Hindawi
Advances in Civil Engineering
Volume 2020, Article ID 8850535, 23 pages
https://doi.org/10.1155/2020/8850535

Research Article

Applications of Gene Expression Programming for Estimating Compressive Strength of High-Strength Concrete

Fahid Aslam,[1] Furqan Farooq,[2] Muhammad Nasir Amin,[3] Kaffayatullah Khan,[3] Abdul Waheed,[2] Arslan Akbar,[4] Muhammad Faisal Javed,[2] Rayed Alyousef,[1] and Hisham Alabdulijabbar[1]

[1]*Department of Civil Engineering, College of Engineering in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia*
[2]*Department of Civil Engineering, COMSATS University Islamabad, Abbottabad 22060, Pakistan*
[3]*Department of Civil and Environmental Engineering, College of Engineering, King Faisal University (KFU), P.O. Box 380, Al-Hofuf, Al Ahsa 31982, Saudi Arabia*
[4]*Department of Architecture and Civil Engineering, City University of Hong Kong, Kowloon 999077, Hong Kong*

Correspondence should be addressed to Furqan Farooq; furqanfarooq_33@yahoo.com, Arslan Akbar; aakbar4-c@my.cityu.edu.hk, Muhammad Faisal Javed; arbabfaisal@cuiatd.edu.pk, and Rayed Alyousef; r.alyousef@psau.edu.sa

Received 14 June 2020; Revised 25 August 2020; Accepted 9 September 2020; Published 26 September 2020

Academic Editor: Shengwen Tang

The experimental design of high-strength concrete (HSC) requires deep analysis to get the target strength. In this study, machine learning approaches and artificial intelligence python-based approaches have been utilized to predict the mechanical behaviour of HSC. The data to be used in the modelling consist of several input parameters such as cement, water, fine aggregate, and coarse aggregate in combination with a superplasticizer. Empirical relation with mathematical expression has been proposed using engineering programming. The efficiency of the models is assessed by statistical analysis with the error by using MAE, RRMSE, RSE, and comparisons were made between regression models. Moreover, variable intensity and correlation have shown that deep learning can be used to know the exact amount of materials in civil engineering rather than doing experimental work. The expression tree, as well as normalization of the graph, depicts significant accuracy between target and output values. The results reveal that machine learning proposed adamant accuracy and has elucidated performance in the prediction aspect.

1. Introduction

High-strength concrete (HSC) production in the construction industry has been adamantly upsurge in recent years for use in modern construction work [1–3]. Improving concrete performance ultimately enhances the overall effectiveness of modern concrete structures. HSC has significant strength in concrete media, greater than 40 MPa compared with the conventional concrete system [4]. HSC is a modified form of concrete that requires vibrating media and nonvibrating media for its placement; moreover, it is dense and homogenous concrete with adamant high strength and superior durability properties as compared with traditional concrete making it extensively applicable to the concrete industry [5, 6]. For example, it is adamantly used for high-rise buildings, long-span bridges, piers, etc. American Concrete Institute (ACI) defines HSC as “concrete that possesses specific requirement for its working which cannot be achieved by conventional concrete” [7].

Their use in construction improves the working environment and unlocks the way for concrete construction automation. However, the major problem lies with its design procedure due to the complex nature of HSC. Various researchers have reported different guidelines and standards for design mixture, which compromises the use of chemical and mineral admixtures [8–10]. Due to its complex nature rather than conventional strength concrete, it requires experience and adamant knowledge of the constituent used in the mixture process. The HSC complex structure requires an arduous mix design procedure for attaining its essential properties. Concrete strength is an important aspect in high-strength concrete; however, variation in constituents, chemical and mineral admixtures, and design specifications may vary from source to source [11–14]. This creates ambiguity in the general relationship between cement ratio to mineral admixtures, chemical admixtures, w/b ratio, and aggregate grain sizes. These variations in constituent somehow, if not properly managed, will produce deficiency in concrete strength. These constituents can be properly and adamantly managed by using their desire (optimized) quantities that will produce the utmost aspect of strength rather than using experimental work. As these experimental works cost resources and time by using hit and trial of taking desire quantities to achieve maximum effect on ultimate strength. In this aspect, numerous researchers have used traditional methods by using linear and nonlinear equations to give prediction measures of (HPC) strength. These methods were based on statistical analysis; however, accurate prediction from equation-based approaches is difficult and thus requires a lot of research to overcome these obstacles. In recent years, concepts of machine learning neural-based approaches overwhelm these difficulties and provide an accurate prediction of concrete strength.

Machine learning approaches such as genetic engineering programming (GEP) [15–17], artificial neural networks (ANN) [18–21], support vector machine (SVM), decision tree (DT), adaptive boost algorithm (ABA), and adaptive neuro-fuzzy interference (ANFIS) [22–26] have been widely used and publicized in civil engineering domain [27]. Dong et al. used machine learning approaches like ANN and ANFIS for prediction of compressive strength of geopolymer concrete at 28 days with 210 data samples. The authors concluded that these approaches give better prediction; however, ANFIS approach outbreaks with the coefficient of determination (R^2) and model performance from ANN [28]. Nour and Güneyisi [29] used genetic engineering programming (GEP) for prediction of compressive strength of recycled aggregate (RA) concrete filled with steel tube columns with 97 test datasets and concluded that GEP provides an accurate prediction of (RACFSTC) with empirical relation. The authors observed and concluded the coefficient of determination (R^2) for testing and training is 0.996 and 0.995, respectively, providing accurate behavior of model [29]. Bingöl et al. model the compressive strength of lightweight exposed to high temperature by employing ANN approach [30]. The authors concluded that ANN is an advanced predictive approache; however, the model predicts the strength with adequate accuracy. Moreover, researchers used ANN and other machine learning approaches for the prediction properties of recycled aggregate concrete and high-performance concretes [31–36]. Pala et al. investigated the long-term impact of replacing silica and fly ash on cured concrete performance. Their experiments included concrete mixtures of different water-cement ratios, including the lowest and highest fly ash concentrations, with or without additional small silica fume amounts. Based on the results, ANNs have tremendous potential as a suitable means to examine the effect of secondary raw materials on the compressive strength of concrete [37]. Iqbal et al. used genetic engineering programming machine learning approach for the prediction of green concrete with 234 data samples. The authors reported that gene programming gives adamant prediction accuracy with an empirical relationship [32]. Javed et al. [15] conducted experimental program predict the strength of sugarcane bagasse ash using different machine learning approaches. The authors obtained a strong correlation between input and output by using GEP approach. Moreover, the same trend was also observed by Azim et al. [31]. The authors used GEP for the prediction of reinforced concrete structure with adamant accuracy. GEP is superior to existing methods like feature selection, ANN, and M5P methods.

The choice of features is an essential step in data processing and is seen in many areas, like genetics, medicine, and bioinformatics. The selection of the key elements (genes) is necessary in order to uncover new information concealed inside the genetic code and to recognise relevant biomarkers. Although the proposed algorithms can help sort by large numbers of genes relating to the problem at hand, the results generated appear to be unstable and thus cannot be reconstructed in other studies. It is vital to emphasize that the two most widely employed Machine learning models in previous studies, i.e., the ANN and the M5P models, sometimes face challenges to reliably predict outcomes in data domains that have complicated input(s)-output(s) feature(s) (i.e., highly nonlinear or non-monotonic) [17, 38–41]. That is because the ANN models, as well as their variants such as MLP-ANN, are predicated on local optimization and search algorithms (e.g., the back-propagation technique used in many neural ML-models based on a network to maximize the activation function parameters), which are highly susceptible to local (or around) minima instead of converging to the globally relevant.

This paper aims to build a GEP-based model for accurate prediction for high-strength concrete with an empirical equation. For this aspect, data have been acquired from previously published work compromising of 357 data points as shown in Table 1. It is worth mentioning that this research is primarily based on estimating the compressive strength of the high-strength concrete using a genetic engineering approach. The parameters used in the modelling of HSC consist of (cement, water, fine aggregate, coarse aggregate, and superplasticizer). Section 2 represents data input to output (strength) with optimal quantities with graphical representation (Kde contour graph), which was done by using python programmable software. Section 3 then shows the importance of each variable on its output by conducting sensitivity analysis (SA) or permutation features importance (PFI). Section 4 represents the statistical measures for model

Table 1

S. no	Cement	Coarse aggregate	Fine aggregate	Water	Superplasticizer	Compressive strength
(1)	360	845	900	160	1.5	48
(2)	320	950	782	160	1.5	46.1
(3)	356	845	951	160	1	46
(4)	463	845	750	180	0.75	52.7
(5)	300	720	787	168	4.8	52.7
(6)	486	950	714	170	1	56.1
(7)	360	950	797	160	1.5	48.7
(8)	500	769.18	740.16	154	10.5	86.99
(9)	510	950	628	170	2	61.7
(10)	400	950	811	160	1	49.4
(11)	284	898	874	160	1.5	43.8
(12)	356	898	899	160	0.5	44.5
(13)	330.85	739.44	857.57	180.05	6.02	57.3
(14)	350	500	1050	178.5	3.8	49.5
(15)	425	845	868	170	0	47.7
(16)	193	900	1024	136.89	2.808	44
(17)	340	845	838	170	1	45
(18)	331.33	739.44	875.16	180.05	6.02	57.3
(19)	302	845	880	170	0.5	41.5
(20)	411	898	680	180	0.25	47.3
(21)	486	898	766	170	1.5	57.8
(22)	360	898	738	180	0	43.2
(23)	160	900	886	156	3.2	44
(24)	160	900	886	156	3.2	44
(25)	330.85	768.88	840	180.05	6.08	58.45
(26)	340	898	786	170	1	45.1
(27)	360	898	848	160	1.5	48
(28)	567	898	700	170	1	63.9
(29)	284	845	926	160	1.5	43.7
(30)	567	845	751	170	1.5	64.6
(31)	383	950	749	170	0	45
(32)	320	950	835	160	0.5	42.6
(33)	480	898	604	180	1.75	57
(34)	340	950	734	170	0.5	43.3
(35)	320	687	1016	174	5.21	53.6
(36)	220	900	916	156	3.2	47
(37)	220	900	916	156	3.2	47
(38)	350	883	815	183.54	3.864	55.3
(39)	320	845	938	160	1.5	45
(40)	425	845	853	170	0	47.1
(41)	427	844	779	194.578	4.336	59.4
(42)	540	845	677	180	1.25	61.1
(43)	411	845	732	180	1.25	51
(44)	279.5	630	1135	200	2.1	44.344
(45)	320	898	886	160	0.5	42.6
(46)	302	950	776	170	0.5	41
(47)	237	900	960	159.1	2.96	46
(48)	237	900	960	159.1	2.96	46
(49)	400	845	914	160	1	49.6
(50)	365.5	630	1135	200	3.11	50.982
(51)	180	829	788	198	3	42.5
(52)	198	900	874	146.2	3.44	46
(53)	400	989	863	160	1	49.1
(54)	340	950	790	170	0	41.3
(55)	302	898	828	170	0.5	40.8
(56)	540	950	573	180	0.75	60
(57)	350	621	939	173.5	6.75	57.3
(58)	383	898	801	170	0.5	45.7
(59)	540	898	625	180	0.75	59.9
(60)	514	898	717	180	0.25	54.6

Table 1: Continued.

