	85.1768
	ErrStd	0.1828	0.1834	0.1764	**0.1740**	0.1748	0.1845	0.1883
	MAPE(%)	4.4281	4.4539	4.1127	**3.9917**	4.0453	4.4527	4.4641
B_3	MSE	5374.4014	6972.9893	15364.1620	10418	15004.1478	**4672.0745**	8514.9592
	RAE	0.4319	0.4640	0.7560	0.5611	0.7463	**0.3980**	0.5544
	RRSE	0.4954	0.5643	0.8377	0.6898	0.8278	**0.4619**	0.6236
	VAF	0.7570	0.6820	0.3494	0.5380	0.3606	**0.7866**	0.6304
	R2(%)	75.8728	69.7023	60.8803	0.5414	61.1017	**78.6973**	70.8960
	ErrStd	7.3347	7.6022	9.7037	8.3597	9.6415	**7.0411**	8.3098
	MAPE(%)	38.0678	34.4512	71.9117	47.2345	70.9166	**37.8091**	45.2176
B_4	MSE	317.7183	352.6731	621.3291	636.6495	643.5271	**296.1729**	568.7602
	RAE	0.3986	0.4093	0.5486	0.5441	0.5446	**0.3954**	0.5399
	RRSE	0.4566	0.4810	0.6385	0.6463	0.6498	**0.4408**	0.6109
	VAF	0.7930	0.7690	0.6162	0.5887	0.6016	**0.8075**	0.6465
	R^2(%)	79.3931	77.6592	65.7831	59.1944	64.5274	**80.8082**	72.5591
	ErrStd	3.6433	3.6915	4.2740	4.2563	4.2581	**3.6284**	4.2400
	MAPE(%)	21.1144	21.078	31.3301	30.7349	30.8448	**20.7656**	26.2821
B_5	MSE	5.5697	4.7380	5.0355	**3.7591**	4.4918	6.0158	8.0857
	RAE	0.4767	0.4322	0.4533	**0.3952**	0.4337	0.5049	0.6136
	RRSE	0.4918	0.4536	0.4676	**0.4040**	0.4416	0.5111	0.5925
	VAF	0.7595	0.7945	0.7853	**0.8368**	0.8050	0.7388	0.6573
	R^2(%)	75.9892	79.4538	81.6469	**84.1274**	82.2525	73.8852	66.2104
	ErrStd	1.3368	1.2729	1.3036	**1.2171**	1.2751	1.3758	1.5167
	MAPE(%)	13.4001	**12.0108**	12.7238	12.7176	13.1138	15.0793	18.3903

Note: in this table, the first and second better results are shown by bold and underline states, respectively

over the other models in predicting B_5. These results indicate the good level of accuracy for predicting brittleness indices exploiting machine learning methods. For more clarity, the error bar of various models in 3 folds have been illustrated in Fig. 10. As it shown, the best models in predicting brittleness indices have an acceptable standard deviation compared to other models, except for the SVM model in predicting B_1, which is not robust enough. With a detailed examination of the results, it is found that the prediction error in two folds is much lower than the reported average, and only the high error in the third fold has caused an increase in the average and standard deviation of the error. Therefore, with the increase in the number of samples, it is expected that the average and standard deviation of the error will decrease. Besides, the detailed results indicate the superiority of SVM_POL over other models in predicting B_5.

Fig. 11 illustrates the results of proposed predictive models in addition to observed values. It appears that the predicted values are close to measured ones in different methods.

In order to validate the prediction models suggested in this paper, the scatter plot, R^2, and the linear relation between predicted and observed values are shown in Figs. 12-16. The distance of each data point from the diagonal line of 1: 1, indicates the error of the predicted values.

According to the obtained results, it was concluded that the SVM model can outperform other developed models to

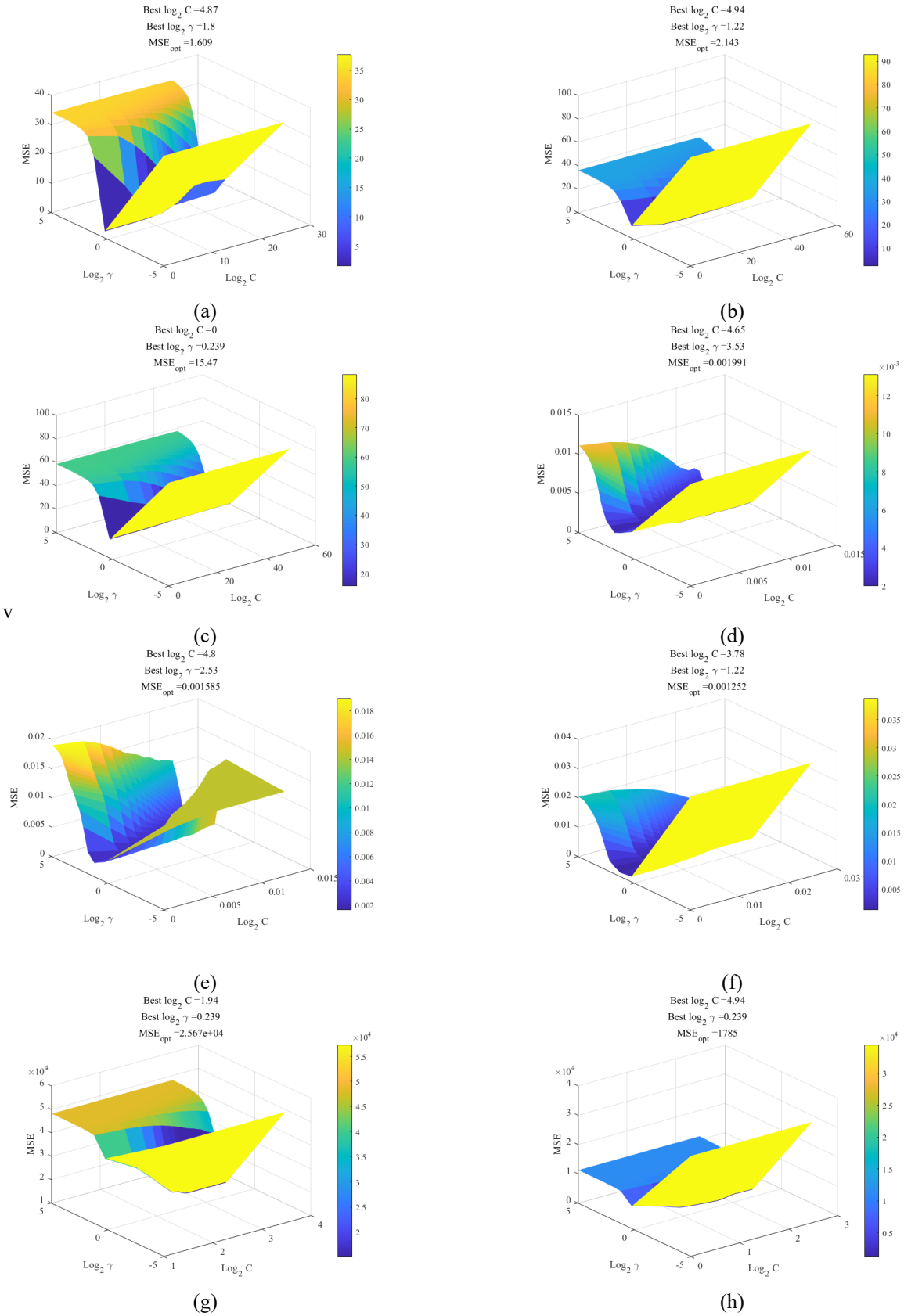
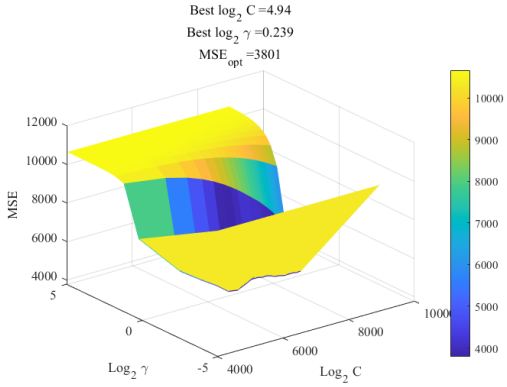
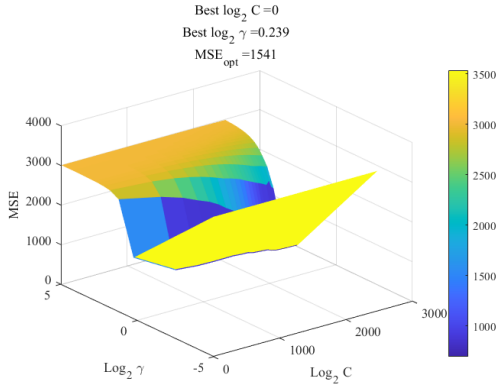


