CrossMark

Prediction of coal ash fusion temperatures using computational intelligence based models

Sanjeev S. Tambe[1] · Makarand Naniwadekar[1,2,3] · Shishir Tiwary[1] ·
Ashis Mukherjee[4] · Tarit Baran Das[5]

Abstract In the coal-based combustion and gasification processes, the mineral matter contained in the coal (predominantly oxides), is left as an incombustible residue, termed *ash*. Commonly, ash deposits are formed on the heat absorbing surfaces of the exposed equipment of the combustion/gasification processes. These deposits lead to the occurrence of slagging or fouling and, consequently, reduced process efficiency. The *ash fusion temperatures* (AFTs) signify the temperature range over which the ash deposits are formed on the heat absorbing surfaces of the process equipment. Thus, for designing and operating the coal-based processes, it is important to have mathematical models predicting accurately the four types of AFTs namely *initial deformation temperature*, *softening temperature, hemispherical temperature,* and *flow temperature*. Several linear/nonlinear models with varying prediction accuracies and complexities are available for the AFT prediction. Their principal drawback is their applicability to the coals originating from a limited number of geographical regions. Accordingly, this study presents computational intelligence (CI) based nonlinear models to predict the four AFTs using the oxide composition of the coal ash as the model input. The CI methods used in the modeling are *genetic programming* (GP), *artificial neural networks*, and *support vector regression*. The notable features of this study are that the models with a better AFT prediction and generalization performance, a wider application potential, and reduced complexity, have been developed. Among the CI-based models, GP and MLP based models have yielded overall improved performance in predicting all four AFTs.

✉ Sanjeev S. Tambe
ss.tambe@ncl.res.in

1 Chemical Engineering and Process Development Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pashan, Pune 411008, India

2 Chemical Engineering Department, All India Shri Shivaji Memorial Society's College of Engineering (AISSMSCOE), Kennedy Road, Pune 411001, India

3 Savitribai Phule Pune University (SPPU), Ganeshkhind Road, Pune 411007, India

4 Combustion Science and Technology, CSIR-Central Institute of Mining and Fuel Research (CIMFR), Dhanbad 826015, India

5 Mineral Technology, CSIR-Central Institute of Mining and Fuel Research (CIMFR), Dhanbad 826015, India

1 Introduction

Coal as a feedstock is used in processes such as combustion, gasification, and liquefaction. It is a complex substance mainly comprising carbon, hydrogen, nitrogen, sulfur, oxygen, and mineral matter that can be intrinsic and/or extraneous with differing form and composition (Ozbayoglu and Ozbayoglu 2006). Being a natural resource, coal exhibits a large variation in its composition.

In coal-based processes, coal's mineral matter experiences a wide variety of complex physical and chemical transformations. These result in the formation of ash that possesses a tendency of depositing on the surfaces of the heat-transfer and other exposed process equipment

(Seggiani and Pannocchia 2003). The phenomena responsible for this ash deposition are termed *slagging* and *fouling*. Slagging forms significantly viscous or fused deposits of ash in zones that are directly exposed to the hottest parts of the boiler (radiant heat exchange). The result of fouling is deposition of species in the vapor form and condensing on the surfaces due to the convective heat exchange. These effects typically take place in the cooler parts of the boiler at temperatures below the melting point of the bulk coal ash (Seggiani and Pannocchia 2003). The coal ash fusion at a low temperature, encourages the formation of a clinker that gathers around the heat transfer pipes and thereby corrodes the furnace components. It is well-known that the ash clinkering may lead to channel burning, pressure drop, and an unstable gasifier operation (Van Dyk et al. 2001). Thus, coal-based power stations have to necessarily take periodic shut-downs to remove the clinker from the ash recovery shuts and heat transfer pipes (Yin et al. 1998). The occurrence of ash slag flows in the *Integrated Gasification Combined Cycle* (IGCC) and other slagging reactors is directly attributed to the formation of the liquid slag and to the stabilities of the solid crystalline phases (Patterson and Hurst 2000; Skrifvars et al. 2004). The property that governs the behavior of ash in various coal-utilizing processes is termed *ash fusion temperature* (AFT). Its analysis consists of the determination of four temperatures signifying four phases in the melting of the ash, as described below.