S. no	Cement	Coarse aggregate	Fine aggregate	Water	Superplasticizer	Compressive strength
(61)	340	845	893	170	0	41.1
(62)	320	845	834	180	0	40.3
(63)	360	750	900	200	12	70.67
(64)	284	950	822	160	1	42.9
(65)	356	950	847	160	0.5	43.6
(66)	463	898	698	180	0.25	50.3
(67)	540	898	625	180	1.25	59.9
(68)	248	900	808	175.89	3.608	50
(69)	248	900	808	175.89	3.608	50
(70)	350	500	1050	143.5	8.9	60.1
(71)	387	630	1135	200	3.29	51.039
(72)	360	950	686	180	0	42.4
(73)	378	845	907	170	0.5	44.9
(74)	540	845	677	180	1.75	61.1
(75)	552	1486	342	160	5.52	91.3
(76)	514	845	769	180	0.25	54.2
(77)	340	898	842	170	0	40.8
(78)	383	898	801	170	0	45.7
(79)	320	898	782	180	0	40
(80)	540	950	573	180	1.25	60
(81)	284	950	822	160	1.5	42.9
(82)	458	698	748	190	8.4	68.2
(83)	467	762	865	182	8.53	74.2
(84)	408.5	630	1135	200	3.68	53.006
(85)	302	898	828	170	0	40.8
(86)	400	845	914	160	1.5	49.6
(87)	258	630	1135	200	1.68	43.184
(88)	480	898	604	180	1.25	57
(89)	440	927	721	176	9.78	79.6
(90)	567	845	751	170	1	64.6
(91)	301	630	1135	200	2.26	44.003
(92)	320	845	938	160	1	45
(93)	302	950	776	170	0	41
(94)	320	898	886	160	1	42.6
(95)	567	950	648	170	1	62.4
(96)	567	898	700	170	1.5	63.9
(97)	340	950	734	170	0	43.3
(98)	360	845	789	180	0	42
(99)	302	845	880	170	1	41.5
(100)	330.36	768.88	857.57	179.74	6.02	56.6
(101)	383	950	749	170	0.5	45
(102)	400	989	863	160	0.5	49.1
(103)	340	898	786	170	0.5	45.1
(104)	340	845	838	170	0.5	45
(105)	425	845	868	170	0.5	47.7
(106)	360	898	848	160	1	48
(107)	486	845	818	170	1.5	58.8
(108)	463	845	750	180	0.25	52.7
(109)	463	845	750	180	1.25	52.7
(110)	425	950	764	170	0	46
(111)	425	898	816	170	0	46
(112)	320	950	835	160	1	42.6
(113)	220	900	916	156	3.2	45
(114)	220	900	916	156	3.2	45
(115)	330.36	754.18	840	180.38	6.02	59.26
(116)	275	900	827	184.9	3.44	48
(117)	438	723	774	191	8.1	69.5
(118)	510	950	628	170	1.5	61.7
(119)	356	898	899	160	1	44.5
(120)	360	845	900	160	1	48

Table 1: Continued.

S. no	Cement	Coarse aggregate	Fine aggregate	Water	Superplasticizer	Compressive strength
(121)	356	845	951	160	0.5	46
(122)	356	845	951	160	1.5	46
(123)	440	924	720	176	8	76.9
(124)	486	950	714	170	1.5	56.1
(125)	378	898	855	170	0	42.5
(126)	348	848	783	174.46	4.576	58.6
(127)	411	898	680	180	0.75	47.3
(128)	302	845	880	170	0	41.5
(129)	400	950	811	160	0.5	49.4
(130)	344	630	1135	200	2.75	50.37
(131)	360	898	738	180	0.5	43.2
(132)	486	950	714	170	0.5	56.1
(133)	480	950	552	180	1.75	59.6
(134)	360	845	848	180	0	41.5
(135)	486	898	766	170	1	57.8
(136)	320	950	782	160	1	46.1
(137)	284	898	874	160	1	43.8
(138)	437	950	697	170	1.5	55.9
(139)	567	845	751	170	2	64.6
(140)	360	950	797	160	1	48.7
(141)	510	898	679	170	2	63.4
(142)	540	845	677	180	0.75	61.1
(143)	411	845	732	180	0.75	51
(144)	400	950	811	160	1.5	49.4
(145)	284	845	926	160	1	43.7
(146)	250	853	787	192.66	4.056	51.5
(147)	425	845	853	170	0.5	47.1
(148)	237	900	1034	133.2	2.96	49
(149)	360	750	900	200	12	73.7
(150)	350	840	768	207.872	4.096	51.7
(151)	320	845	886	160	1.5	47.7
(152)	340	950	734	170	1	43.3
(153)	400	845	914	160	0.5	49.6
(154)	320	845	834	180	0.5	40.3
(155)	302	898	828	170	1	40.8
(156)	302	950	776	170	1	41
(157)	514	898	717	180	0.75	54.6
(158)	331.33	768.88	857.57	179.74	6.08	55.9
(159)	328.5	533.7	625.1	225	0	40
(160)	383	898	801	170	1	45.7
(161)	220	881	686	176	4.4	47.5
(162)	340	845	893	170	0.5	41.1
(163)	340	950	790	170	0.5	41.3
(164)	220	900	916	156	3.2	49
(165)	220	900	916	156	3.2	49
(166)	220	900	916	156	3.2	49
(167)	220	900	916	156	3.2	49
(168)	220	900	916	156	3.2	49
(169)	220	900	916	156	3.2	49
(170)	378	950	803	170	0	41.8
(171)	417	759	828	182.4	4.56	61.82
(172)	463	898	698	180	0.75	50.3
(173)	220	900	916	156	3.2	44
(174)	540	898	625	180	1.75	59.9
(175)	400	989	863	160	1.5	49.1
(176)	330.36	768.88	840	180.05	5.97	55.3
(177)	514	845	769	180	0.75	54.2
(178)	427	844	779	243.9	4.336	59.4
(179)	510	750	900	209	12	77.15
(180)	378	845	907	170	1	44.9

Table 1: Continued.

S. no	Cement	Coarse aggregate	Fine aggregate	Water	Superplasticizer	Compressive strength
(181)	405	845	805	180	0	43.6
(182)	533	950	701	160	2	67.8
(183)	411	950	628	180	0.5	45.7
(184)	320	898	834	160	1.5	48.5
(185)	320	898	782	180	0.5	40
(186)	360	950	686	180	0.5	42.4
(187)	333	500	1050	158	7.3	52.9
(188)	284	950	822	160	0.5	42.9
(189)	540	950	573	180	1.75	60
(190)	340	898	842	170	0.5	40.8
(191)	480	898	604	180	0.75	57
(192)	356	950	847	160	1	43.6
(193)	180	828	788	198	4.8	44.9
(194)	220	896	697	176	8	64.3
(195)	350	500	1050	161	6.9	58.7
(196)	437	898	749	170	1.5	56.6
(197)	360	845	789	180	0.5	42
(198)	412.5	520.1	612.7	203.5	0	44
(199)	320	845	938	160	0.5	45
(200)	322.5	630	1135	200	2.58	43.984
(201)	330.85	768.88	875.16	180.38	6.02	61.1
(202)	567	898	700	170	2	63.9
(203)	340	845	838	170	0	45
(204)	425	845	868	170	1	47.7
(205)	486	845	818	170	1	58.8
(206)	320	898	886	160	1.5	42.6
(207)	567	950	648	170	1.5	62.4
(208)	340	898	786	170	0	45.1
(209)	514	950	665	180	0.25	52.1
(210)	300	618	935	162	6.75	52.3
(211)	405	898	754	180	0	43
(212)	383	950	749	170	1	45
(213)	425	898	816	170	0.5	46
(214)	360	898	848	160	0.5	48
(215)	275	840	775	183.75	4.2	54.5
(216)	326	500	1050	158	5.1	54.3
(217)	425	950	764	170	0.5	46
(218)	510	845	731	170	2	64.4
(219)	330	898	699	176	9.11	67.2
(220)	360	898	797	180	0	40
(221)	330.36	754.18	875.16	180.05	6.08	61.3
(222)	411	898	680	180	1.25	47.3
(223)	360	845	848	180	0.5	41.5
(224)	437	845	801	170	1.5	56.8
(225)	360	845	900	160	0.5	48
(226)	360	898	738	180	1	43.2
(227)	500	788.07	758.33	140	10.5	88.98
(228)	320	950	835	160	1.5	42.6
(229)	333	766	835	180.84	4.384	50.24
(230)	510	950	628	170	1	61.7
(231)	378	898	855	170	0.5	42.5
(232)	315	673	1025	173	5.51	50.7
(233)	411	845	732	180	0.25	51
(234)	356	898	899	160	1.5	44.5
(235)	220	900	916	156	3.2	43
(236)	480	950	552	180	1.25	59.6
(237)	440	916	713	176	8.22	79.2
(238)	325	777	611	221	5.2	50.07
(239)	350	883	815	251.16	3.864	55.3
(240)	486	898	766	170	0.5	57.8

Table 1: Continued.

S. no	Cement	Coarse aggregate	Fine aggregate	Water	Superplasticizer	Compressive strength
(241)	401.5	518.6	610.9	203.5	0	50
(242)	510	898	679	170	1.5	63.4
(243)	330.85	739.44	840	179.74	5.97	53.6
(244)	425	845	853	170	1	47.1
(245)	220	880	685	176	8.89	59.9
(246)	480	845	655	180	1.75	60.8
(247)	357	742	878	182	7.98	67.5
(248)	437	950	697	170	1	55.9
(249)	399	882	814	174.65	3.992	55
(250)	389	950	680	170	1.5	54.3
(251)	284	898	874	160	0.5	43.8
(252)	284	845	926	160	0.5	43.7
(253)	400	950	759	180	0	42.1
(254)	320	845	834	180	1	40.3
(255)	330	913	711	176	7.5	62.2
(256)	320	950	782	160	0.5	46.1
(257)	500	769.52	740.48	154	10.5	90.99
(258)	360	950	745	180	0	39.5
(259)	310	667	1018	170	6	51.2
(260)	500	769.85	740.8	154	10.5	83.15
(261)	360	950	797	160	0.5	48.7
(262)	220	900	982	132	3.2	51
(263)	514	898	717	180	1.25	54.6
(264)	320	845	886	160	1	47.7
(265)	350	840	768	302.08	4.096	51.7
(266)	340	845	893	170	1	41.1
(267)	407	761	815	181	7.5	70.4
(268)	225	652	908	175	4	41.42
(269)	514	845	769	180	1.25	54.2
(270)	463	898	698	180	1.25	50.3
(271)	540	750	900	209	12	77.82
(272)	405	845	805	180	0.5	43.6
(273)	340	950	790	170	1	41.3
(274)	378	845	907	170	1.5	44.9
(275)	440	775	866	182	9.35	77.9
(276)	378	950	803	170	0.5	41.8
(277)	300	613	927	175.5	6.75	59.1
(278)	320	898	782	180	1	40
(279)	411	950	628	180	1	45.7
(280)	400	898	811	180	0	41.5
(281)	360	950	686	180	1	42.4
(282)	360	845	789	180	1	42
(283)	340	898	842	170	1	40.8
(284)	320	898	834	160	1	48.5
(285)	457	950	764	160	1.75	61.5
(286)	600	898	646	180	0.75	60.5
(287)	533	950	701	160	1.5	67.8
(288)	400	845	863	180	0	41.3
(289)	437	898	749	170	1	56.6
(290)	331.33	754.18	840	179.74	6.02	53.1
(291)	533	898	753	160	2	69.4
(292)	280	900	946	156	3.2	45
(293)	486	845	818	170	0.5	58.8
(294)	405	950	702	180	0	41.5
(295)	356	950	847	160	1.5	43.6
(296)	453	950	608	170	2	61.9
(297)	405	898	754	180	0.5	43
(298)	450	845	821	180	0	44.5
(299)	330	899	700	176	4.4	60.9
(300)	360	845	848	180	1	41.5

Table 1: Continued.