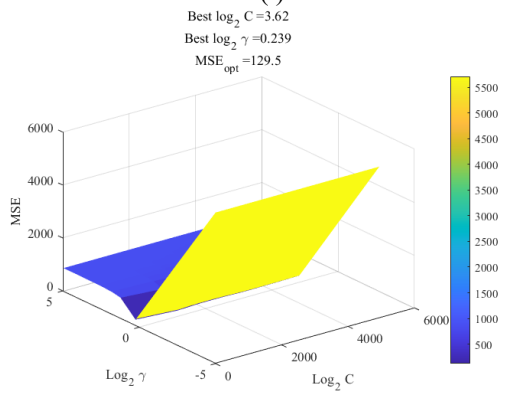
Fig. 9 3D view of MSE of predicting Brittleness indices by SVM using Polynomial kernel versus $\log_2 C$ and $\log_2 \gamma$. (a), (b) and (c), 1st, 2nd fold and 3rd folds of predicting B_1, respectively. (d), (e) and (f), 1st, 2nd fold and 3rd folds of predicting B_2, respectively. (g), (h) and (i), 1st, 2nd fold and 3rd folds of predicting B_3, respectively. (j), (k) and (l), 1st, 2nd fold and 3rd folds of predicting B_4, respectively. The last row: (m), (n) and (o), 1st, 2nd fold and 3rd folds of predicting B_5, respectively



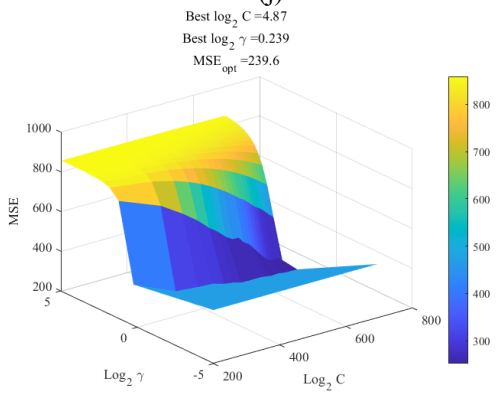
(i)



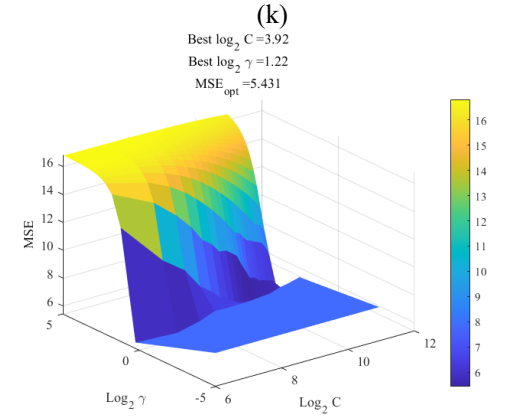
(j)



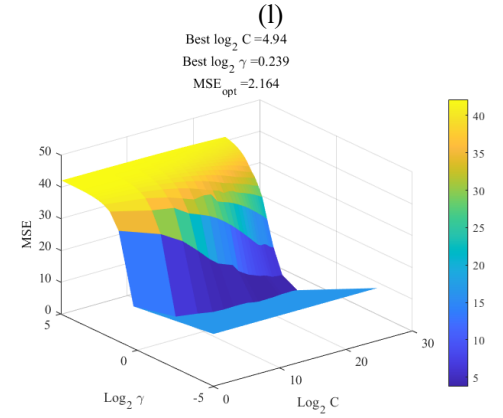
(k)



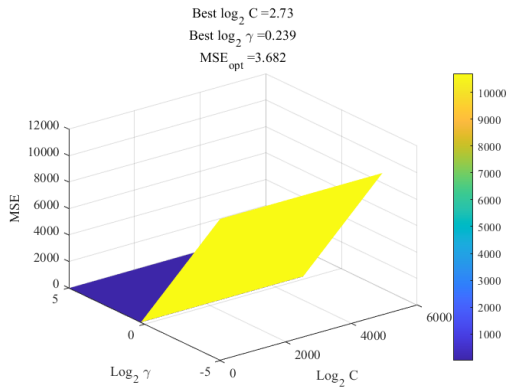
(l)



(m)



(n)



(o)

Fig. 9 Continued-

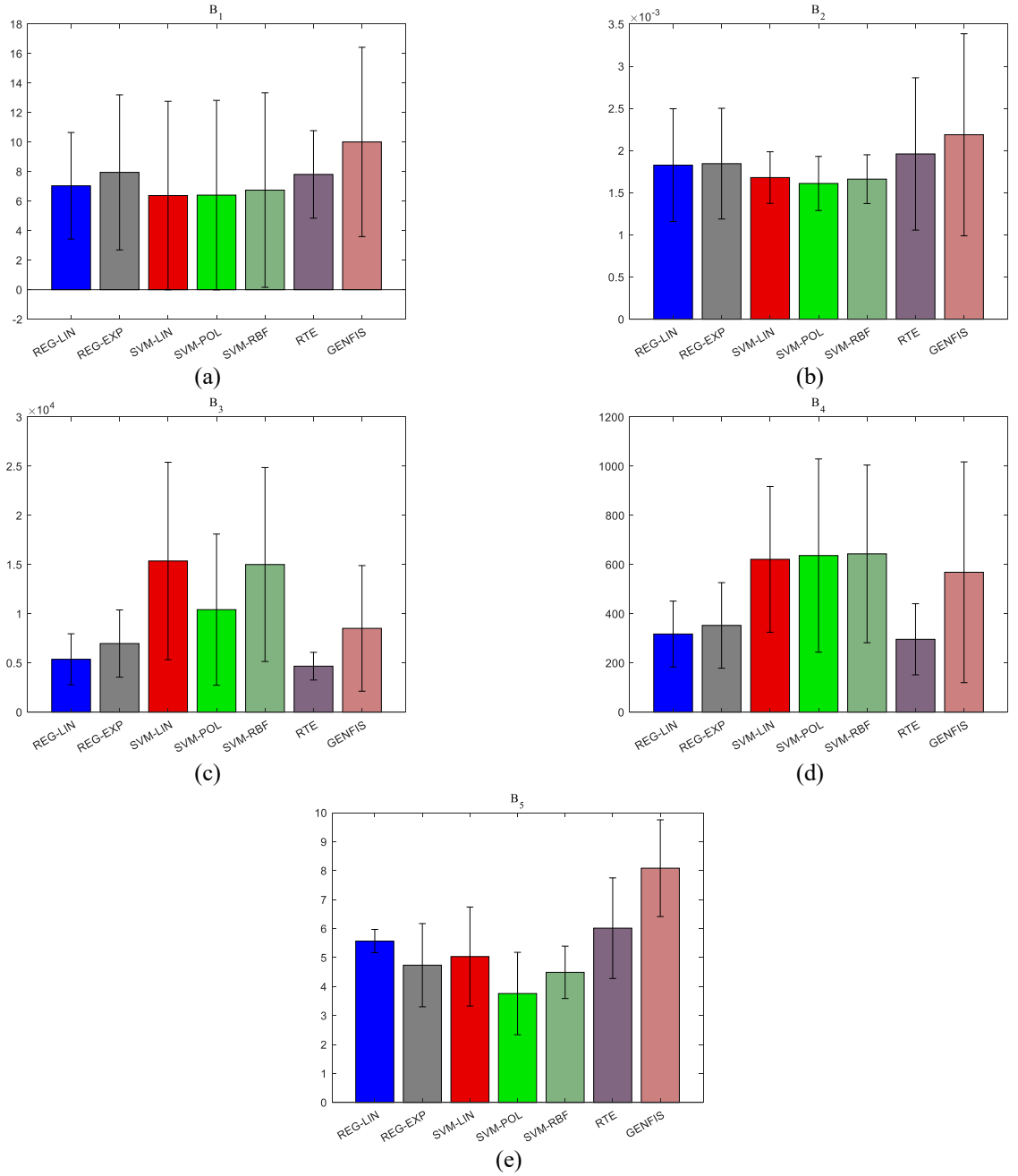


Fig. 10 Error bars of various models for predicting (a) B_1, (b) B_2, (c) B_3, (d) B_4, and (e) B_5 in 3 folds

predict both B_1,B_2 and B_5 on the premise of both low error and high correlation. However, as expected, the type of kernel has an essential role in the performance measures; polynomial kernel and linear kernels achieve the best results in predicting B_1 (polynomial: R^2=0.8433 and MSE=6.4077, linear: R^2=0.8367 MSE=6.3796); polynomial and RBF kernels achieve the best results in predicting B_5 (polynomial: R^2=0.8413 and MSE=3.7591, RBF: R^2=0.8225 and MSE=4.4918) and polynomial kernel has better results than other methods in predicting B_2 (R^2=0.8806 and MSE=0.0016). In addition, boosting regression trees achieved the most accurate values in forecasting both B_3 (R_2=0.7870 and MSE=4672.0745), and B_4 (R^2=0.8081 and MSE=296.1729).

Consequently, REG_LIN can predict all brittleness indices in an acceptable precision rate, and it has second best results in predicting B_3 and B_4. However, it can be found that an applied machine learning method possessed superior predictive ability than the other models in forecasting all brittleness indices in this study.