(1) *Initial deformation temperature* (*IDT*): Temperature at which ash just begins to flow.
(2) *Softening temperature* (*ST*): Refers to the temperature at which the ash softens and becomes plastic.
(3) *Hemispherical temperature* (*HT*): Denotes the temperature yielding a hemispherically shaped droplet.
(4) *Fluid temperature* (*FT*): At this temperature, ash becomes a free-flowing fluid (Slegeir et al. 1988).

These AFTs possess following attributes and applications.

(1) They indicate the temperature range for a possible formation of the deposits on the heat adsorbing surfaces (Ozbayoglu and Ozbayoglu 2006).
(2) Provide important clues regarding the extent to which the ash agglomeration and clinkering are likely to occur within the combustor/gasifier (Alpern et al. 1984; Seggiani 1999; Van Dyk et al. 2001).
(3) They are of particular significance to the operation of all types of gasifiers (Bryers 1996; Wall et al. 1998). For instance, to allow continuous slug tapping, it is necessary that the operating temperature in the entrained flow gasifiers is above the flow temperature (Hurst et al. 1996). In the case of fluid-bed gasifiers, AFTs set the upper limit for the operating temperature at which the ash agglomeration is initiated (Song et al. 2010).
(4) The knowledge of AFTs is routinely utilized by the furnace and boiler operators and engineers in power generation stations for predicting the melting and sticking behavior of the coal ash (Seggiani and Pannocchia 2003).

Conventionally, characterization of the ash fusibility is conducted using ASTM D1857 procedure. It comprises monitoring cones or pyramids of ash—prepared in a muffle furnace at 815 °C—in an oven operated under a reducing atmosphere and whose temperature is continuously increased steadily past 1000 °C to as high as possible, preferably 1600 °C (2910 °F). It may be noted that for a given coal, the AFT analysis conducted by different laboratories may vary by ± 20–100 °C (Jak 2002; Winegartner and Rhodes 1975).

Owing to the importance of the knowledge of the AFTs in designing, operating, and optimizing the coal-based combustion, liquefaction and gasification processes, a wide variety of mathematical models have been developed for their prediction. Efforts to develop models with higher prediction accuracy and wider applicability still continue. The currently available AFT prediction models possess certain limitations as stated below.

(1) Most of the existing models have been developed using data pertaining to coals from a single or a few geographical regions. Since coals from different regions/countries may exhibit significantly different chemical and physical characteristics, the AFT models of coals from a single or a few geographies possess limited applicability.
(2) Some of the existing AFT prediction models do consider coals from multiple geographical regions (see, for example, Seggiani 1999; Seggiani and Pannocchia 2003). These models also possess reasonably good prediction accuracies. However, they are based upon a large number of predictors (input variables) and as a result, suffer from the following undesirable characteristics: (a) the models are complex, which adversely affects their generalization ability, and (b) costly and tedious experimentation needed for compiling the predictor data.

Accordingly, the main objective of this study is to develop AFT prediction models that are parsimonious (i.e., with lower complexity) and applicable to coals from a large number of geographical regions. Towards this objective, the present study reports the results of the development of *computational intelligence* (CI) based models for the prediction of *IDT, ST, HT,* and *FT*. The three CI paradigms

used in this modeling are *genetic programming* (GP), *multi-layer perceptron* (MLP) neural network, and *support vector regression* (SVR). The results of the CI-based modeling of AFT prediction models indicate that the GP and MLP based models predicting *IDT, ST, HT,* and *FT* have outperformed the existing linear models with relatively wider applicability in terms of possessing better generalization capability. Also, the said GP and MLP based models require a lower number of predictors than the stated linear models thus reducing the effort and cost involved in compiling the predictor data.

The remainder of this paper is structured as follows. An overview of the existing models for AFT prediction is provided in Sect. 2. The necessity to develop data-driven nonlinear models is explained in Sect. 3. Next, a brief overview of the GP, MLP and SVR formalisms is provided in Sect. 4 titled "CI-based modelling." The Sect. 5 titled "Results and discussion" presents the development of the CI-based models for the prediction of four AFTs. This section also provides a comparison of the prediction and generalization performance of the CI-based models. Finally, "Concluding remarks" summarize the principal findings of this study.