S. no	Cement	Coarse aggregate	Fine aggregate	Water	Superplasticizer	Compressive strength
(301)	514	950	665	180	0.75	52.1
(302)	567	950	648	170	2	62.4
(303)	510	845	731	170	1.5	64.4
(304)	330.85	754.18	875.16	179.74	6.08	52.9
(305)	425	898	816	170	1	46
(306)	360	898	797	180	0.5	40
(307)	437	845	801	170	1	56.8
(308)	300	663	923	175	4	54.69
(309)	412	752	887	182	8.25	73.4
(310)	480	845	655	180	1.25	60.8
(311)	378	898	855	170	1	42.5
(312)	480	950	552	180	0.75	59.6
(313)	389	845	783	170	1.5	55.3
(314)	453	898	659	170	2	62
(315)	377	562	861	227	3.7	56.1
(316)	510	898	679	170	1	63.4
(317)	350	500	1050	196	3.2	43
(318)	600	845	698	180	0.75	59.5
(319)	540	750	900	200	12	78.05
(320)	457	898	816	160	1.75	62.1
(321)	600	950	594	180	0.75	59.7
(322)	400	950	759	180	0.5	42.1
(323)	360	812	813	168	6	56
(324)	198	900	872	154.8	3.44	52
(325)	198	900	872	154.8	3.44	52
(326)	437	950	697	170	0.5	55.9
(327)	389	950	680	170	1	54.3
(328)	570	750	900	200	12	80.42
(329)	450	898	770	180	0	43.8
(330)	360	950	745	180	0.5	39.5
(331)	275	880	691	187	4.4	57.9
(332)	320	845	886	160	0.5	47.7
(333)	405	845	805	180	1	43.6
(334)	457	845	867	160	1.75	62
(335)	500	753	820	192.32	4.808	70.93
(336)	540	750	900	192	12	79.18
(337)	450	950	718	180	0	43.5
(338)	400	898	811	180	0.5	41.5
(339)	411	950	628	180	1.5	45.7
(340)	378	950	803	170	1	41.8
(341)	290	837	913	175.5	3.12	42.7
(342)	400	845	863	180	0.5	41.3
(343)	331.33	739.44	840	180.38	6.08	63.8
(344)	600	898	646	180	1.25	60.5
(345)	440	917	714	176	4.4	69.8
(346)	331.33	768.88	875.16	180.38	5.97	51.53
(347)	320	898	834	160	0.5	48.5
(348)	437	898	749	170	0.5	56.6
(349)	344	881	814	171.85	3.928	48.75
(350)	457	950	764	160	1.25	61.5
(351)	405	950	702	180	0.5	41.5
(352)	533	898	753	160	1.5	69.4
(353)	405	898	754	180	1	43
(354)	533	950	701	160	1	67.8
(355)	510	845	731	170	1	64.4
(356)	375	673	938	175	4	60.8
(357)	360	898	797	180	1	40

performance, and in the end, an empirical model for prediction of strength is also developed.

2. Research Methodology

2.1. Genetic Programming Machine Learning Approach. GP was firstly developed by Jone Koza in 1988, which generates a computer-based model to solve the problem by using the Darwinian selection principle [42]. GP is a predictive tool based on artificial intelligence that develops a program by emulating the progression of living organisms [42]. GP is the generalization form that comes from the genetic algorithm (GA) [43]. These two approaches are somehow different from one another, which is distinguishing based on solution representation. GA represents the solution in the form of a string of numbers (chromosomes), whereas GP represents the solution of given data in the form of a tree-like structure by using the programming language [44]. GA provides linear fixed-length binary strings (chromosomes), whereas GP provides alternative strings of different shapes and sizes of nonlinear entities, thus making GP a versatile approach in the prediction of properties. In other words, the solution of the representation is expressed in the form of a parse tree with varying string size and shape. The hierarchy of problems in GP is similar to GA. The computer program then searches for the optimized solution of the problem in an independent manner [45–47].

The overall chain of GP in solving a problem by programming language consists of the following steps:

(1) Generate and produce individual chromosomes (population set) by selecting in the random way of the problem in the form of function sets and terminal sets. These sets chose their individuals at random and build computer models in tree form with roots (branches) reaching to the end in the terminal set as shown in Figure 1.

(2) The GP algorithm than performing iteratively measures for the selection of best fitness chromosomes and generates new individual chromosomes by three measures, namely, reproduction, mutation, and crossover. GP works in the same way as a human analogy.

  (A) Reproduction: During this procedure, the parts of individuals (chromosomes) are copied without any modification into the next process in a new population [44].

  (B) Crossover: During this operation, a node is randomly selected on one of the roots of each program and the function set with the terminal set of each program is then swapped to create a new offspring program as shown in Figure 2. It can be seen that two new offsprings are generated from two parental computer-based programs [42, 44].

  (C) Mutation: During this procedure, node of individuals in terminal sets and function sets are selected at random and replaced by same parity. This creates new offsprings by randomly choosing sets and best generation appeared in the form of tree as shown in Figure 3 [46].

(3) Genetic programming then finalized its best solution to problem by solving computer based program [48, 49].

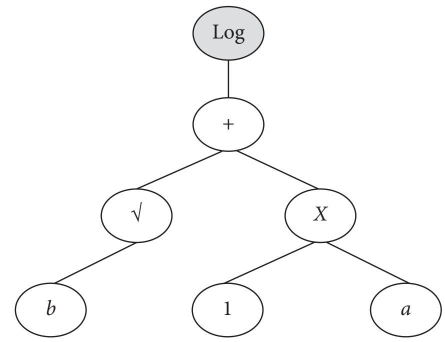


Figure 1: Representation of class tree of GP (Log (sqrt(b) + (1 × a))).

In recent years, approaches like linear genetic programming (LGP), multi expression programming (MEP), and genetic expression programming (GEP) have been used in prediction properties of many domains including civil engineering. These approaches are mainly roots of genetic algorithms and genetic programming. Moreover, these processes diminish the limitation like genetic operation on tree, code growth with complexity, and implementation difficulties. Owing to their extreme benefits, these methods are a favourable candidate in execution complex forecast problems. However, in this paper, genetic expression programming was used for prediction of high-strength concrete.

2.2. Genetic Expression Programming (GEP) Approach. Ferreira [50] proposed a new algorithm, which is the modified development form of GA and GP known as GEP. It incorporates both the linear string of fixed length and parse tree. The linear variant utilizes same genetic operator as used in GP with some minor modifications. The GEP model consists of five parameters having same analogy to GP, i.e., fitness function, terminal set, control parameters, terminal conditions, and function set. GEP algorithm creates population set of randomly selected individual chromosomes and afterward converts each individual into expression tree of different forms (shapes and sizes) to represent its solutions with mathematical expression. Later the target is then compared with the predicted one, and the fitness score of each individual entity is determined. The model stops if it gives best fitness; otherwise, individuals are selected on the basis of roulette wheel sampling. This then extracts the best survival chromosomes from individuals and passes them to the next generation. This loop goes on until the best survival chromosome with adamant fitness score is achieved. The basic step involves in representation of solution is shown in Figure 4.

Each chromosome (gene) of GEP contains a list of symbols with fixed-length variable, arithmetic operations {+,

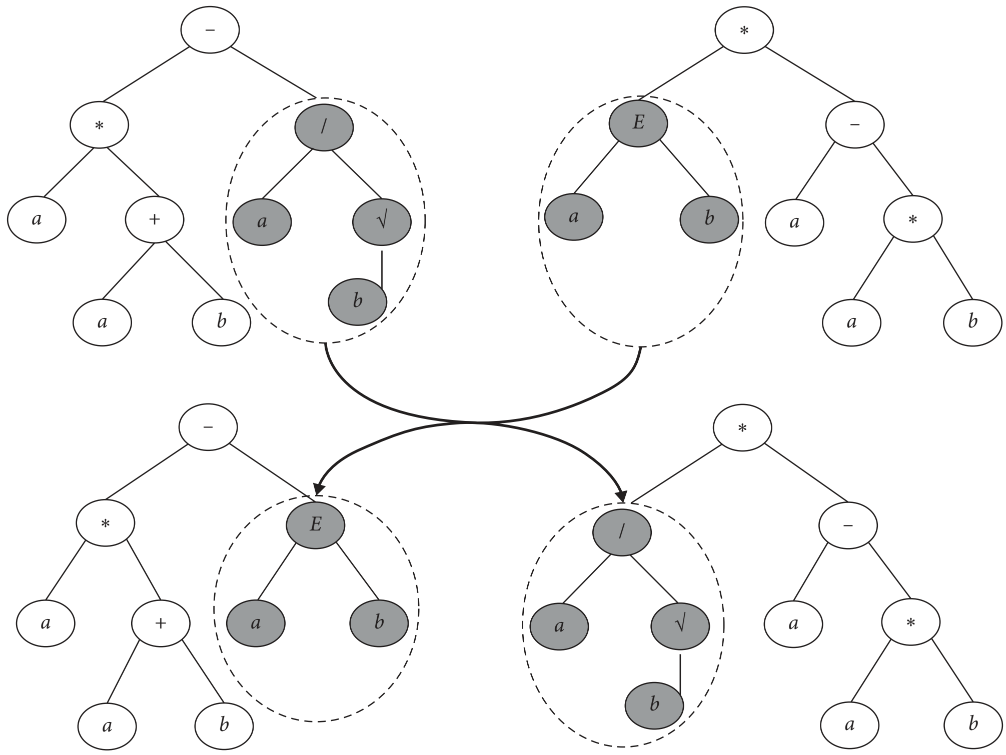


Figure 2: Crossover example based on genetic programming.

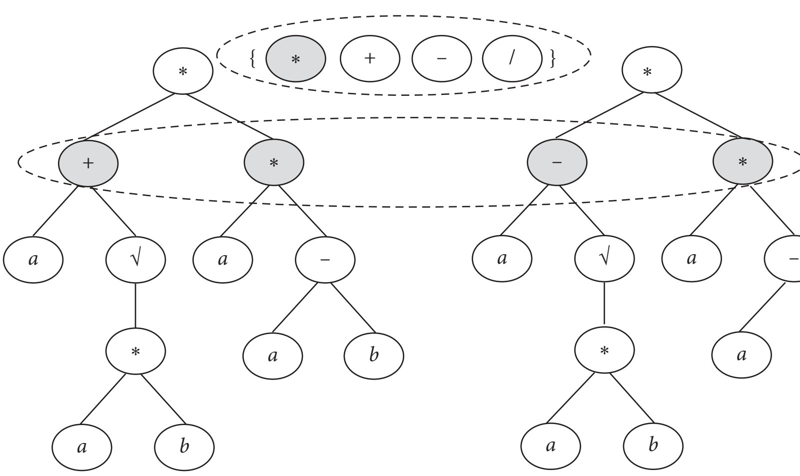


Figure 3: Mutation example based on genetic programming.

×, −, /, sqrt} as set of functions, and constants as terminal sets like {A, B, C, D, 4}. There exists a linear relationship between individual (chromosomes) and function set and terminal set in the genetic code operator. GEP gene with the given function and terminal sets is

$$+.x.\sqrt{A}.\mathrm{A}.-.+.\mathrm{B.A.C.3.B.C.4}, \quad (1)$$

where A, B, C, D are variables (terminal set) and 3, 4 are constants. This term is expressed as K expression (Karva notation) which is used to develop empirical relationship between sets and individual chromosomes [51]. This Karwa expression can also be represented by expression tree (ETs) diagram [52]. For example, the ETs diagram of above mentioned expression is expressed in Figure 5. The transformation of K-notation to ETs starts from the first position which resembles to the roots of ETs and continues through the string [29]. Similarly ETs also transform into K-expression by recording the ties from the base level to the adamant deepest layer. The GEP gene in mathematical form can also be expressed as

$$A(A+3)-(BXC)+\sqrt{(B+A)}. \quad (2)$$

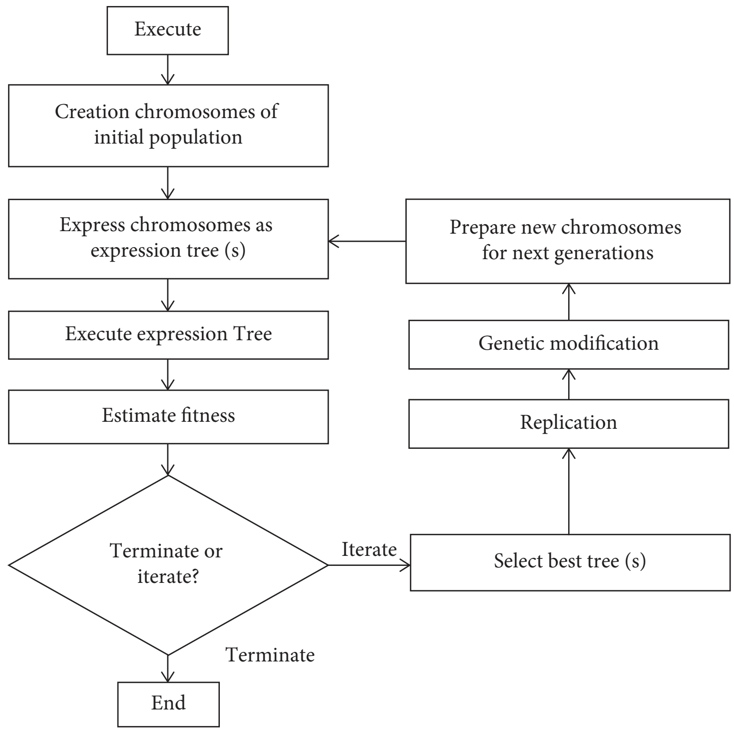


Figure 4: Schematic flowchart of the GEP algorithm [15, 16].