For more analysis, the above experiments for the best two obtained models have been repeated 10 times (for each 3-fold cross validation setting). By regarding that some models such as SVM and RTE have some random steps in the training phase, these experiments explore the stability of the obtained results with various seed values. The error bars are plotted in Figure 17. The second-best model for predicting B_3 and B_4 is REG_LIN that has no any random

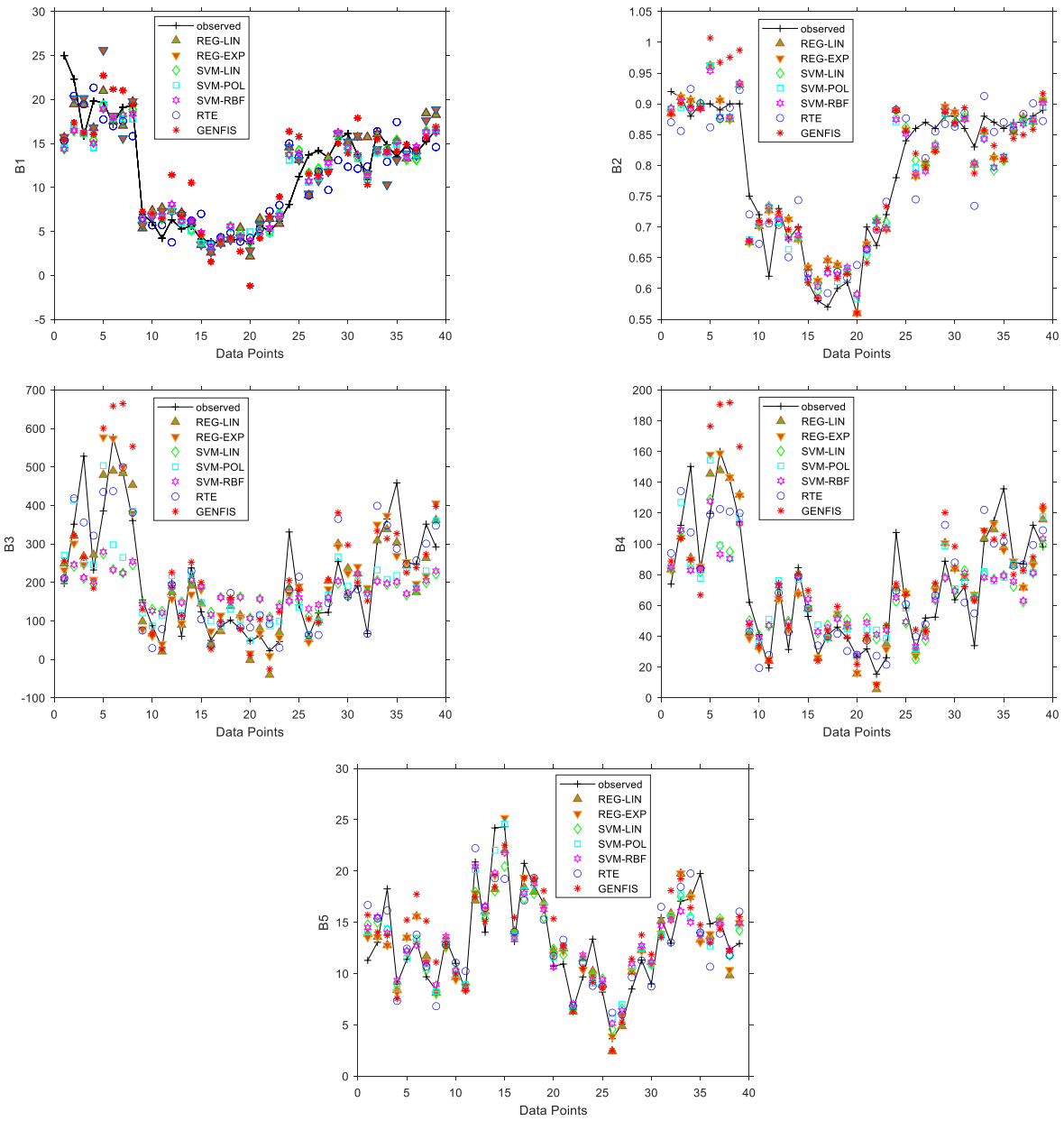


Fig. 11 Prediction of B_1, B_2, B_3, B_4, and B_5 by REG_LIN, REG_EXP, SVM_LIN, SVM_RBF, SVM_POL, RTE, and GENFIS models

steps in the training phase, therefore, their error bars are not given in the Figure. As it shown, SVM_POL is a stable method for predicting B_1 and B_5. SVM_POL and SVM_RBF are also stable for predicting B_2. RTE yields relatively stable results for predicting B_3 and B_4.

3.6 The sensitivity analysis

The sensitivity analysis is conducted to study the correlation between independent and dependent variables and specify the contribution of independent variables to the model output. The method of relative strength of effect (RSE) is applied. The RSE measure between ith input, X_i, and jth output, X_j, is computed using below equation

$$RSE = \frac{\sum_{k=1}^{m} x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^{m} x_{ik}^2 \sum_{k=1}^{m} x_{jk}^2}} \quad (28)$$

Fig. 18 depicts the RSE values of all dependent variables. H_S has the most impact on B_1, B_3, and B_4. B_2 is affected by Id and γ_d more than other inputs. B_5 is more affected by V_p, Id, and γ_d, respectively. The results indicate that the RSE values of independent variables for B_3 and B_4 have more variance in comparison to B_1, B_2, and B_5. This can justify the suitable performance of boosting regression trees for predicting B_3 and B_4 in comparison to other models. The regression trees are doing the feature selection simultaneously with prediction and the ensemble of regression trees can deal with the various importance of input features.

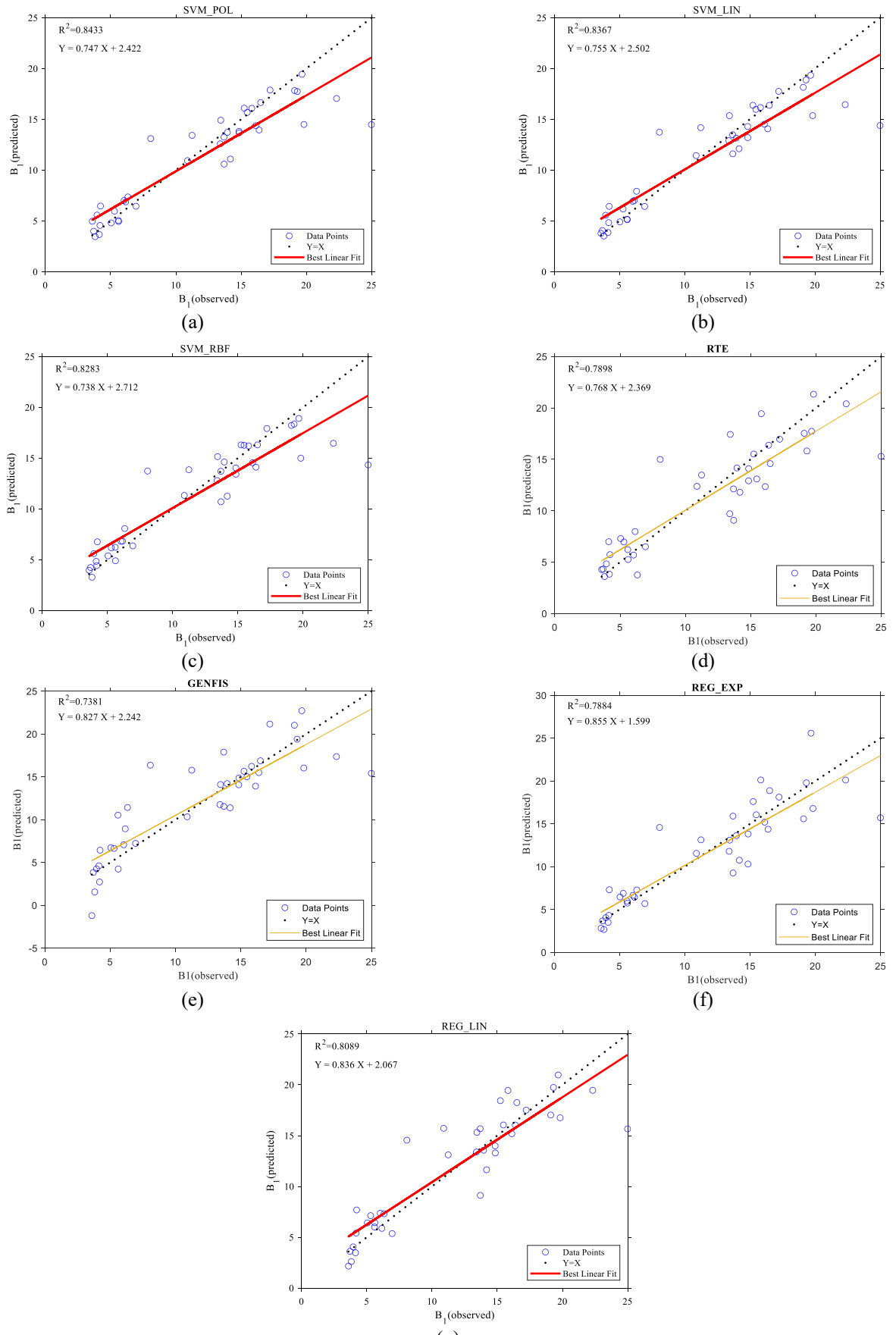


Fig. 12 Scatter plots of B_1 output for (a) SVM_POL, (b) SVM_LIN, (c) SVM_RBF, (d) RTE, (e) GENFIS, (f) REG-EXP, and (g) REG_LIN