2 Models for predicting AFT

There exists a number of studies, demonstrating that the chemical and mineral composition of the coal ash governs its melting characteristics and fusion temperatures (Gray 1987; Kucukbayrak et al. 1993; Vorres 1979; Vassilev et al. 1995; Winegartner and Rhodes 1975). Since the composition of the ash influences the AFT magnitudes and, thereby, performance of a coal-based process, it is essential to establish quantitative relationships between the ash composition and the corresponding four AFTs. The conventional methods are unable to accurately predict the high-temperature behaviour of the coal ash, slag, and blends in the combustion and gasification technologies (Goni et al. 2003; Gray 1987; Huggins et al. 1981; Lloyd et al. 1995; Yin et al. 1998; Wall et al. 1998). Thus, several studies on the prediction of the AFTs have been conducted using a variety of methods such as thermodynamic, statistical, empirical, and more recently artificial intelligence based data-driven modeling techniques. These correlations/models are widely used for assessing the deposition characteristics of the coal ashes. Specifically, the models employ the mineral content of a number of oxides (expressed as weight percentages), for correlating with the data obtained from the standard ash fusion tests. It may be noted that the stated models simulate the ash formation under controlled conditions and, therefore, do not portray closely the conditions in the real combustors/gasifiers.

The AFT models, however, compensate for the said deficiency by providing predictions under a consistent set of test conditions (Lolja et al. 2002). The AFT predicting models have following applications: (a) based upon the chemical and mineral composition of the coal ashes, the models offer a method to calculate coal ashes' thermal properties (Seggiani 1999), and (b) provide a method to evaluate the outcome of an addition of minerals (such as CaO) for modifying the slag behaviour (Wall et al. 1998). A representative compilation of the currently available AFT models is presented in Table 1.

3 Experimental data and need for nonlinear AFT models

In the present study, weight percentages (wt%) of the seven principal oxides appearing in the coal ashes (i.e. SiO_2, Al_2O_3, Fe_2O_3, CaO, MgO, TiO_2, $Na_2O + K_2O$), which are the most frequently used parameters for correlating the ash fusibility with the mineral composition (Gao et al. 2011; Liu et al. 2007; Yin et al. 1998), have been used as the predictors (model inputs) of AFTs. Large datasets containing the information of the composition of the stated oxides and the corresponding magnitudes of the four AFTs, pertaining to the ash samples of coals from several countries listed below were compiled from a number of research articles.

(1) *IDT* Albania, Australia, China, Colombia, India, Indonesia, Russia, South Sumatra, South Africa, USA, Venezuela.
(2) *ST* Australia, China, Colombia, Indonesia, Russia, South Africa, South Sumatra, USA, Venezuela.
(3) *HT* Albania, Australia, Colombia, India, Indonesia, Russia, South Africa, South Sumatra, USA, Venezuela.
(4) *FT* Albania, Australia, China, Colombia, India, Indonesia, Russia, South Africa, South Sumatra, USA, Venezuela.

A portion of the data on Indian coal ashes was sourced from *Central Institute of Mining and Fuel Research* (CIMFR), Dhanbad, India. The data sets consisting of the seven model predictors, the corresponding AFTs, and their sources are listed in Tables S1–S4 in *Supplementary Material*. Specifically, these four tables contain the oxide composition and the corresponding AFT data pertaining to the *IDT* (184 samples), *ST* (92 samples), *HT* (82 samples), and *FT* (94 samples), respectively. Most of these data were collected by conducting experiments in the reducing conditions.