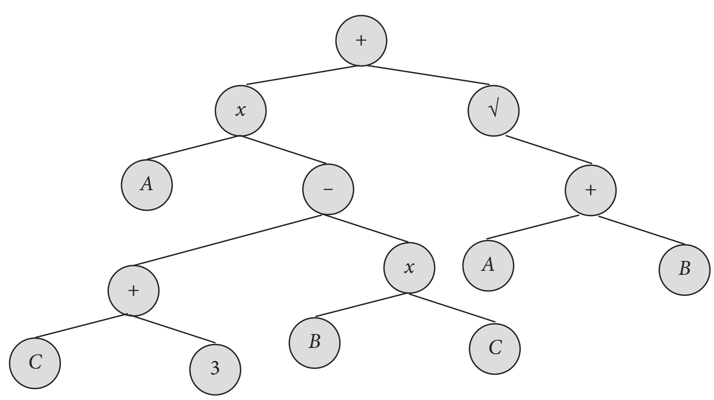


Figure 5: Example tree-expression (ETs).

3. Representation of Experimental Data

3.1. Experimental Datasets. In this paper, 357 data samples have been utilized in modelling of high-strength concrete, which was acquired from previously published papers (see Table 1). However, the aim is to utilize these values to predict the optimized quantities rather than going for hit and trial in experimental work. The database consisting of 357 samples is randomly divided into sets of training, validation, and testing. This scaling is mainly done in machine learning approaches to avoid the overfitting of data, giving us more reliable results in the determination of coefficient (R^2). Moreover, training is done to train the model for the upcoming validation aspect, and in the end, testing was mainly done on unseen data for forecasting of high-strength concrete properties. Out of 357 datasets, 251 (70% data) were assigned to training set and remaining 53 (15%) data to testing and validation sets [53, 54].

3.2. Python Measures for Presenting Database. Representation of the database was done by using anaconda based python programming version 3.7. The data obtained from literature consist of five parameters starting from cement, water, fine aggregate, and coarse aggregate with superplasticizer concentration in the modelling of strength. Every parameter has an influence on strength properties. Python measures were done to find the correlation of each variable to its compressive strength and also to find the optimal dosage and influential effect of variables by conducting

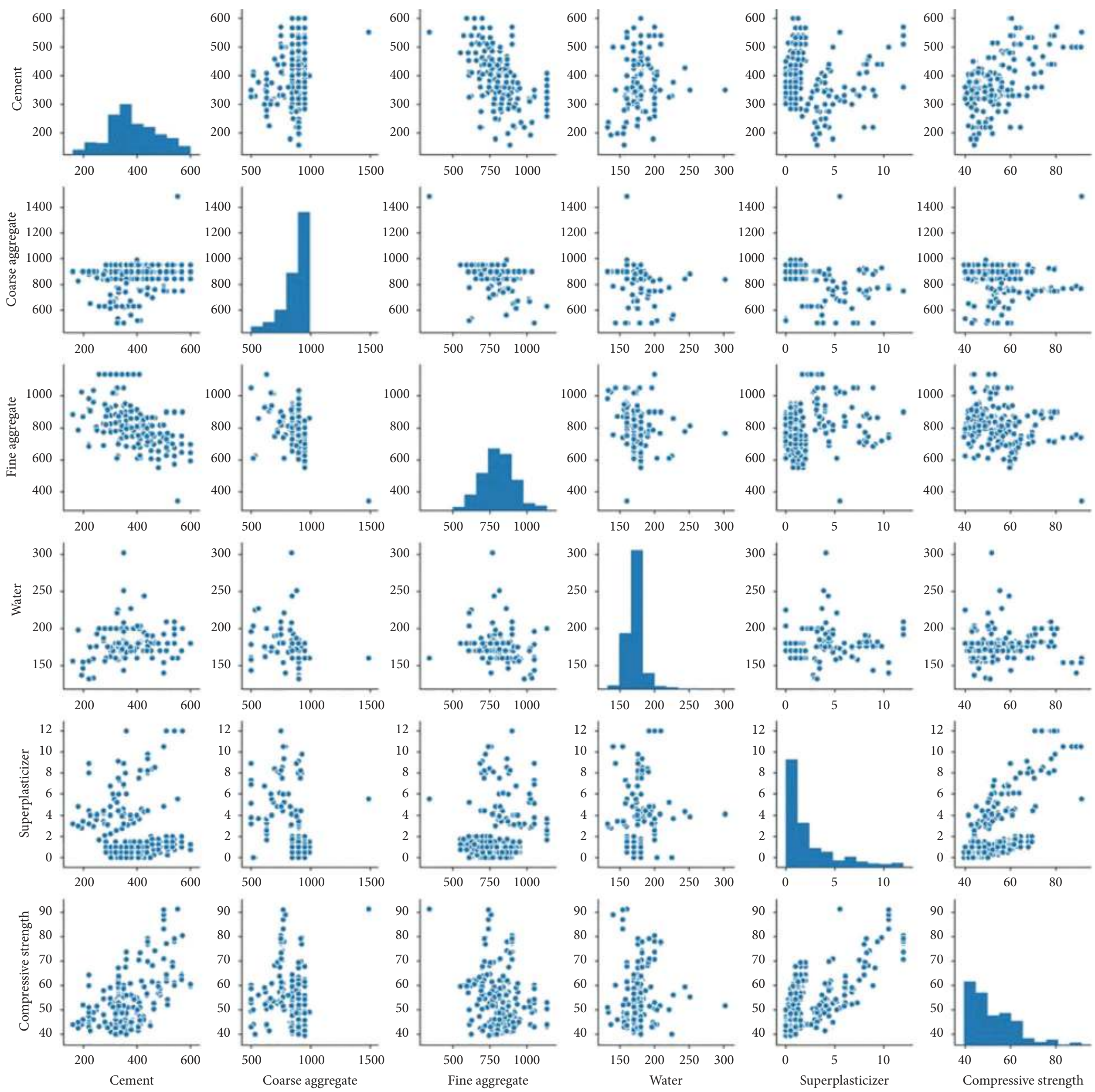


Figure 6: Relation of input variable to its output strength.

permutation features importance. The correlation and distribution with of the variables are shown in Figure 6. It is well stated that model performance is adamantly affected by its variables [55]. Deep leaning is a handful tool in neuron-based artificial approach to predict the mechanical properties by knowing its actual concentration of variables. Python deals with machine learning approach and this correlation plot is made by using seaborn command. The description of data variable used in model is listed in Table 2.

3.2.1. Design of HSC Using Python. This section deals with the parameters in the process of gaining its optimal goal. It is important to state that variables in the modelling of any model have an adamant and significant role in determining its goal. So, the variable study is conducted by using python programming.

(1) Contour Maps. Five contour plots obtained from the python model is illustrated in Figures 7(a)–7(j). As previously mentioned, that model performance is dependent on its variables, so the optimal quantity of variables is important to know rather than using experimental work. This provides us a useful graph to predict the strength at 28 days.

(a) *Effect of Binder on Compressive Strength*. Binder is an adamantly important variable in the domain of civil engineering. It provides strength and setting to cement. Figure 7(a) shows the effect of binder to

Table 2: Statistical description of data in model (kg/m^3).

Parameters	Cement	Fine/coarse aggregate	Water	Superplasticizer
(a) Total data				
Mean	384.35	0.964	173.57	2.35
Standard error	4.92	0.014	0.82	0.14
Median	360	0.922	170	1.25
Mode	360	1.018	170	1
Standard deviation	93.01	0.261	15.56	2.69
Sample variance	8650.50	0.068	242.20	7.25
Kurtosis	0.36	6.453	15.59	2.88
Skewness	0.15	2.129	2.46	1.80
Range	440	1.870	170.08	12
Minimum	160	0.230	132	0
Maximum	600	2.1	302.08	12
Sum	137212.84	344.08	61963.8	837.61
Count	357	357	357	357
(b) Train data				
Mean	383.29	0.971	173.73	2.43
Standard error	6.07	0.017	1.09	0.17
Median	360	0.922	170	1.38
Mode	320	1.018	170	1
Standard deviation	95.95	0.276	17.18	2.75
Sample variance	9206.58	0.076	295.08	7.55
Kurtosis	0.60	5.826	14.42	2.97
Skewness	0.20	2.089	2.49	1.83
Range	420	1.870	170.08	12
Minimum	180	0.230	132	0
Maximum	600	2.1	302.08	12
Sum	95823.1	242.79	43431.75	606.44
Count	250	250	250	250
(c) Validation data				
Mean	387.04	0.922	172.19	1.98
Standard error	12.47	0.025	1.35	0.33
Median	400	0.903	170	1
Mode	360	0.756	170	1
Standard deviation	95.76	0.189	10.36	2.56
Sample variance	9170.57	0.036	107.26	6.55
Kurtosis	0.23	6.821	0.18	4.75
Skewness	0.18	1.663	0.34	2.19
Range	440	1.221	45.2	12
Minimum	160	0.581	154.8	0
Maximum	600	1.802	200	12
Sum	22835.54	54.380	10159.18	117.09
Count	59	59	59	59
(d) Test data				
Mean	390.53	0.909	173.07	2.11
Standard error	12.58	0.022	1.21	0.35
Median	378	0.903	175	1
Mode	360	1.041	180	0.5
Standard deviation	89.87	0.156	8.67	2.47
Sample variance	8076.30	0.024	75.21	6.12
Kurtosis	1.08	0.527	−0.18	2.17
Skewness	0.17	0.612	−0.63	1.66
Range	440	0.733	38.32	10.5
Minimum	160	0.661	154	0
Maximum	600	1.394	192.32	10.5
Sum	19916.87	46.346	8826.8	107.57
Count	51	51	51	51

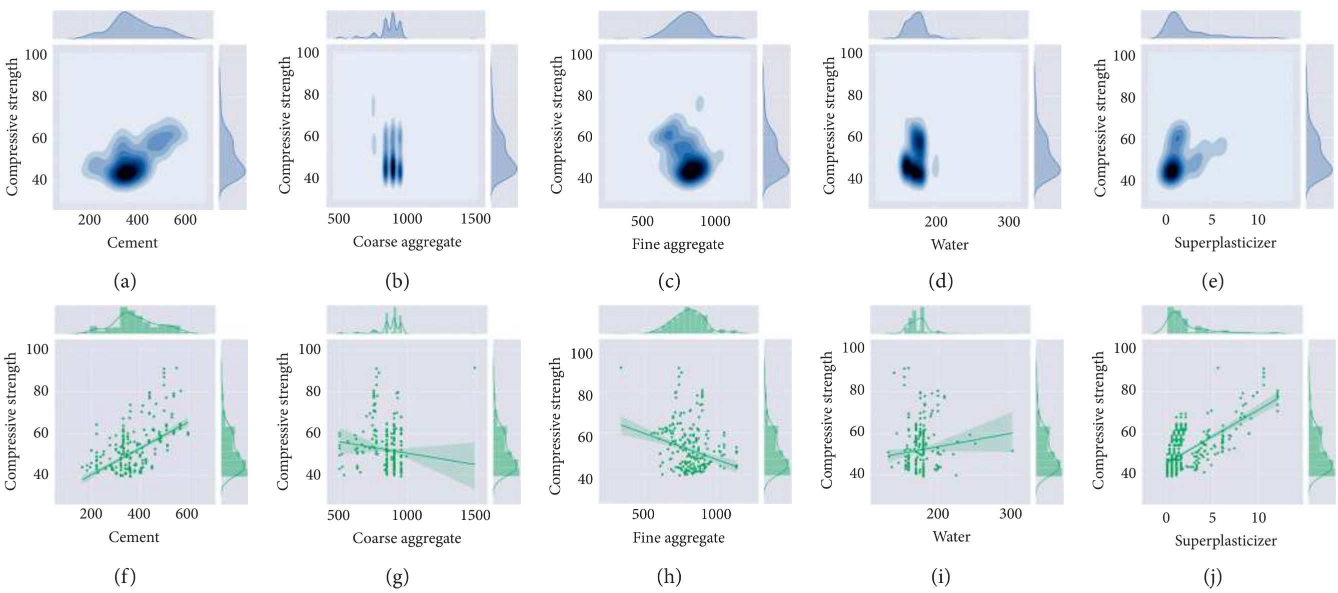


FIGURE 7: (a–j) Contour plots of input variables with the regression graph.