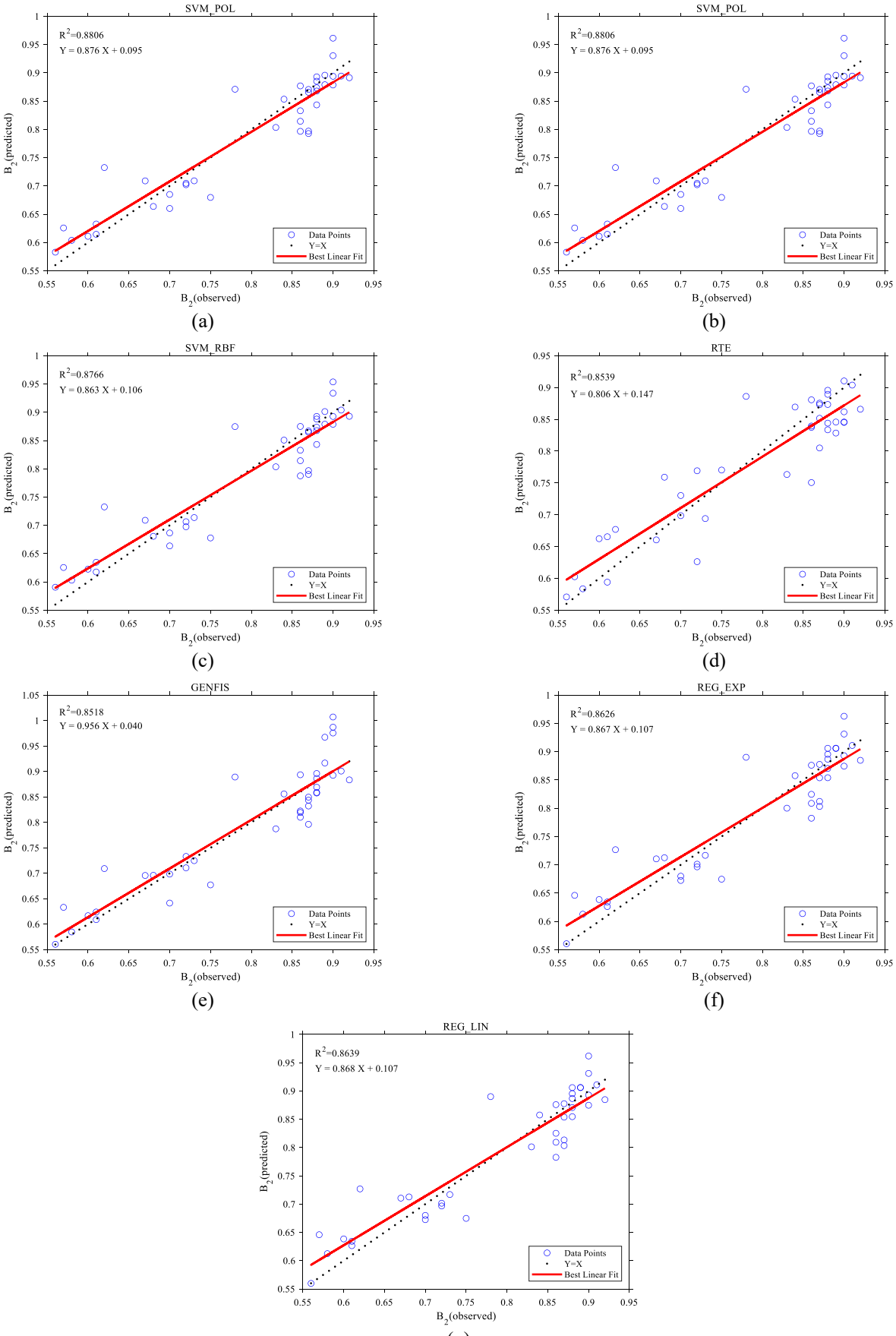
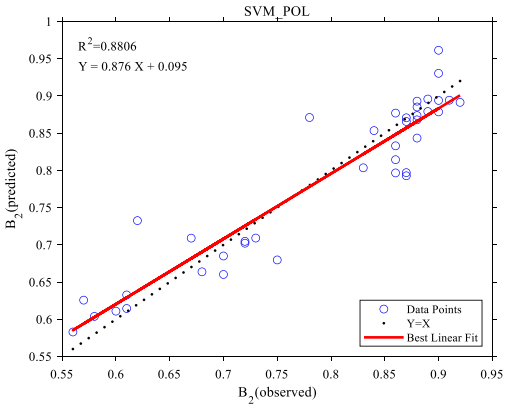
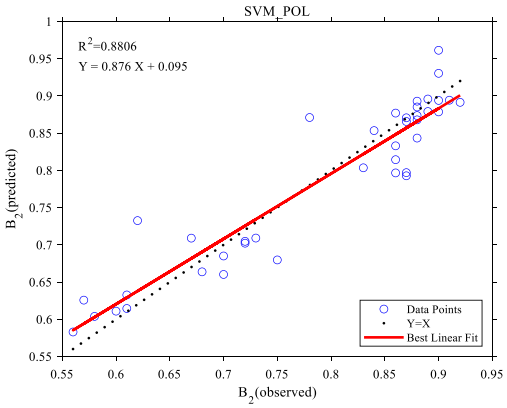


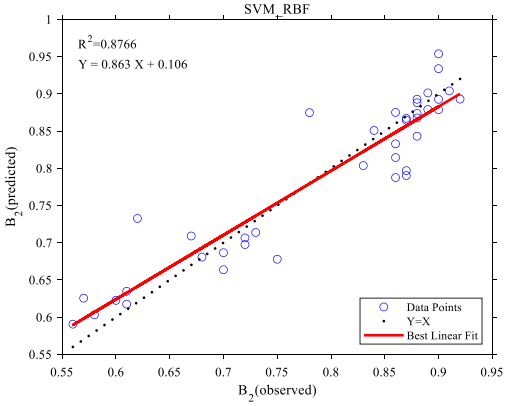
Fig. 13 Scatter plots of B_2 output for (a) SVM_POL, (b) SVM_LIN, (c) SVM_RBF, (d) RTE, (e) GENFIS, (f) REG_EXP, and (g) REG_LIN



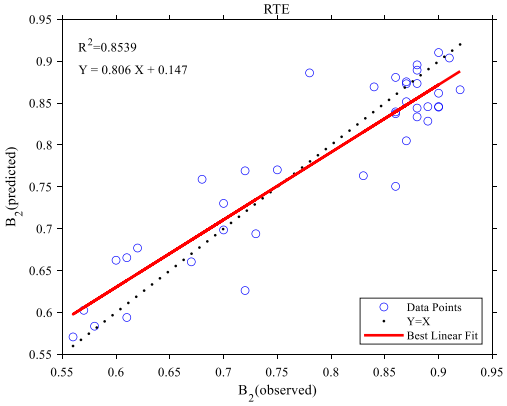
(a)



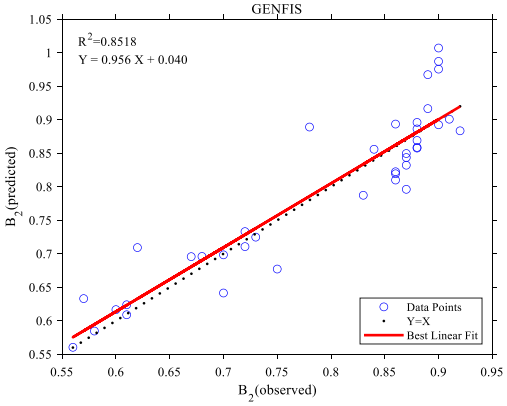
(b)



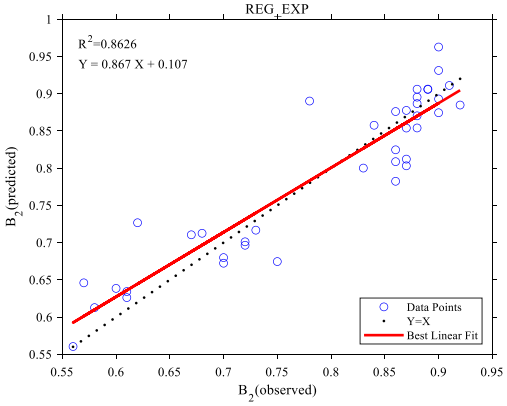
(c)



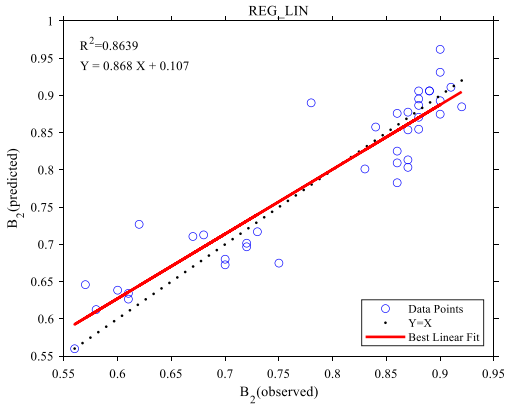
(d)