The regression studies, such as those by Lolja et al. (2002), Ozbayoglu and Ozbayoglu (2006), and Seggiani (1999), have reported that there exists a linear dependence

Table 1 A representative compilation of AFT predicting correlations/models

No.	Authors	Type of model	Model inputs (composition of ash constituents and other parameters)	Predicted AFT	Coal region	Statistical analysis[a] of model predictions
1	Winegartner and Rhodes (1975)	Stepwise regression	$Fe_2O_3\%$, FeO%, FeO $\times$ CaO%, bases/acids, silica value	ST	USA	$CC > 0.7$
2	Gray (1987)	Multiple regression	Different combinations of metal oxides	Reducing IDT and HT	New Zealand	$R^2 > 53\%$
3	Rhinehart and Attar (1987)	Thermo-dynamic modelling	Metal and other (P_2O_5, TiO_2, SO_3) oxides	IDT, ST, FT	US coals	$R^2 > 0.67$
4	Kucukbayrak et al. (1993)	Least square regression analysis	Combinations of metal oxides	HT	Turkish lignite	$CC > 0.26$
5	Yin et al. 1998	Back-propagation neural network	SiO_2, Al_2O_3, Fe_2O_3, CaO, MgO, TiO_2, $K_2O + Na_2O$	ST	Chinese	Average fractional error: 0.049
6	Kahraman et al. (1998)	Empirical models	Al_2O_3, Fe_2O_3, CaO	IDT, spherical, FT	Australian	$R^2 > 0.84$
7	Seggiani (1999)	Linear regression	49 parameters comprising concentrations of nine metal oxides, their squares, and combinations of these values, base, acid, Silica value, and dolomite ratio.	Reducing IDT, ST, HT, and FT for coal and biomass ashes	American, Australian, African, German, Italian, Polish, Spanish, etc.	CC range: 0.84–0.92
8	Lolja et al. (2002)	Linear regression	Metal oxides, acids, bases, crystal components and fluxing agents	IDT, ST, HT, FT	Albanian	$0.93 \leq CC \leq 0.95$
9	Jak (2002)	Thermo-dynamic modelling (FACT package)	SiO_2, Al_2O_3, Fe_2O_3, CaO	IDT, Spherical, HT, FT	Australian	Liquidus temperature and AFT strongly correlated
10	Seggiani and Pannocchia (2003)	Partial least squares regression	11–13 parameters comprising concentrations of nine metal oxides in ash, and their various combinations	Reducing IDT, ST, HT, FT	American, Italian, Spanish, German, Australian, Polish, African, French, Albanian	CC (training set) range: 0.75–0.82; CC (validation set) range: 0.76–0.84
11	Ozbayoglu and Ozbayoglu (2006)	Linear and nonlinear regression	Chemical composition of ash and coal parameters	ST, FT	Turkey	Regression coefficient > 0.93
12	Liu et al. (2007)	Back propagation neural network-ant colony optimization	SiO_2, Al_2O_3, Fe_2O_3, CaO, MgO, TiO_2, $K_2O + Na_2O$	ST	Chinese	Average training (Test) error: 1.55 (1.85)%
13	Zhao et al. (2010)	Least-squares support vector regression	SiO_2, Al_2O_3, Fe_2O_3, CaO, MgO, TiO_2, K_2O, Na_2O, SO_3	ST	Chinese	$CC = 0.927$, $MSE = 0.0128$
14	Gao et al. (2011)	Support vector regression by ACO algorithm	SiO_2, Al_2O_3, Fe_2O_3, CaO, MgO, TiO_2, $K_2O + Na_2O$	ST	China	$MSE = 1.52$, $CC = 0.999$ (training), and $= 0.9716$ (test)

Table 1 continued

No.	Authors	Type of model	Model inputs (composition of ash constituents and other parameters)	Predicted AFT	Coal region	Statistical analysis[a] of model predictions
15	Karimi et al. (2014)	Adaptive neuro fuzzy inference system	Different combinations of metal oxides	*IDT, ST, FT*	USA	*CC* = 0.97, 0.98 and 0.99, respectively
16	Miao et al. (2016)	Back-propagation neural net	SiO_2, Al_2O_3, Fe_2O_3, CaO, MgO	*ST*	Chinese	–

[a]*CC* correlation coefficient, *MSE* mean squared error, R^2 variance

between the mineral composition of the coal ashes and their AFTs. On the other hand, a number of studies (e.g. Gao et al. 2011; Karimi et al. 2014; Liu et al. 2007; Miao et al. 2016; Yin et al. 1998) listed in Table 1 have proposed nonlinear AFT prediction models. Towards analyzing the true nature of the dependencies (whether linear or nonlinear) between the individual oxide components in the coal ashes and the corresponding AFTs, cross-plots were generated as shown in Figs. 1, 2, 3 and 4. In these plots, magnitudes of the four AFTs (*IDT, ST, HT,* and *FT*) are plotted against the weight percentages of individual oxide constituents. The observations drawn from the cross-plots are given below.