TABLE 3: Range of input and output variables.

Parameters	Abbreviation	Minimum	Maximum
Input variables			
Cement	CEM	160	600
Fine/Coarse aggregate ratio	F/C ratio	0.23	2.1
Water	W	132	302.08
Superplasticizer	SP	0	12
Output variable			
Compressive strength	F_c	39.5	91.3
Predicted compressive strength	F_c	38.33	92.8

compressive strength in the form of contour giving us the required quantity of cement and Figure 7(f) shows the regression graph of cement versus strength. It can be seen that maximum data point used in the literature lies between 300 and 400 kg/m^3. However, significant strength was also achieved by the binder in a range of 500 kg/m^3. Moreover, the deep contour of cement lies in the range of 300 to 400 kg/m^3. It is worth mentioning here that machine deep learning provides us the range in achieving our desire goal.

(b) *Effect of Fine and Coarse Aggregate on Compressive Strength*. Fine and coarse aggregate is used to fill the void and to impart strength in making concrete, however, their concern dosage, type, and condition will affect concrete strength. It is clear from Figures 7(b) and 7(c) that maximum strength was achieved, when using coarse and fine aggregate in the range of about 800 to 1000 and 800 to 900 kg/m^3, respectively. Moreover, Figures 7(g) and 7(h) correlate strength with aggregate.

(c) *Effect of Water and Superplasticizer on Compressive Strength*. Water and superplasticizer have major influence on its strength. Water quantity has direct and indirect relation to strength. Moreover, superplasticizer dosage is used to alter the quantity of water in strength achievement. Figures 7(a)–7(j) represent the required values graphs and these values with their range are also reported in Table 3. In other words, using this much of concentration in HSC yields maximum output, thus eliminating its need for using experimental work.

4. Development of Model Using Gene Expression

This paper aims to develop a generalized equation for the compressive strength of high-strength concrete. Therefore, a set of terminals and function set is used. These variables and function sets have an adamant effect on the performance of the model. For modelling strength of HSC, four variables are selected as input parameters in gene expression programming d_0: cement, d_1: fine to coarse aggregate, d_2: water, and d_3: superplasticizer. Simple division multiplication summation and subtraction operation are used as the function set in model setup. Therefore, the mechanical strength of HSC is dependent on the given relation (see equation (3)):

$$f_c' = f\left(\text{cement}, \frac{\text{fine aggregate}}{\text{coarse aggregate}}, \text{water, superplasticizer}\right). \tag{3}$$

The selection of variables has significant effect in generalization fitness of the GEP-based model. The variables used in the model are presented in Table 4. The model time is controlled by the basic arithmetic process, head size, chromosomes, population size, and complexity. It is better to select those sets which will give a generalized model in due time. Furthermore, the selection of these sets was determined by using hit and trial basis. The model performance is done by utilizing (RMSE) error. Afterward, GEP evaluates its model by presenting architectures structure with head size and number of genes [53].

5. Model Performance Analysis

The performance of any model in learning, training, and testing set is evaluated by the coefficient of determination (R^2) and also by using regression measures and error like relative root mean square error (RMSE), means absolute error (MAE), relative mean square error (RSE), and relative root mean square error (RRMSE). The calculated expressions are given as equations for these error functions which are listed below:

$$\begin{aligned}
\text{RMSE} &= \sqrt{\frac{\sum_{i=1}^{n}\left(ex_i - mo_i\right)^2}{n}},\\
\text{MAE} &= \frac{\sum_{i=1}^{n}\left|ex_i - mo_i\right|}{n},\\
\text{RSE} &= \frac{\sum_{i=1}^{n}\left(mo_i - ex_i\right)^2}{\sum_{i=1}^{n}\left(\overline{ex} - ex_i\right)^2},\\
\text{RRMSE} &= \frac{1}{e}\sqrt{\frac{\sum_{i=1}^{n}\left(ex_i - mo_i\right)^2}{n}},\\
R &= \frac{\sum_{i=1}^{n}\left(ex_i - \overline{ex}_i\right)\left(mo_i - \overline{mo}_i\right)}{\sqrt{\sum_{i=1}^{n}\left(ex_i - \overline{ex}_i\right)^2\sum_{i=1}^{n}\left(mo_i - \overline{mo}_i\right)^2}},\\
\rho &= \frac{\text{RRMSE}}{1+R},
\end{aligned} \tag{4}$$

where ex_i, mo_i are experimental actual strength and model strength, whereas $\overline{ex}_i$ and $\overline{mo}_i$ are average values of experimental and predicted outcome, respectively. The accuracy of the model is defined by its determination of coefficient (R^2). For the effective model, its value should be close to 1 and a value greater than 0.8 presents a high accuracy of the model [56]. This value shows the correlation between experimental and predicted outcomes. An R^2 value close to 1 and lower values of errors (MAE, RRMSE, RMSE, and RSE) indicate higher accuracy of the model. Moreover, an output index or performance index (ρ) is proposed to measure model efficiency as a result of both R^2 and RRMSE [55]. Lower value of the index indicates better performance of the model between experimental and prediction outcomes.

Table 4: Input parameters assigned in GEP model.

Parameters	Settings
General	f_c'
Genes	4
Chromosomes	30
Linking function	Addition
Head size	10
Function set	+, −, ×, ÷
Numerical constants	
Constant per gene	10
Lower bound	−10
Data type	Floating number
Upper bound	10
Genetic operators	
Mutation rate	0.00138
Inversion rate	0.00546
RIS transposition rate	0.00546
IS transposition rate	0.00546
One-point recombination rate	0.00277
Gene recombination rate	0.00277
Two-point recombination rate	0.00277
Gene transposition rate	0.00277

In deep and machine learning approaches, overfitting of data is a major concern. To counter fall this, researchers used objective function (OBF) for their model accuracy (equation (5)) OBF takes the overall data with error and regression coefficient into it to give the best-generalized model [55]. This is achieved by the following equation as presented by Gandomi et al. higher value of R and lower values of errors result in a significantly lower value of index and OBF.

$$\text{OBF} = \left\langle\frac{n_{\text{Train}} - n_{\text{Test}}}{n}\right\rangle\rho_{\text{Train}} + 2\left(\frac{n_{\text{Test}}}{n}\right)\rho_{\text{Test}}. \tag{5}$$

6. Results and Discussion

6.1. Formulation of Compressive Strength of HSC Using GEP. Genetic expression algorithm is used to predict the mechanical response of HSC in the form of empirical relation. This formulation is the function of variables expressed in equation (6). Expression resulting in the form of a relationship comes from expression trees as shown in Figure 8. It can be seen that GEP used both linear

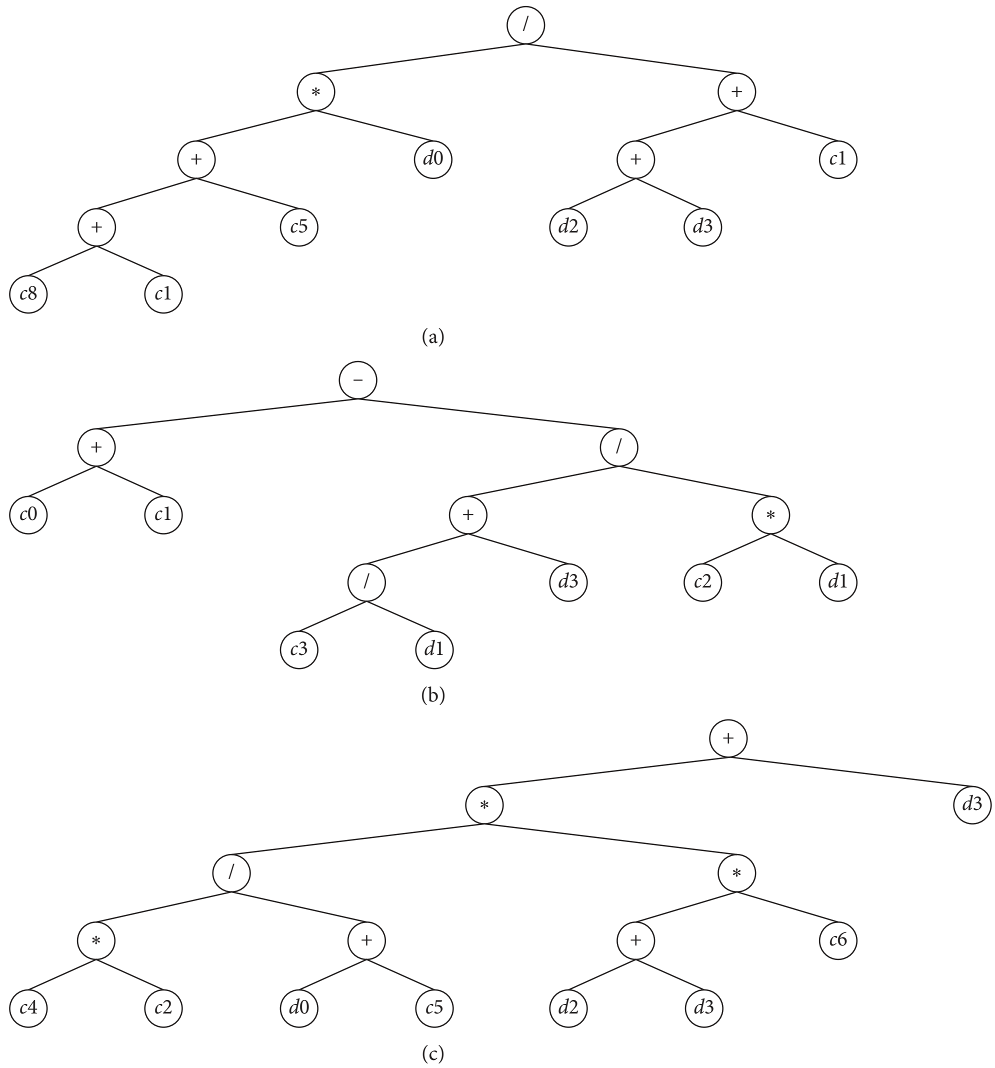


Figure 8: Gene expression tree on high-strength concrete. (a) Sub-ET 1. (b) Sub-ET 2. (c) Sub-ET 3.

as well as nonlinear algorithms by forming a tree structure. Moreover, this complex architectural tree utilizes arithmetic operators, variables, and somehow constants in prediction of strength. Basic operator is employed by GEP in solving three sets of expressions. Each sub program or chromosomes reflect specific features of the problem, which in turn develops functionalized solution to the problem [50].