(e)



(f)



(g)

Fig. 14 Scatter plots of B_2 output for (a) SVM_POL, (b) SVM_LIN, (c) SVM_RBF, (d) RTE, (e) GENFIS, (f) REG_EXP, and (g) REG_LIN

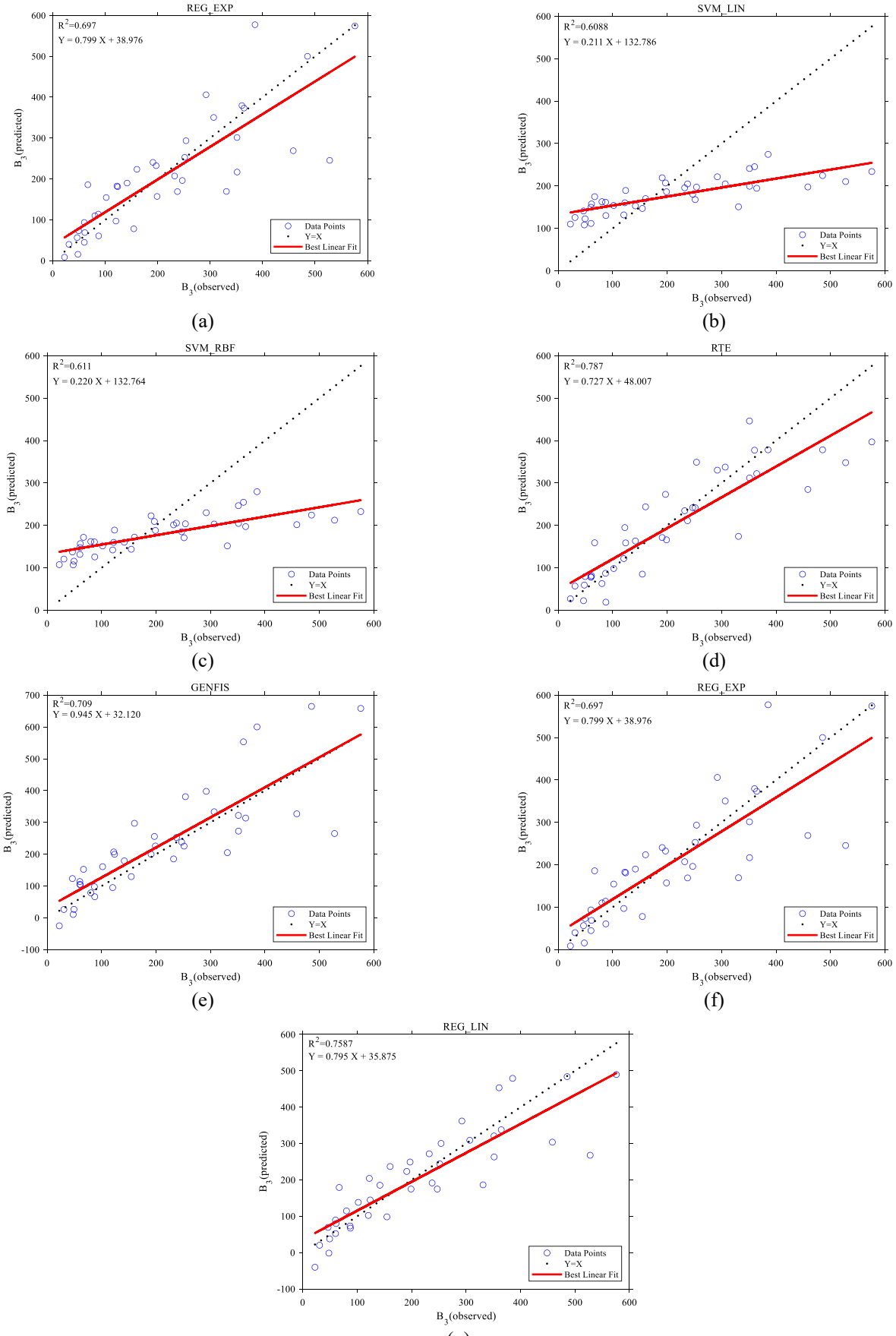


Fig. 15 Scatter plots of B_3 output for (a) SVM_POL, (b) SVM_LIN, (c) SVM_RBF, (d) RTE, (e) GENFIS, (f) REG_EXP, and (g) REG_LIN

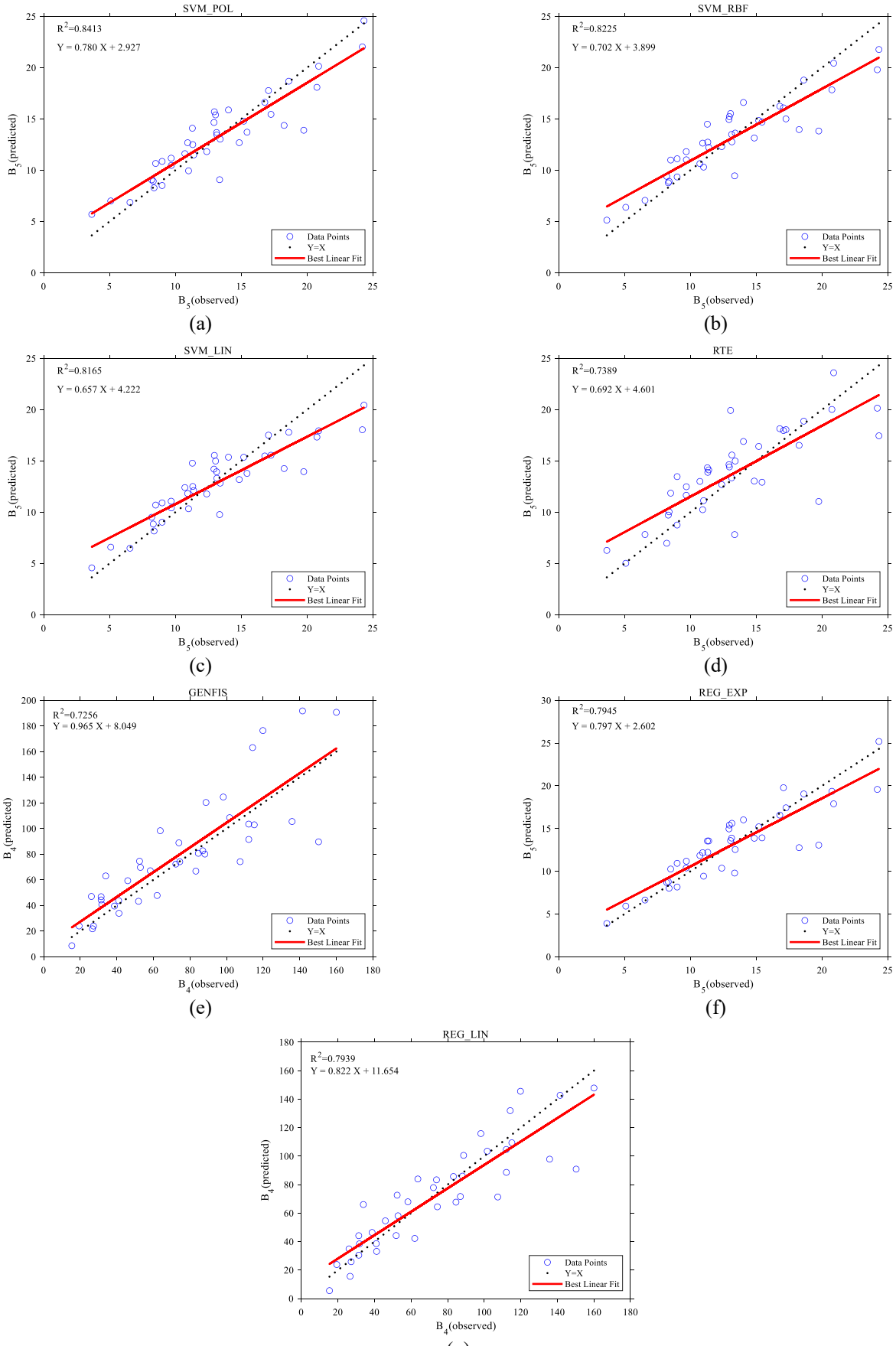


Fig. 16 Scatter plots of B_4 output for (a) SVM_POL, (b) SVM_LIN, (c) SVM_RBF, (d) RTE, (e) GENFIS, (f) REG_EXP, and (g) REG_LIN