(1) The four panels (1a–c, g) of Fig. 1, show that there exists an approximately linear relation between the *IDT* and the weight percentages of SiO_2, Al_2O_3, Fe_2O_3, and $K_2O + Na_2O$; however, due to the presence of a significant scatter in the corresponding data, a similar conclusion cannot be drawn from the cross-plots (Fig. 1d–f) involving other three ash components (CaO, MgO, and TiO_2).

(2) Notwithstanding the high scatter seen in all panels of Fig. 2, there exists a high probability of linear dependencies between the softening temperature and Al_2O_3 (Fig. 2b), Fe_2O_3 (Fig. 2c), MgO (Fig. 2e), and TiO_2 (Fig. 2f); whereas, possibly the individual relationships between *ST* and SiO_2 (Fig. 2a), CaO (Fig. 2d) and $K_2O + Na_2O$ (Fig. 2g) are nonlinear.

(3) Cross-plots in Fig. 3 suggest a high probability of nonlinear relationships between *HT* and weight percentages of three ash components, namely SiO_2 (Fig. 3a), Al_2O_3 (Fig. 3b) and CaO (Fig. 3d). Whereas, the individual relationships between *HT* and Fe_2O_3 (Fig. 3c), MgO (Fig. 3e), TiO_2 (Fig. 3f), and $K_2O + Na_2O$ (Fig. 3g), could be linear.

(4) In Fig. 4, linear dependencies are indicated with a high probability between the flow temperature (*FT*) and SiO_2 (Fig. 4a), Al_2O_3 (Fig. 4b), Fe_2O_3 (Fig. 4c), and $K_2O + Na_2O$ (Fig. 4g). However, in the remaining three panels of Fig. 4, the relationships between

FT and CaO (Fig. 4d), MgO (Fig. 4e) and TiO_2 (Fig. 4f), appear to be nonlinear.

As can be seen from the above observations, there exist several probable cases of the nonlinear relationships between various AFTs and the weight percentages of the individual oxides present in the coal ashes. Thus, it is necessary to explore nonlinear models for the prediction of the four AFTs. Such models are expected to better capture the relationships between the AFTs and the seven oxides in the coal ash and, thereby, make more accurate predictions than the linear models. Towards this objective, in the present study, three computational intelligence (CI) based data-driven modeling formalisms (GP, ANN, and SVR) have been employed for the prediction of the four AFTs from the knowledge of the mineral composition (oxides) of the coal ashes. The objective of developing multiple models for each AFT is to afford a comparison of their prediction and generalization performances and thereby selecting the best prediction model.

The general forms of the CI-based models developed in this study are given as:

$$IDT = f_1(x_1, x_2, x_3, x_4, x_5, x_6, x_7, \beta_{IDT}) \tag{1}$$

$$ST = f_2(x_1, x_2, x_3, x_4, x_5, x_6, x_7, \beta_{ST}) \tag{2}$$

$$HT = f_3(x_1, x_2, x_3, x_4, x_5, x_6, x_7, \beta_{HT}) \tag{3}$$

$$FT = f_4(x_1, x_2, x_3, x_4, x_5, x_6, x_7, \beta_{FT}) \tag{4}$$

where, *IDT, ST, HT,* and *FT* represent *Initial deformation temperature* (°C), *Softening temperature* (°C), *Hemispherical temperature* (°C), and *Fluid temperature* (°C), respectively; β_{IDT}, β_{ST}, β_{HT}, and β_{FT}, represent the parameter vectors associated with the functions f_1, f_2, f_3, and f_4, respectively. The seven predictors (inputs) of the four AFT models are defined as: (1) x_1: weight percentage of SiO_2, (2) x_2: weight percentage of Al_2O_3, (3) x_3: weight percentage of Fe_2O_3, (4) x_4: weight percentage of CaO, (5) x_5: weight percentage of MgO, (6) x_6: weight percentage of TiO_2, and (7) x_7: weight percentage of $K_2O + Na_2O$.
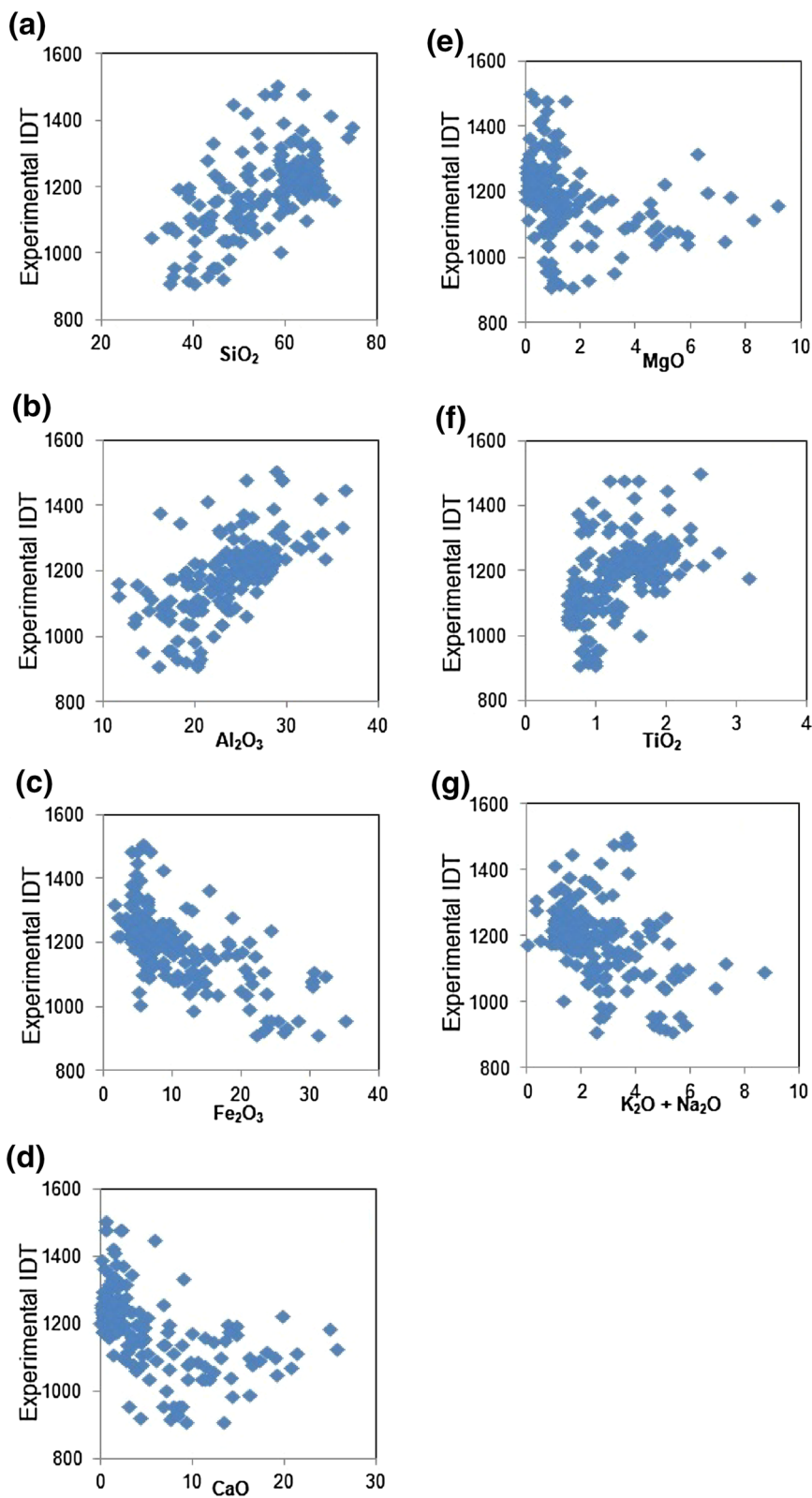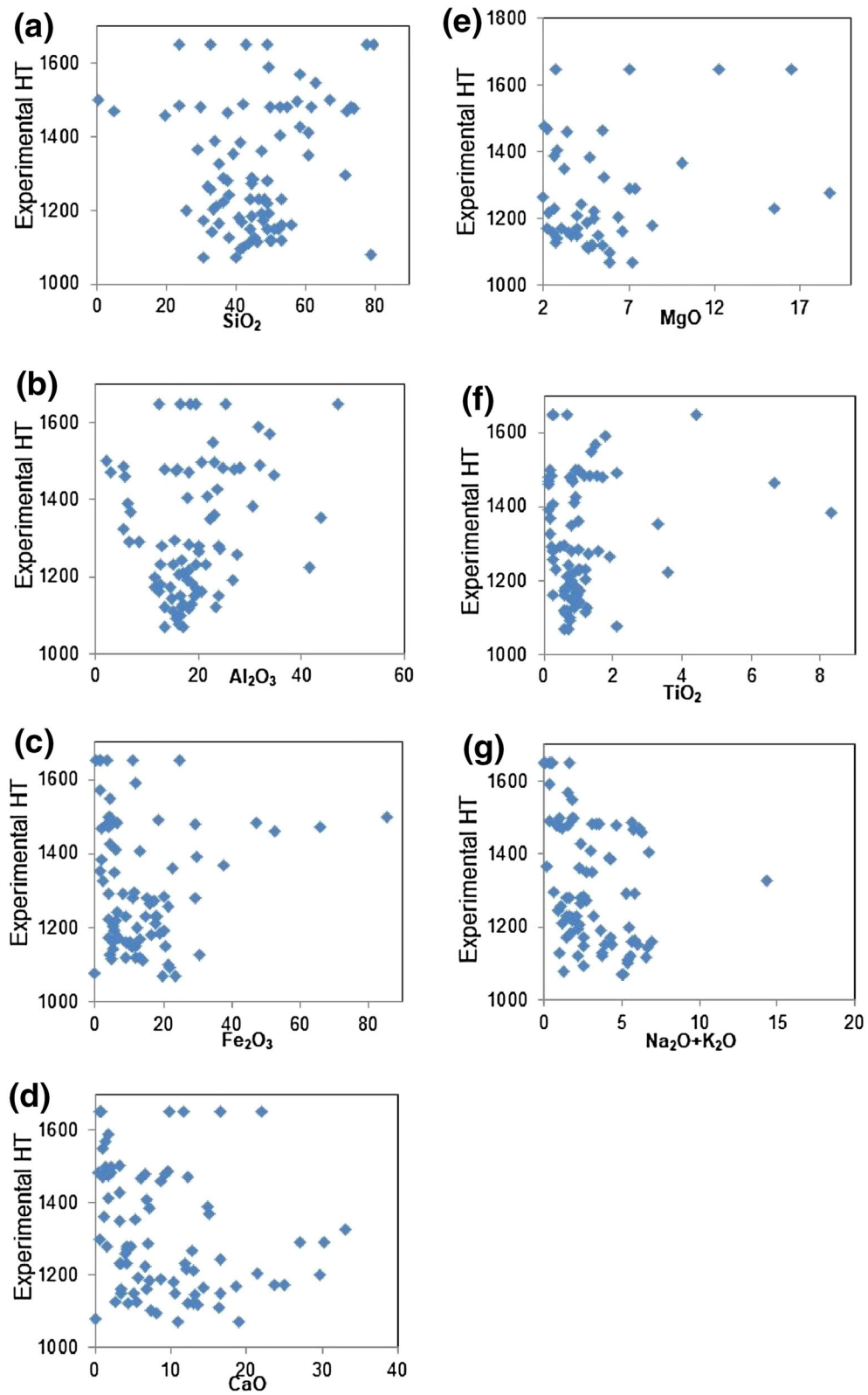
Fig. 1 Cross-plots of weight percentages (wt%) of individual metal oxides versus IDT magnitudes