$$f_c\,(\text{MPa}) = X + Y + Z, \tag{6}$$

$$\begin{aligned} X &= \left(\frac{((9.77 + 15.30) + (-5.11)) * \text{cement}}{(\text{water} + \text{superplasticizer}) + 15.30}\right) + Y \\ &= \left((-5.30 + (-2.40)) - \left(\frac{(-(0.59/(F/C)\ \text{agg})) + \text{superplasticizer}}{-0.50} * \left(\frac{F}{C}\ \text{agg}\right)\right)\right) + Z \\ &= \left(\left(-0.76 * \frac{-4.76}{\text{cement} + 32.4} * ((\text{water} + \text{superplasticizer}) * 8.65)\right) + \text{superplasticizer}\right). \end{aligned} \tag{7}$$

The structural gene, number of chromosomes, and operators are selected prior running the GEP algorithm. The best selection of model is based on several trials by varying its head size, gene numbers, and chromosomes with operational operators. The GEP algorithm selects the best generation and gene within the population set. Figure 8 presents the best

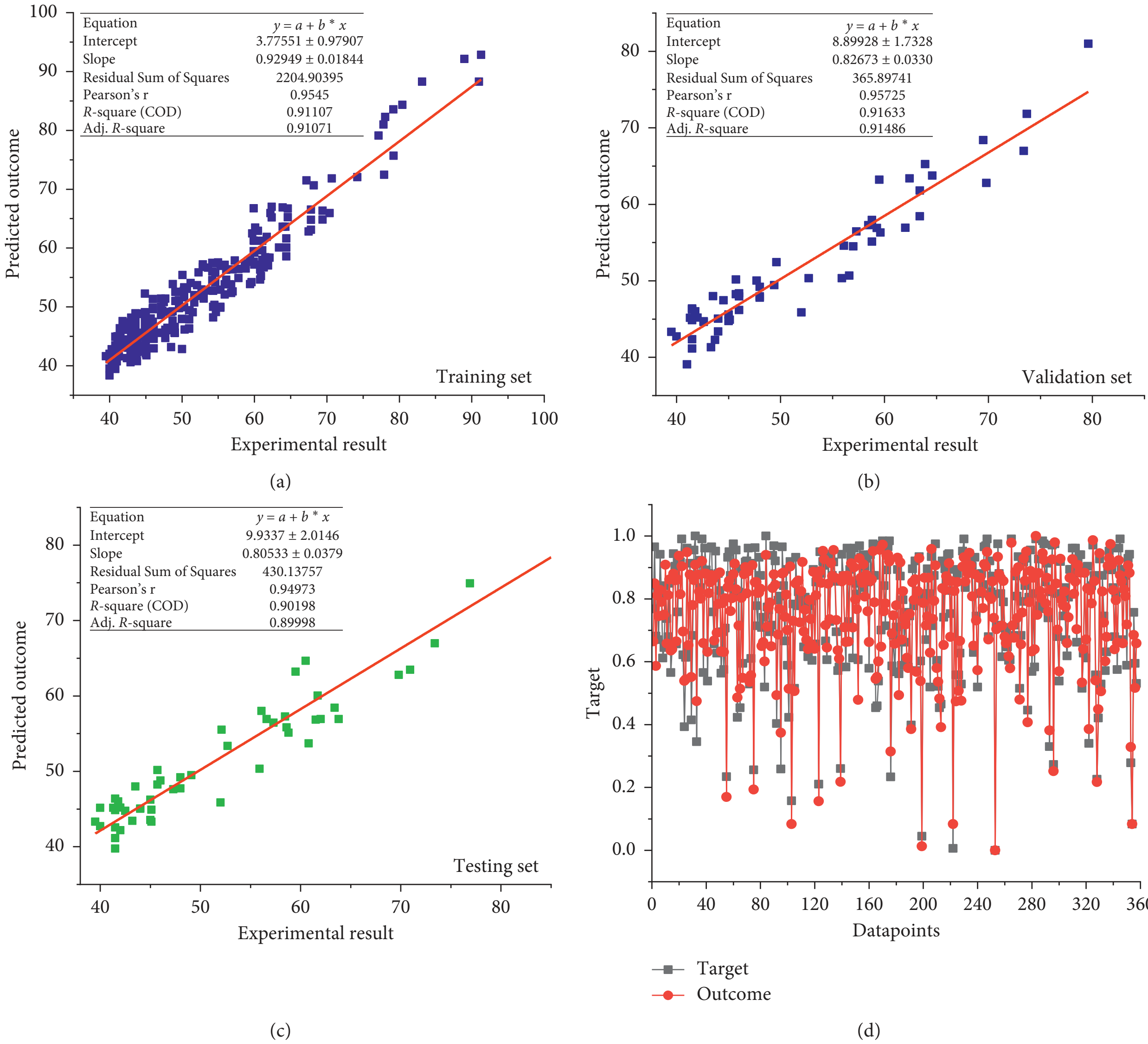


FIGURE 9: (a–c) Regression analysis of model; (d) normalized graph of all data.

outcome of f_c. It can be seen that linkage function employed in GEP is the basic operator in which c represents constant values and d represents the input variables. The basic fitness function used in modelling perspective is RMSE.

6.2. Evaluation of Model and Analysis. The evaluation of model between actual and predicted one is shown in Figure 9. It is clear that the GEP-based algorithm in prediction aspect is a prominent tool in assessment of strength. It can be seen that the regression line for data samples in training set, testing, and validation set approaches to 1. Model accuracy and validity can be judged by its coefficient of determination (R^2). Figures 9(a)–9(c) represent the model accuracy by depicting its R^2 value greater than 0.8; however, in our case, it is 0.910, 0.914, and 0.9 for testing training and validation set, respectively. These sets consist of approximately 360 data samples, out of which testing training and validation set consist of 70/15/15 data points. This outfitted data modelled in the GEP algorithm indicate good relation between output and target values. Moreover, normalization of data was also done to give a generalized relation in the range of 0 and 1. The model accuracy of overall data can also be seen in the normalized graph as shown in Figure 9(d).

The model performance can also be evaluated by checking from statistical analysis such as MAE, R^2, and RRMSE with RSE. The statistical measures of the proposed GEP-based model for testing, training, and validation set are shown in Table 5. Moreover, further analysis can be done by determining covariance (COV) and standard deviation (SD) of predicted to actual targets. Values of covariance and SD of training set are 0.16 and 0.059, respectively. The statistical analysis gives an accurate idea of model accuracy by its R^2 and error values with the adamant low objective function. Furthermore, the model accuracy can also be judged from its R^2 and statistical error values of all sets. Thus, proposed model give high accuracy of actual and predicted values.

Table 5: Statistical calculation of the proposed models.

Model		RMSE			MAE			RSE	
	Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
	3.057	1.42	1.62	2.602	0.575	0.595	0.089	0.092	0.023
		RRMSE			R^2			P (row)	
F_c	Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
	0.058	0.0286	0.031	0.954	0.957	0.031	0.03	0.014	0.015

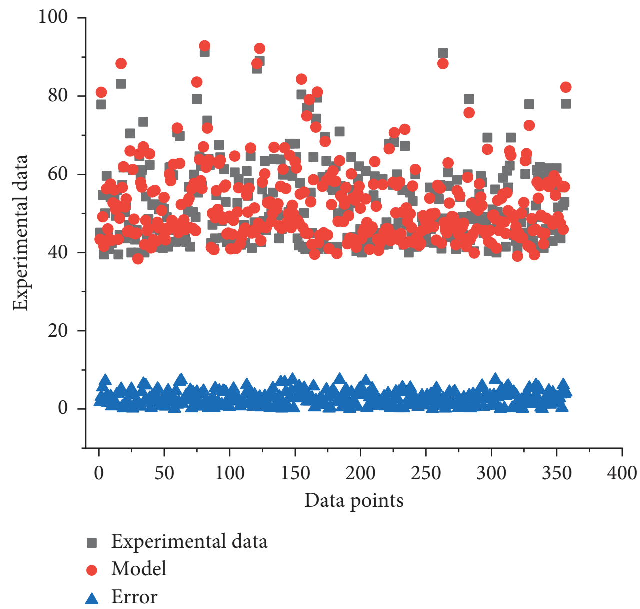


Figure 10: Relation of actual target to outcome values of strength.

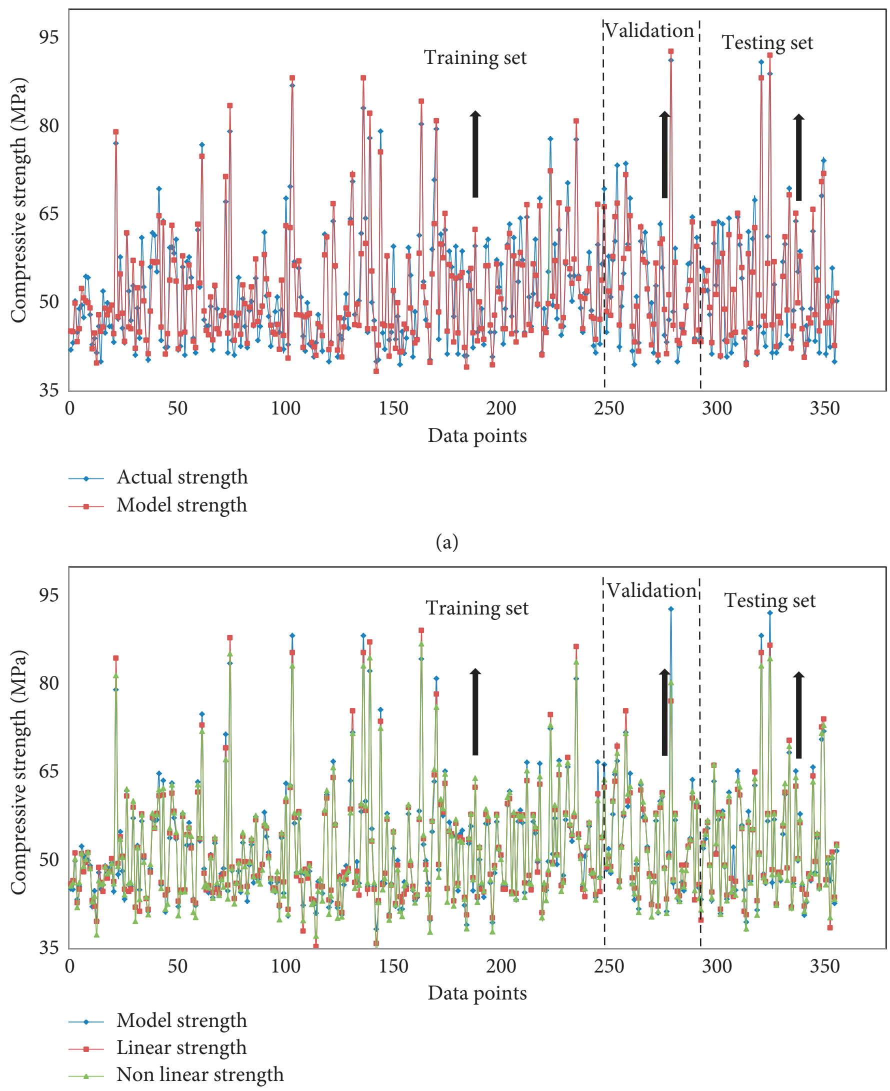
The accuracy of the proposed model in a broader aspect can also be evaluated by checking the absolute error difference between predicted and actual targets as shown in Figure 10. It is adamantly clear from the figure showing its accuracy between predicted and actual ones with maximum average error of 2.64. Majority of the predicted data lie in the range of 0.029 MPa to 7.5 MPa. These values are of absolute error with minimum and maximum of predicted datasets. Moreover, adamant difference between experimental and model values with less error depicts the adaptive nature of gene expression programming.