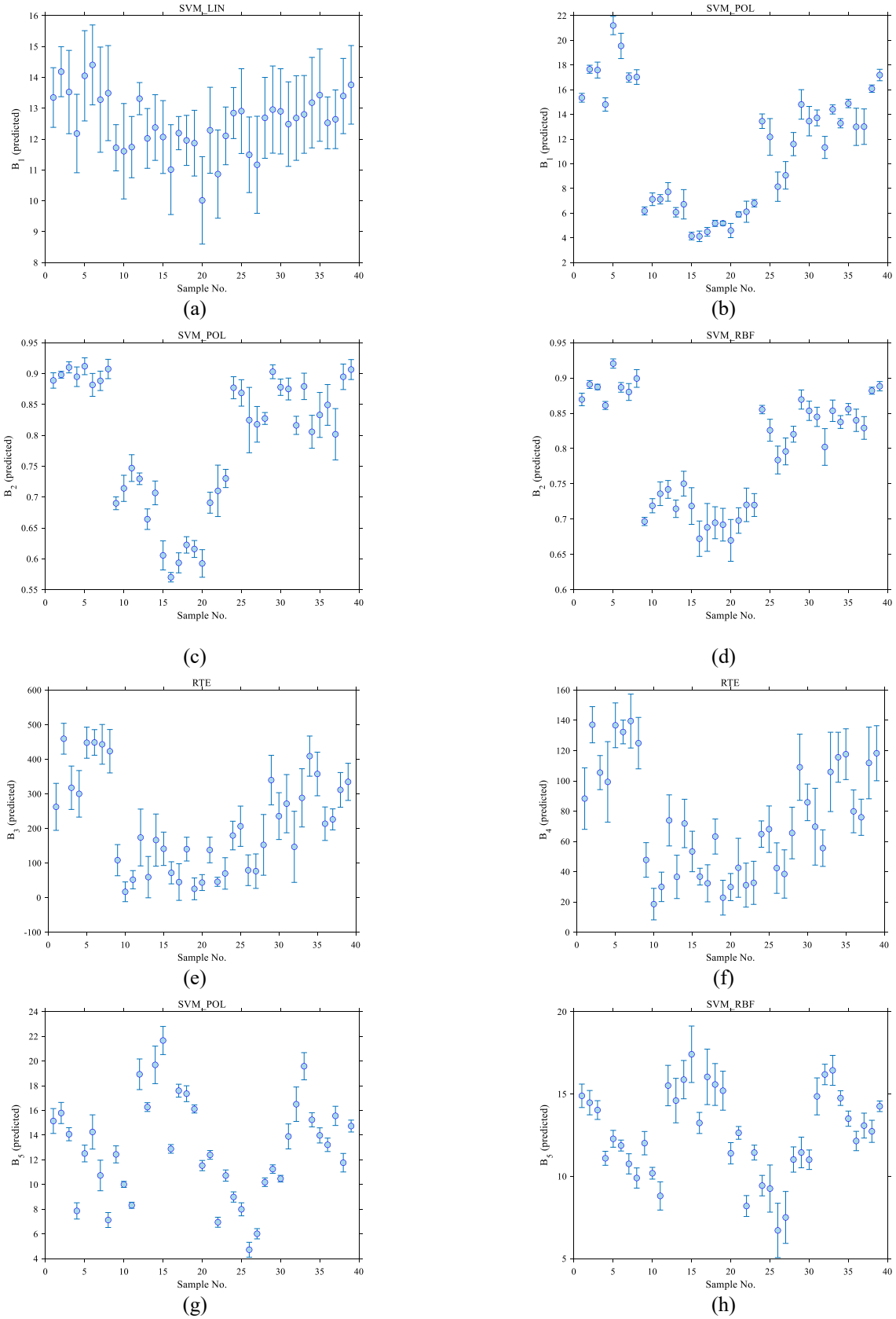


Fig. 17 Error bars of (a) the estimated B_1 by SVM_LIN, (b) the estimated B_1 by SVM_POL, (c) the estimated B_2 by SVM_POL, (d) the estimated B_2 by SVM_RBF, (e) the estimated B_3 by RTE, (f) the estimated B_4 by RTE, and (g) the estimated B_5 by SVM_POL, and (f) the estimated B_5 by SVM_RBF

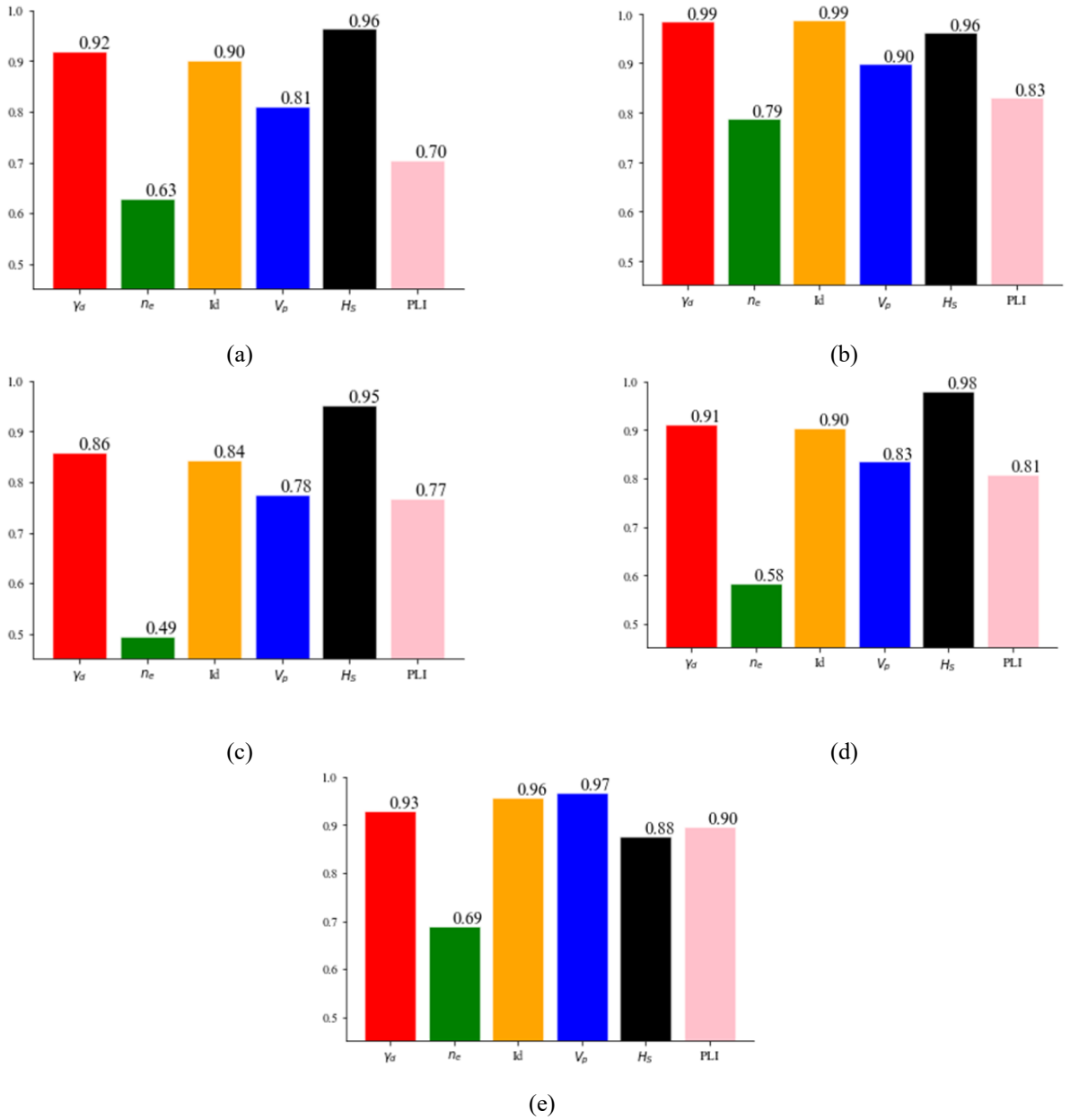


Fig. 18 RSE values for (a) B_1, (b) B_2, (c) B_3, (d) B_4, and (e) B_5

4. Conclusions

In the present study, the number of 39 samples including igneous, sedimentary, and metamorphic rocks were collected from different regions of Iran. Mineralogical, physical and mechanical properties as well as five well known rock brittleness indices (i.e., B_1, B_2, B_3, B_4, and B_5) were measured for the selected rock samples. Then, the rock brittleness indices were predicted from the results of simple, inexpensive, and quick laboratory tests namely dry unit weight, porosity, slake-durability index, P-wave velocit y, Schmidt rebound hardness, and point load strength index using multiple linear regression, exponential regression, support vector machine (SVM) with various kernels, generating fuzzy inference system, and regression tree ensemble (RTE) with boosting framework. This research conducts a comparative analysis of diverse machine learning models. The study examines the efficacy of single models vs. ensemble models (RTS), fuzzy methods (FIS) vs. non-fuzzy methods, and linear method (linear regression model) in contrast to non-linear methods.

The performance of the developed models was evaluated based on several statistical metrics such as mean square error, relative absolute error, root relative absolute error, determination coefficients, variance account for, mean absolute percentage error and standard deviation of the error. The comparison of the obtained results revealed that among the studied methods, SVM is the most suitable one for predicting B_1, B_2 and B_5. SVM is a powerful model which is highly sensitive to hyper-parameters. This study involves fine-tuning the SVM's hyper-parameters in the k-fold cross validation setting. This approach distinguishes our research from previous studies employing SVM for rock brittleness prediction, resulting in superior

performance compared to alternative models. Besides, RTE predicts B_3 and B_4 better than other models. It yields to R^2 of 78.70% and 80.81% for predicting B_3 and B_4, respectively, which improved the results in comparison to standalone models without considering feature selection. Furthermore, the sensitivity analysis was also performed to evaluate the relative importance of input variables on model's outputs. The sensitivity analysis indicates that there is a greater degree of variability in the RSE values of the independent variables corresponding to B_3 and B_4 as opposed to B_1, B_2, and B_5. This observation provide justification for the appropriate efficacy of boosting regression trees in the prediction of B_3 and B_4 as compared to alternative models, since RTE performs feature selection at the same time as learning the model.

Authors' contributions

The main idea of the research belongs to Davood Fereidooni who analyzed the obtained laboratory results as well as wrote the related parts and improved the main manuscript. Zohre Karimi has developed the machine learning algorithms used in the research and wrote the related parts of the text.

Acknowledgements

The authors acknowledge the official supports of the Engineering Geology and Rock Mechanics Laboratory of Damghan University for performing all laboratory tests of the research.

References

Altindag, R. (2000), "The role of rock brittleness on the analysis of percussive drilling performance (in Turkish)", *Proceedings of the 5th Turkish National Rock Mechanics Symposium*, Isparta, Turkey.

Altindag, R. (2002), "The evaluation of rock brittleness concept on rotary blasthole drills", *J. Southern African Inst. Min. Metallurgy*, **102**, 61-66. https://hdl.handle.net/10520/AJA0038223X_2763.

Altindag, R. (2003), "Correlation of specific energy with rock brittleness concepts on rock cutting", *J. Southern African Inst. Min. Metallurgy*, **103**, 163-171. https://hdl.handle.net/10520/AJA0038223X_2948.

Altindag, R. (2010), "Assessment of some brittleness indexes in rock-drilling efficiency", *Rock mechanics and rock engineering*, **43**(3), 361–370. https://doi.org/10.1007/s00603-009-0057-x.

Andreev, G.E. (1995), "Brittle failure of rock materials: test results and constitutive models", *Brookfield Press: Rotterdam The Netherlands*, p. 446.

ASTM, (1990), "Standard test method for slake durability of shales and similar weak rocks (D4644)", Annual Book of ASTM Standards, vol. 4.08. ASTM, Philadelphia. 863-865.

ASTM, (1995), "Standard test method for unconfined compressive strength of intact rock core specimens", ASTM standards on disc 04.08, Designation D2938.

ASTM (1996), "Standard test method for laboratory determination of pulse velocities and ultrasonic elastic constants of rock", Designation: D2845-95.

ASTM (2001a), "Standard test method for determination of rock hardness by rebound hammer method", ASTM standards on disc 04.09, D5873-00.

ASTM (2001b), "Standard method for determination of the point load strength index of rock", ASTM standards on disc 04.08, Designation, D5731.

ASTM (2001c), "Standard test method for splitting tensile strength of intact rock core specimens", ASTM standards on disc 04.08, Designation, D3967.

Aubertin, M. and Gill, D.E. (1988), "A methodology for assessing the potential for rock bursts in Abitibi mine", *Proceedings of the Colloque sur le Controle de Terrain (AMMQ)*, Val d'Or, 47-77.

Aubertin, M., Gill, D.E. and Simon, R. (1994), "On the use of the brittleness index modified (BIM) to estimate the post-peak behaviour of rocks, rock mechanics", *In: Nelson and Laubach (Eds.) Balkema*, 945-952.

Baron, L.I. (1962), "Determination of properties of rocks (in Russian)", *Gozgotekhizdat*, Moscow.

Bezdek, J.C. (2013), "Pattern recognition with fuzzy objective function algorithms", Springer Science & Business Media.

Cao, J., Gao, J., Nikafshan Rad, H., Mohammed, A.S., Hasanipanah, M. and Zhou, J. (2021), "A novel systematic and evolved approach based on XGBoost-firefly algorithm to predict Young's modulus and unconfined compressive strength of rock", *Eng. with Comput*., 1-17. https://doi.org/10.1007/s00366-020-01241-2.

Chen, P.H., Fan, R.E. and Lin, C.J. (2006), "A study on SMO-type decomposition methods for support vector machines", *IEEE T. Neural Netw.*, **17**(4), 893-908.

Coates, D.F. and Parsons, R.C. (1966), "Experimental criteria for classification of rock substances", *Int. J. Rock Mechanics and Min. Sci.*, **3**(3), 181-189. https://doi.org/10.1016/0148-9062(66)90022-2.

Evans, I. and Pomeroy, C.D. (1966), "The strength fracture and workability of coal", Pergamon Press, Oxford.

Gamble, J.C. (1971), "Durability-Plasticity classification of shales and other argillaceous rocks", Ph.D. Thesis, University of Illinois, Urbana-Champaign, IL, 161.

Ghadernejad, S., Nejati, H.R. and Yagiz, S. (2020), "A new rock brittleness index on the basis of punch penetration test data", *Geomech. Eng.*, **21**(4), 391-399. https://doi.org/10.12989/gae.2020.21.4.391.

Ghobadi, M.H. and Naseri, F. (2016), "Rock brittleness prediction using geomechanical properties of Hamekasi limestone: regression and artificial neural networks analysis", *Geopersia*, **6**(1), 19-33. https://doi.org/10.22059/JGEOPE.2016.57819.

Goktan, R.M. (1991), "Brittleness and micro-scale rock cutting efficiency", *Min. Sci. Tech.*, **13**, 237-241. https://doi.org/10.1016/0167-9031(91)90339-E.

Goodman, R.E. (1989), "Introduction to Rock Mechanics", *John Wiley & Sons Inc*, New York, p. 562.

Gunaydin, O., Kahraman, S. and Fener, M. (2004), "Sawability prediction of carbonate rocks from brittleness indexes", *Journal of the Southern African Institute of Mining and Metallurgy*, **104**(4), 239–243. https://hdl.handle.net/10520/AJA0038223X_2828.

Hajiabdolmajid, V. and Kaiser, P. (2003), "Brittleness of rock and stability assessment in hard rock tunneling", *Tunn. Undergr. Sp. Tech.*, **18**, 35-48. https://doi.org/10.1016/S0886-7798(02)00100-1.

Hasanipanah, M., Jamei, M., Mohammed, A.S., Amar, M.N., Hocine, O. and Khedher, K. M. (2022), "Intelligent prediction of rock mass deformation modulus through three optimized cascaded forward neural network models", *Earth Sci. Inform*., **15**(3), 1659-1669. https://doi.org/10.1007/s12145-022-00823-6.