The reliability of any model is greatly dependent on its data set. Adamant data point increases the accuracy of the model with input variables. However, the validity of data to variables in relation making is quite a major concern in its modelling. To counter fall and to check the validation of the dataset, Frank and Todeschini [57] stated that the ratio of input data set to its variables should be equal to 5. This scenario presented by the author is for an ideal model. However, the current paper significantly outfits this ratio which is equal to $357/4 = 89.25$ as compared to the available literature. Moreover, validation of the GEP model can also be checked by external statistical measures on the testing set. Golbraikh and Tropsha [58] proposed a generalized relationship that the slope of line regression (k' or k) in the model should approach to 1. Similarly, various scholars have suggested that the squared relationship coefficient (origin) between the output and target values (Ro'^2) or the coefficient amongst expected and tentative values (Ro^2) should be near to 1 [44]. These external checks on the GEP-based model are presented in Table 6. Hence, it can be concluded that the models hold the expectation capability which is not just a connection amongst the input and output variables.

The prediction of the mechanical behaviour of high-strength concrete by genetic expression algorithm is adamantly reliable in using data samples to its variables. The behaviour of the GEP based model can also be compared with the linear and nonlinear based model by presenting an empirical relationship between predicted and experimental results. The empirical relation of both results in the form of expression is shown in equations (5) and (6). Moreover, Figure 11 represents the behaviour of modelled data. It can be seen that GEP based model outfits in data presentation of testing, validation, and training set from linear and nonlinear ones with R greater than both modelled [59]. This is due to one of the advantages of GEP, as it takes both linear and nonlinear data into its database which ultimately generates accuracy of predicted data by showing expression tree and then simplified its data by decoding it in the form of the generalized equation as shown in Figure 8. Moreover, its simplest nature can help researchers to calculate the compressive

Table 6: Statistical variables of GEP models from externally validation.

S. no	Equation	Condition	Model
1	$k = \sum_{i=1}^{n} (e_i \times m_i)/e_i^2$	$0.85 < k < 1.15$	0.98
2	$k' = \sum_{i=1}^{n} (e_i \times m_i)/m_i^2$	$0.85 < k < 1.15$	1.00
3	$R_o^2 - (\sum_{i=1}^{n} (m_i - e_i^o)^2 / \sum_{i=1}^{n} (m_i - m_i^o)^2)$, $e_i^o = k \times m_i$	$R_o^2 \cong 1$	0.97
4	$R_o^{'2} - (\sum_{i=1}^{n} (e_i - m_i^o)^2 / \sum_{i=1}^{n} (e_i - e_i^o)^2)$, $m_i^o = k' \times e_i$	$R_o^2 \cong 1$	0.99



(a)

(b)

Figure 11: (a) Evaluation of sets with experimental and model data. (b) Evaluation of sets with model set with linear as well as nonlinear approach.

strength by doing hand calculations. These algorithms help in predesign design to forecast prediction close to experimental work [56]. The accuracy of this can also be checked by residual error as shown in Figure 12. It represents the accuracy of data with frequency of data present in GEP model and its regression accuracy.

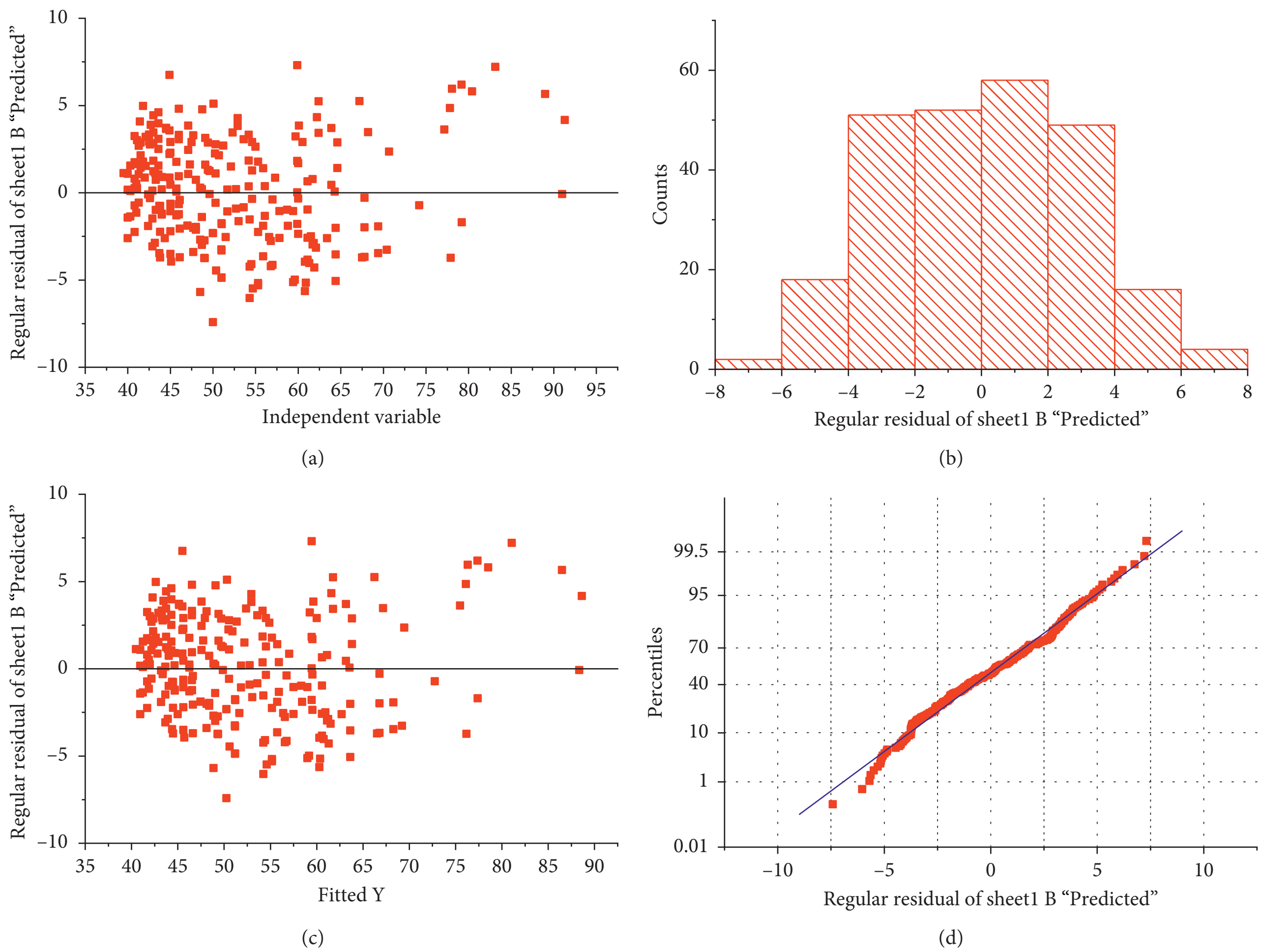


Figure 12: Accuracy of GEP based model with less residual error.

$$f'_c = 0.065\,\text{cement} - 4.9\,\frac{\text{fine}}{\text{coarse}}\,\text{aggregate} - 0.089\,\text{water} - 2.95\,\text{superplasticizer} + 40.1,$$
$$f'_c = 0.0042\,\text{cement}^{1.41} - 27.7\,\frac{\text{fine}}{\text{coarse}}\,\text{aggregate}^{0.16} - 0.5\,\text{water}^{0.73} + 5.03\,\text{superplasticizer}^{0.77} + 72.6. \tag{8}$$

6.3. Compression of GEP Model with Other Model. The performance of the GEP model is compared with other models available in the literature [7, 60, 61]. Al-Shamiri et al. [61] used extreme learning machine (ELM) and compared its model prediction accuracy with back-propagation neural network (BP-NN). The authors predict R^2 of testing set of about 0.9937 and 0.9938 for ELM and BP-NN [61]. Similarly, Öztaş et al. [7] predicted the compressive strength and slump of HSC using neural network. The author reported strong correlation between input and output result of testing set which is about 0.99 for both slump test and compressive strength. Baykasoğlu et al. [60] predicted the parameters of high-strength concrete using machine learning techniques. Regression analysis, genetic engineering programming, and neural network were first employed to make generalized equation. Afterwards, a multiobjective optimization model is made to predict the outcome and comparison was also made between prediction and optimization results. Singh et al. [62] predicted the compressive strength of HSC using random forest and M5P techniques. The authors achieved a good relation by using random forest rather than M5P which is $R^2 = 0.876$ and 0.814 for testing set, respectively. It can be seen that prediction of HSC was evaluated using different approaches but none of the method give a diesrable equation which predicts the strength by using hand calculation. Thus, employing GEP approach gives not only $R^2 = 0.90$ but also an equation with parameters involved.

7. Conclusion

The machine learning approach provides adamant accuracy between the modelled and experimental data. This will help in the predesign phase rather than conducting experimental tests by doing trials. The following conclusion has been drawn by utilizing GEP.

(i) Artificial intelligence using anaconda Jupiter notebook python-based is conducted on the input variables with compressive strength. This programming technique provides the optimal values of all these influential variables which will help the researcher to design their experimental work by just taking these optimized values.

(ii) GEP approach provides a simplified formulation of compressive strength with adamant accuracy between modelled and experimental results. This shows its diversity by considering linear and nonlinear data.

(iii) The statistical analysis gives significant accuracy between training, testing, and validation set with the coefficient of determination greater than 0.9. Moreover, an error like MAE, RSE, and RRMSE shows low value with high *R*-value. This contra values adamantly provides the accuracy of modelled data.

(iv) The GEP model is compared with linear analysis and nonlinear analysis. However, GEP model outfits both analyses. Moreover, the current model was also compared with other published models, but GEP model gave us the required equation which helps in prediction with current parameters via hand calculations

(v) Permutation feature importance was done by using python on variables to show the influential one in the modelling aspect. In another word, which parameter influences the compressive strength of HSC is check by PMI.

Data Availability

The data used in the study were collected from different research papers in modelling aspect.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research project was supported by the deanship of scientific research at Prince Sattam bin Abdulaziz University under the research project no. 2020/01/16810.

References

[1] X. Zhang and J. Han, "The effect of ultra-fine admixture on the rheological property of cement paste," *Cement and Concrete Research*, vol. 30, no. 5, pp. 827–830, 2000.

[2] A. Khaloo, M. H. Mobini, and P. Hosseini, "Influence of different types of nano-SiO_2 particles on properties of high-performance concrete," *Construction and Building Materials*, vol. 113, pp. 188–201, 2016.

[3] R. D. Hooton and J. A. Bickley, "Design for durability: the key to improving concrete sustainability," *Construction and Building Materials*, vol. 67, pp. 422–430, 2014.

[4] R. Carrasquillo, A. Nilson, and F. O. Slate, "Properties of high strength concrete subjectto short-term loads," *Journal Proceedings*, vol. 78, no. 3, pp. 171–178, 1981.

[5] K. Celik, C. M. Akgul, A. P. Gursel, P. K. Mehta, A. Horvath, and P. J. M. Monteiro, "Mechanical properties, durability, and life-cycle assessment of self-consolidating concrete mixtures made with blended portland cements containing fly ash and limestone powder," *Cement and Concrete Composites*, vol. 56, pp. 59–72, 2017.

[6] M. Mbessa and J. Péra, "Durability of high-strength concrete in ammonium sulfate solution," *Cement and Concrete Research*, vol. 31, no. 8, pp. 1227–1231, 2001.

[7] A. Öztaş, M. Pala, E. Özbay, E. Kanca, N. Çağlar, and M. A. Bhatti, "Predicting the compressive strength and slump of high strength concrete using neural network," *Construction and Building Materials*, vol. 20, pp. 769–775, 2006.

[8] B. Persson, "A comparison between mechanical properties of self-compacting concrete and the corresponding properties of normal concrete," *Cement and Concrete Research*, vol. 31, no. 2, pp. 193–198, 2001.

[9] M. Sahmaran, A. Yurtseven, and I. Ozgur Yaman, "Workability of hybrid fiber reinforced self-compacting concrete," *Building and Environment*, vol. 40, no. 12, pp. 1672–1677, 2005.