Hetenyi, M. (1966), "*Handbook of experimental stress analysis*",

Wiley, New York.
Howell, J.V. (1960), "*Glossary of geology and related sciences*", American Geological Institute, Washington, DC.
Hucka, V. and Das, B. (1974), "Brittleness determination of rocks by different methods", *Int. J. Rock Mech. Min. Sci. Geomech. Abstracts*, **17**(10), 389-392. https://doi.org/10.1016/0148-9062(74)91109-7.
Ishikawa, K. (1986), "Guide to Quality Control", No. TS156.13713 1994.
ISRM, (2007), "The Blue Book: The complete ISRM suggested methods for rock characterization, testing and monitoring", 1974-2006, (Eds., Ulusay, R., and Hudson, J.A.), Compilation arranged by the ISRM Turkish National Group, Ankara, Turkey, Kazan Offset Press, Ankara.
Jahandideh, A. and Jafarpour, B. (2016), "Optimization of hydraulic fracturing design under spatially variable shale fracability", *J. Petroleum Sci. Eng.*, **138**, 174-188. https://doi.org/10.1016/j.petrol.2015.11.032.
Jahed Armaghani, D., Asteris, P.G., Askarian, B., Hasanipanah, M., Tarinejad, R. and Huynh, V.V. (2020), "Examining hybrid and single SVM models with different kernels to predict rock brittleness", *Sustainability*, **12**(6), 2229.
Jamei, M., Mohammed, A.S., Ahmadianfar, I., Sabri, M.M.S., Karbasi, M. and Hasanipanah, M. (2022), "Predicting rock brittleness using a robust evolutionary programming paradigm and regression-based feature selection model", *Appl. Sci.*, **12**(14), 7101. https://doi.org/10.3390/app12147101.
Jang, J.S., Sun, C.T. and Mizutani, E. (1997), "Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]", *IEEE T Automat. Control*, **42**(10), 1482-1484.
Kahraman, S. and Altindag, R. (2004), "A brittleness index to estimate fracture toughness", *Int. J. Rock Mech. Min. Sci. Geomech. Abstracts*, **41**, 343-348. https://doi.org/10.1016/j.ijrmms.07.010.
Kahraman, S., Bilgin, N. and Feridunoglu, C. (2003a), "Dominant rock properties affecting the penetration rate of percussive drills", *Int. J. Rock Mech. Min. Sci. Geomech. Abstracts*, **40**, 711-723. https://doi.org/10.1016/S1365-1609(03)00063-7.
Kahraman, S., Gunaydin, O., Fener, M. and Bilgin, N. (2003b), Correlation between Los Angeles abrasion loss and uniaxial compressive strength", *Proceedings of International Symposium on Industrial Minerals and Building Stones*, Istanbul, Turkey.
Kahraman, S. (2002), "Correlation of TBM and drilling machine performances with rock brittleness", *Eng. Geol.*, **65**(4), 269-283. https://doi.org/10.1016/S0013-7952(01)00137-5.
Karimi, Z. and Farzinfar, M. (2020), "Estimation of power in combined cycle power plant using adaptive neuro-fuzzy inference system (in Persian)", *Proceedings of the 1st national conference of applied water and power industry*.
Koopialipoor, M., Noorbakhsh, A., Noroozi Ghaleini, E., Jahed Armaghani, D. and Yagiz, S. (2019), "A new approach for estimation of rock brittleness based on non-destructive tests", *Nondestruct. Test. Eval.*, **34**(4), 354-375. https://doi.org/10.1080/10589759.2019.1623214.
Koopialipoor, M., Asteris, P.G., Mohammed, A.S., Alexakis, D.E., Mamou, A. and Armaghani, D.J. (2022), "Introducing stacking machine learning approaches for the prediction of rock deformation", *Transport. Geotech.*, **34**, 100756. https://doi.org/10.1016/j.trgeo.2022.100756.
Li, T., Chen, Y. and Zhang, J. (2012), "Logistics service provider segmentation based on improved FCM clustering for mixed data", *J. Comput.*, **7**(11), 2629-2633. https://doi.org/10.4304/jcp.7.11.2629-2633.
Loh, W.Y. (2011), "Classification and regression trees", *Wiley interdisciplinary reviews: data mining and knowledge discovery*, **1**(1), 14–23.
Mahmood, W., Mohammed, A. and HamaHussein, S. (2020), "Predicting mechanical properties and ultimate shear strength of gypsum, limestone and sandstone rocks using Vipulanandan models", *Geomech. Geoeng.*, **15**(2), 90-106. https://doi.org/10.1080/17486025.2019.1632494.
Mamdani, E.H. and Assilian, S. (1975), "An experiment in linguistic synthesis with a fuzzy logic controller", *Int. J. Man-Machine Studies*, **7**(1), 1-13. https://doi.org/10.1016/S0020-7373(75)80002-2.
Meng, F., Wong, L.N. and Zhou, H. (2020), "Rock brittleness indices and their applications to different fields of rock engineering: A review", *J. Rock Mech. Geotech. Eng.*, **13**(1), 221-247. https://doi.org/10.1016/j.jrmge.2020.06.008.
Mohammed, A.S. (2019), "Vipulanandan models to predict the mechanical properties, fracture toughness, pulse velocity and ultimate shear strength of shale rocks", *Geotech. Geol. Eng.*, **37**(2), 625-638. https://doi.org/10.1007/s10706-018-0633-5.
Morley, A. (1944), "Strength of materials: with 260 diagrams and numerous examples", *Longmans, Green and Company*, New York.
Nejati, H.R. and Moosavi, S.A. (2017), "A new brittleness index for estimation of rock fracture toughness", *J. Min. Environ.*, **8**(1), 83-91. https://doi.org/10.22044/jme.2016.579.
Obert, L. and Duvall, W.I. (1967), "Rock mechanics and the design of structures in rock", New York: Wiley.
Parsajoo, M., Mohammed, A.S., Yagiz, S., Armaghani, D.J. and Khandelwal, M. (2021), "An evolutionary adaptive neuro-fuzzy inference system for estimating field penetration index of tunnel boring machine in rock mass", *J. Rock Mech. Geotech. Eng.*, **13**(6), 1290-1299. https://doi.org/10.1016/j.jrmge.2021.05.010.
Ramezan, C.A. Warner, T.A. and Maxwell, A.E. (2019), "Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification", *Remote Sens*., **11**(2), 185. https://doi.org/10.3390/rs11020185.
Ramezani, R., Maadi, M. and Khatami, S.M. (2018), "A novel hybrid intelligent system with missing value imputation for diabetes diagnosis", *Alexandria Eng. J.*, **57**(3), 1883-1891. https://doi.org/10.1016/j.aej.2017.03.043.
Ramsay, J.G. (1967), "Folding and fracturing of rocks", McGraw-Hill Press: London, UK.
Ribacchi, R. (2000), "Mechanical tests on pervasively jointed rock material: insight into rock mass behaviour", *Rock Mech. Rock Eng.*, **33**(4), 243-266. https://doi.org/10.1007/s006030070002.
Shin, Y. (2015), "Application of boosting regression trees to preliminary cost estimation in building construction projects", *Comput. Intel. Neurosci.*, https://doi.org/10.1155/2015/149702.
Singh, S.P. (1986), "Brittleness and the mechanical winning of coal", *J. Min. Sci. Tech.*, **3**, 173-180. https://doi.org/10.1016/S0167-9031(86)90305-1.
Singh, S.P. (1987), "Criterion for the assessment of the cuttability of coal", (Eds., Szwilski, A.B. and Richards, M.J.), Underground mining methods and technology. Amsterdam. https://doi.org/10.1016/B978-0-444-42845-5.50024-3.
Smola, A. and Schölkopf, B. (2004), "A tutorial on support vector regression", *Stat. Comput.*, **14**, 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88.
Sugeno, M. (1985), "An introductory survey of fuzzy control", *Inform. Sci.*, **36**(1-2), 59-83. https://doi.org/10.1016/0020-0255(85)90026-X.
Sun, D., Lonbani, M., Askarian, B., Jahed Armaghani, D., Tarinejad, R., Thai Pham, B. and Huynh, V.V. (2020), "Investigating the applications of machine learning techniques to predict the rock brittleness index", *Appl. Sci.*, **10**(5), 1691.
Tarasov, B. and Potvin, Y. (2013), "Universal criteria for rock brittleness estimation under triaxial compression", *Int. J. Rock Mech Min. Sci.*, **59**, 57-69. https://doi.org/10.1016/j.ijrmms.2012.12.011.

Tiryaki, B. (2008), "Predicting intact rock strength for mechanical excavation using multivariate statistics, artificial neural networks, and regression trees", *Eng. Geol.*, **99**(1-2), 51-60. https://doi.org/10.1016/j.enggeo.2008.02.003.

Vipulanandan, C. and Mohammed, A. (2018), "New Vipulanandan failure model and property correlations for sandstone, shale and limestone rocks", In IFCEE 2018, 365-376.

Vipulanandan, C., Mohammed, A. and Mahmood, W. (2021), "Characterizing rock properties and verifying failure parameters using data analytics with vipulanandan failure and correlation models", *Proceedings of the 55th US Rock Mechanics/Geomechanics Symposium*, OnePetro.

Wang, Y. and Chen, Y. (2014), "A comparison of Mamdani and Sugeno fuzzy inference systems for traffic flow prediction", *J. Comput.*, **9**(1), 12-21. https://doi.org/10.4304/jcp.9.1.12-21.

Wood, D.A. (2020), "Brittleness index predictions from Lower Barnett Shale well-log data applying an optimized data matching algorithm at various sampling densities", *Geosci. Front.*, 101087. https://doi.org/10.1016/j.gsf.2020.09.016.

Yang, S.Q., Yin, P.F. and Ranjith, P.G. (2020), "Experimental study on mechanical behavior and brittleness characteristics of longmaxi formation shale in Changning, Sichuan basin, China", *Rock Mech. Rock Eng.*, **11**, 1-23. https://doi.org/10.1007/s00603-020-02057-8.

Yagiz, S., Yazitova, A. and Karahan, H. (2020), "Application of differential evolution algorithm and comparing its performance with literature to predict rock brittleness for excavatability", *Int. J. Min. Reclamat. Environ.*, **34**(9), 672-685.

Yarali, O. (2007), "Investigation of the relations between rock brittleness and drilling rate index", *Proceedings of the 20th International Mining Congress of Turkey, (in Turkish),* Ankara, Turkey.

Yarali, O. and Soyer, E. (2011), The effect of mechanical rock properties and brittleness on drillability", *Scientific Res. Essays*, **6**(5), 1077-1088. https://doi.org/10.5897/SRE10.1004.

Ye, Y., Tang, S. and Xi, Z. (2020), "Brittleness evaluation in shale gas reservoirs and its influence on fracability", *Energies*, **13**(2), 388. https://doi.org/10.3390/en13020388.

Yilmaz, N.G., Karaca, Z., Goktan, R.M. and Akal, C. (2008), Relative brittleness characterization of some selected granitic building stones: influence of mineral grain size", *Constr. Build. Mater.*, **23**(1), 370-375. https://doi.org/10.1016/j.conbuildmat.11.014.

Zhang, Y., Feng, X.T., Yang, C., Han, Q., Wang, Z. and Kong, R. (2021), "Evaluation method of rock brittleness under true triaxial stress states based on pre-peak deformation characteristic and post-peak energy evolution", *Rock Mech. Rock Eng.*, **54**(3), 1277-1291. https://doi.org/10.1007/s00603-020-02330-w.

Zhou, Z.H. (2019), "Ensemble methods: foundations and algorithms", *Chapman and Hall/CRC*.

CC