[10] Z. J. Grdic, G. A. Toplicic-Curcic, I. M. Despotovic, and N. S. Ristic, "Properties of self-compacting concrete prepared with coarse recycled concrete aggregate," *Construction and Building Materials*, vol. 24, no. 7, pp. 1129–1133, 2010.

[11] F. Farooq, S. U. Rehman, H. Hashim, and H. Khan, "Dispersion of multi wall carbon nano tubes using hybrid surfactants," in *Proceedings of the 1st Conference on Sustainability in Civil Engineering*, Islamabad, Pakistan, August 2019.

[12] F. Farooq, A. Akbar, and R. Arsalan Khushnood, "Effect of hybrid carbon nanotubes/graphite nano platelets on mechanical properties of cementitious composite," in *Proceedings of the 1st Conference on Sustainability in Civil Engineering*, Islamabad, Pakistan, August 2019.

[13] A. Akbar, F. Farooq, M. Shafique, F. Aslam, R. Alyousef, and H. Alabduljabbar, "Sugarcane bagasse ash-based engineered geopolymer mortar incorporating propylene fibers," *Journal of Building Engineering*, vol. 33, Article ID 101492, 2021.

[14] F. Farooq, S. K. U. Rahman, A. Akbar et al., "A comparative study on performance evaluation of hybrid GNPs/CNTs in conventional and self-compacting mortar," *Alexandria Engineering Journal*, vol. 59, no. 1, pp. 369–379, 2020.

[15] M. F. Javed, M. N. Amin, M. I. Shah et al., "Applications of gene expression programming and regression techniques for estimating compressive strength of bagasse ash based concrete," *Crystals*, vol. 10, no. 9, p. 737, 2020.

[16] M. F. Javed, F. Farooq, S. A. Memon et al., "New prediction model for the ultimate axial capacity of concrete-filled steel tubes: an evolutionary approach," *Crystals*, vol. 10, no. 9, p. 741, 2020.

[17] F. Farooq, A. Akbar, R. A. Khushnood, W. L. B. Muhammad, S. K. U. Rehman, and M. F. Javed, "Experimental investigation of hybrid carbon nanotubes and graphite nanoplatelets on rheology, shrinkage, mechanical, and microstructure of SCCM," *Materials*, vol. 13, no. 1, p. 230, 2020.

[18] P. G. Asteris and K. G. Kolovos, "Self-compacting concrete strength prediction using surrogate models," *Neural Computing and Applications*, vol. 31, no. S1, pp. 409–424, 2019.

[19] R. Siddique, P. Aggarwal, and Y. Aggarwal, "Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks," *Advances in Engineering Software*, vol. 42, no. 10, pp. 780–786, 2011.

[20] M. Uysal and H. Tanyildizi, "Estimation of compressive strength of self compacting concrete containing polypropylene fiber and mineral additives exposed to high temperature using artificial neural network," *Construction and Building Materials*, vol. 27, no. 1, pp. 404–414, 2012.

[21] M. Uysal and H. Tanyildizi, "Predicting the core compressive strength of self-compacting concrete (SCC) mixtures with mineral additives using artificial neural network," *Construction and Building Materials*, vol. 25, no. 11, pp. 4105–4111, 2011.

[22] B. Vakhshouri and S. Nejadi, "Predicition of compressive strength in light-weight self-compacting concrete by anfis analytical model," *Archives of Civil Engineering*, vol. 61, no. 2, 2015.

[23] A. Info, "Application of ANN and ANFIS models determining compressive strength of concrete," *Jsoftcivil.Com.*vol. 1, pp. 62–70, 2018.

[24] B. Vakhshouri and S. Nejadi, "Prediction of compressive strength of self-compacting concrete by ANFIS models," *Neurocomputing*, vol. 280, pp. 13–22, 2018.

[25] S. Dutta, A. Ramachandra Murthy, D. Kim, and P. Samui, "Prediction of compressive strength of self-compacting concrete using intelligent computational modeling," *Computers, Materials & Continua*, vol. 53, pp. 167–185, 2017.

[26] B. Vakhshouri and S. Nejadi, "Application of adaptive neuro-fuzzy inference system in high strength concrete," *International Journal of Computer Applications*, vol. 101, no. 5, pp. 39–48, 2014.

[27] Q. Han, C. Gui, J. Xu, and G. Lacidogna, "A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm," *Construction and Building Materials*, vol. 226, pp. 734–742, 2019.

[28] D. Dao, H.-B. Ly, S. Trinh, T.-T. Le, and B. Pham, "Artificial intelligence approaches for prediction of compressive strength of geopolymer concrete," *Materials*, vol. 12, no. 6, p. 983, 2019.

[29] A. I. Nour and E. M. Güneyisi, "Prediction model on compressive strength of recycled aggregate concrete filled steel tube columns," *Composites Part B: Engineering*, vol. 173, Article ID 106938, 2019.

[30] A. F. Bingöl, A. Tortum, and R. Gül, "Neural networks analysis of compressive strength of lightweight concrete after high temperatures," *Materials & Design (1980–2015)*, vol. 52, pp. 258–264, 2013.

[31] I. Azim, J. Yang, M. F. Javed et al., "Prediction model for compressive arch action capacity of RC frame structures under column removal scenario using gene expression programming," *Structures*, vol. 25, pp. 212–228, 2020.

[32] M. F. Iqbal, Q.-f. Liu, I. Azim et al., "Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming," *Journal of Hazardous Materials*, vol. 384, Article ID 121322, 2020.

[33] J.-S. Chou, C.-F. Tsai, A.-D. Pham, and Y.-H. Lu, "Machine learning in concrete strength simulations: multi-nation data analytics," *Construction and Building Materials*, vol. 73, pp. 771–780, 2014.

[34] J.-S. Chou and A.-D. Pham, "Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength," *Construction and Building Materials*, vol. 49, pp. 554–563, 2013.

[35] A. T. A. Dantas, M. Batista Leite, and K. De Jesus Nagahama, "Prediction of compressive strength of concrete containing construction and demolition waste using artificial neural networks," *Construction and Building Materials*, vol. 38, pp. 717–722, 2013.

[36] Z. H. Duan, S. C. Kou, and C. S. Poon, "Prediction of compressive strength of recycled aggregate concrete using artificial neural networks," *Construction and Building Materials*, vol. 40, pp. 1200–1206, 2013.

[37] M. Pala, E. Özbay, A. Öztaş, and M. I. Yuce, "Appraisal of long-term effects of fly ash and silica fume on compressive strength of concrete by neural networks," *Construction and Building Materials*, vol. 21, no. 2, pp. 384–394, 2007.

[38] R. Cook, J. Lapeyre, H. Ma, and A. Kumar, "Prediction of compressive strength of concrete: critical comparison of performance of a hybrid machine learning model with standalone models," *Journal of Materials in Civil Engineering*, vol. 31, 2019.

[39] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks:," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[40] X. Yao, "Evolving artificial neural networks," *Proceedings of the IEEE*, vol. 87, pp. 1423–1447, 1999.

[41] P. Cunningham, J. Carney, and S. Jacob, "Stability problems with artificial neural networks and the ensemble solution," *Artificial Intelligence in Medicine*, vol. 20, no. 3, pp. 217–225, 2000.

[42] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, Cambridge, MA, USA, 1992.

[43] R. Cheng, *Genetic Algorithms and Engineering Design*, Wiley, Hoboken, NJ, USA, 2007.

[44] A. H. Alavi, M. Ameri, A. H. Gandomi, and M. R. Mirzahosseini, "Formulation of flow number of asphalt mixes using a hybrid computational method," *Construction and Building Materials*, vol. 25, no. 3, pp. 1338–1355, 2011.

[45] M. Shi, Q. Lin, Q. Zhang, and H. Zhao, "Hybrid genetic Tabu Search Simulated Annealing algorithm and its application in vehicle routing problem with time windows," in *Proceedings of the 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)*, pp. 4022–4025, Dengleng, China, August 2011.

[46] P. Aminian, M. R. Javid, A. Asghari, A. H. Gandomi, and M. A. Esmaeili, "A robust predictive model for base shear of steel frame structures using a hybrid genetic programming and simulated annealing method," *Neural Computing and Applications*, vol. 20, no. 8, pp. 1321–1332, 2011.

[47] A. Baykasoğlu, T. ürkay Dereli, and S. Taniş, "Prediction of cement strength using soft computing techniques," *Cement and Concrete Research*, vol. 34, pp. 2083–2090, 2004.

[48] A. F. Ashour, L. F. Alvarez, and V. V. Toropov, "Empirical modelling of shear strength of RC deep beams by genetic programming," *Computers & Structures*, vol. 81, no. 5, pp. 331–338, 2003.

[49] J. R. Koza, D. Andre, M. A. Keane, F. H. Bennett III, and J. RKoza, "Evolution of a 60 decibel op amp using genetic programming," in *Evolvable Systems: From Biology to Hardware*, vol. 1259, pp. 453–469, Springer, Berlin, Germany, 1997.

[50] C. Ferreira, "Gene expression programming: a new adaptive algorithm for solving problems," 2001, https://arxiv.org/ftp/cs/papers/0102/0102027.pdf.

[51] A. H. Gandomi and A. H. Alavi, “A new multi-gene genetic programming approach to nonlinear system modeling. Part I: materials and structural engineering problems,” *Neural Computing and Applications*, vol. 21, no. 1, pp. 171–187, 2012.

[52] A. H. Gandomi, A. H. Alavi, M. R. Mirzahosseini, and F. M. Nejad, “Nonlinear genetic-based models for prediction of flow number of asphalt mixtures,” *Journal of Materials in Civil Engineering*, vol. 23, no. 3, pp. 248–263, 2011.

[53] A. Gholampour, A. H. Gandomi, and T. Ozbakkaloglu, “New formulations for mechanical properties of recycled aggregate concrete using gene expression programming,” *Construction and Building Materials*, vol. 130, pp. 122–145, 2017.

[54] M. A. Shahin, H. R. Maier, and M. B. Jaksa, “Data division for developing neural networks applied to geotechnical engineering,” *Journal of Computing in Civil Engineering*, vol. 18, no. 2, pp. 105–114, 2004.

[55] A. H. Gandomi and D. A. Roke, “Assessment of artificial neural network and genetic programming as predictive tools,” *Advances in Engineering Software*, vol. 88, pp. 63–72, 2015.

[56] A. H. Gandomi, S. K. Babanajad, A. H. Alavi, and Y. Farnam, “Novel approach to strength modeling of concrete under triaxial compression,” *Journal of Materials in Civil Engineering*, vol. 24, no. 9, pp. 1132–1143, 2012.

[57] I. Frank and R. Todeschini, “The data analysis handbook,” *Data Handling in Science and Technology*, vol. 14, pp. 1–352, 1994.

[58] A. Golbraikh and A. Tropsha, “Beware of q^2!” *Journal of Molecular Graphics and Modelling*, vol. 20, no. 4, pp. 269–276, 2002.

[59] A. H. Gandomi, S. M. Tabatabaei, M. H. Moradian, A. Radfar, and A. H. Alavi, “A new prediction model for the load capacity of castellated steel beams,” *Journal of Constructional Steel Research*, vol. 67, no. 7, pp. 1096–1105, 2011.

[60] A. Baykasoğlu, A. Öztaş, and E. Özbay, “Prediction and multi-objective optimization of high-strength concrete parameters via soft computing approaches,” *Expert Systems With Applications*, vol. 36, pp. 6145–6155, 2009.

[61] A. K. Al-Shamiri, J. H. Kim, T.-F. Yuan, and Y. S. Yoon, “Modeling the compressive strength of high-strength concrete: an extreme learning approach,” *Construction and Building Materials*, vol. 208, pp. 204–219, 2019.

[62] B. Singh, B. Singh, P. Sihag, A. Tomar, and A. Sehgal, “Estimation of compressive strength of high-strength concrete by random forest and M5P model tree approaches,” *Journal of Materials and Engineering Structures*, vol. 6, no. 4, pp. 583–592, 2019.