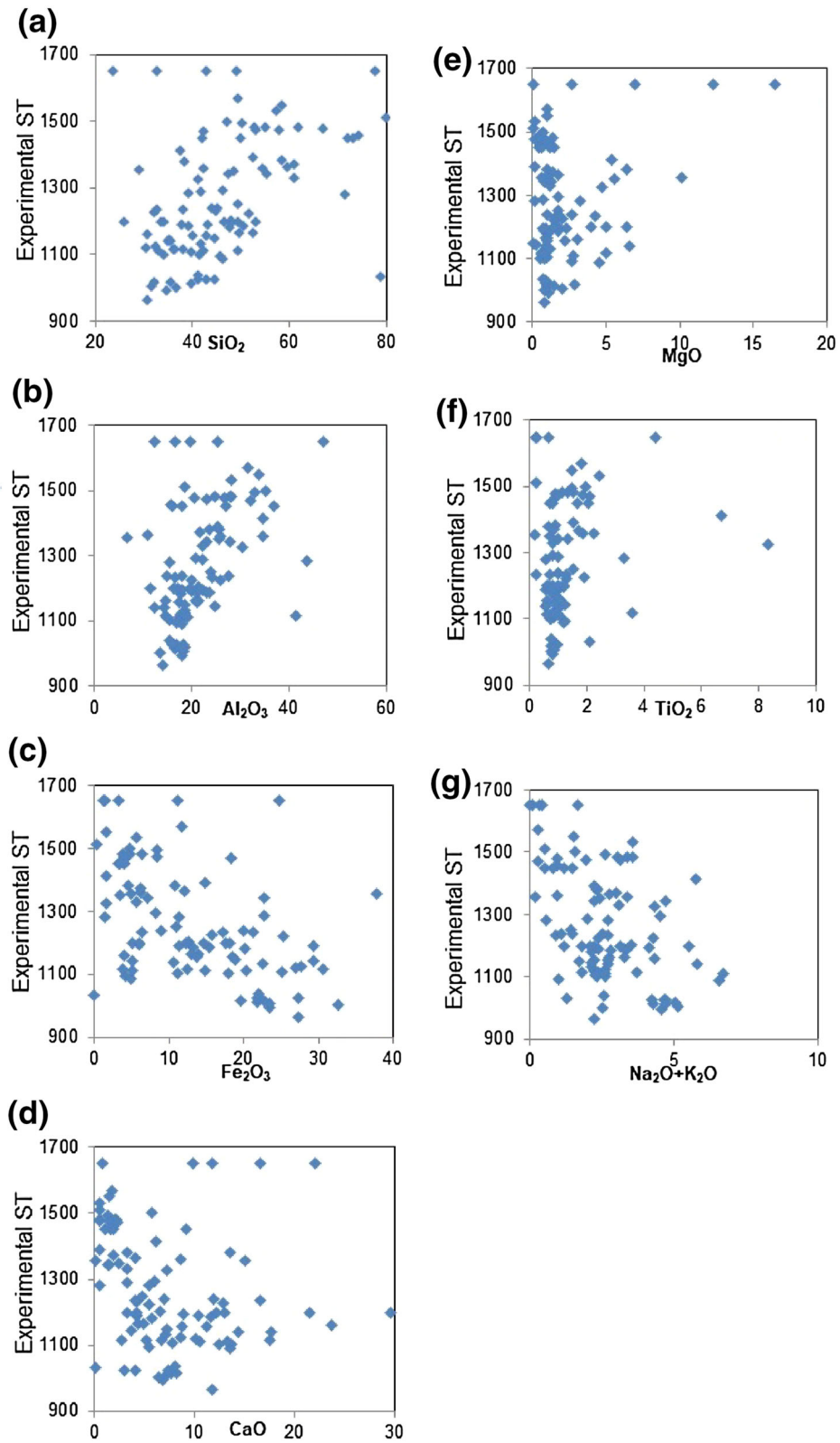Fig. 2 Cross-plots of weight percentages (wt%) of individual metal oxides versus ST magnitudes

Fig. 3 Cross-plots of weight percentages (wt%) of individual metal oxides versus HT magnitudes

Fig. 4 Cross-plots of weight percentages (wt%) of individual metal oxides versus FT magnitudes
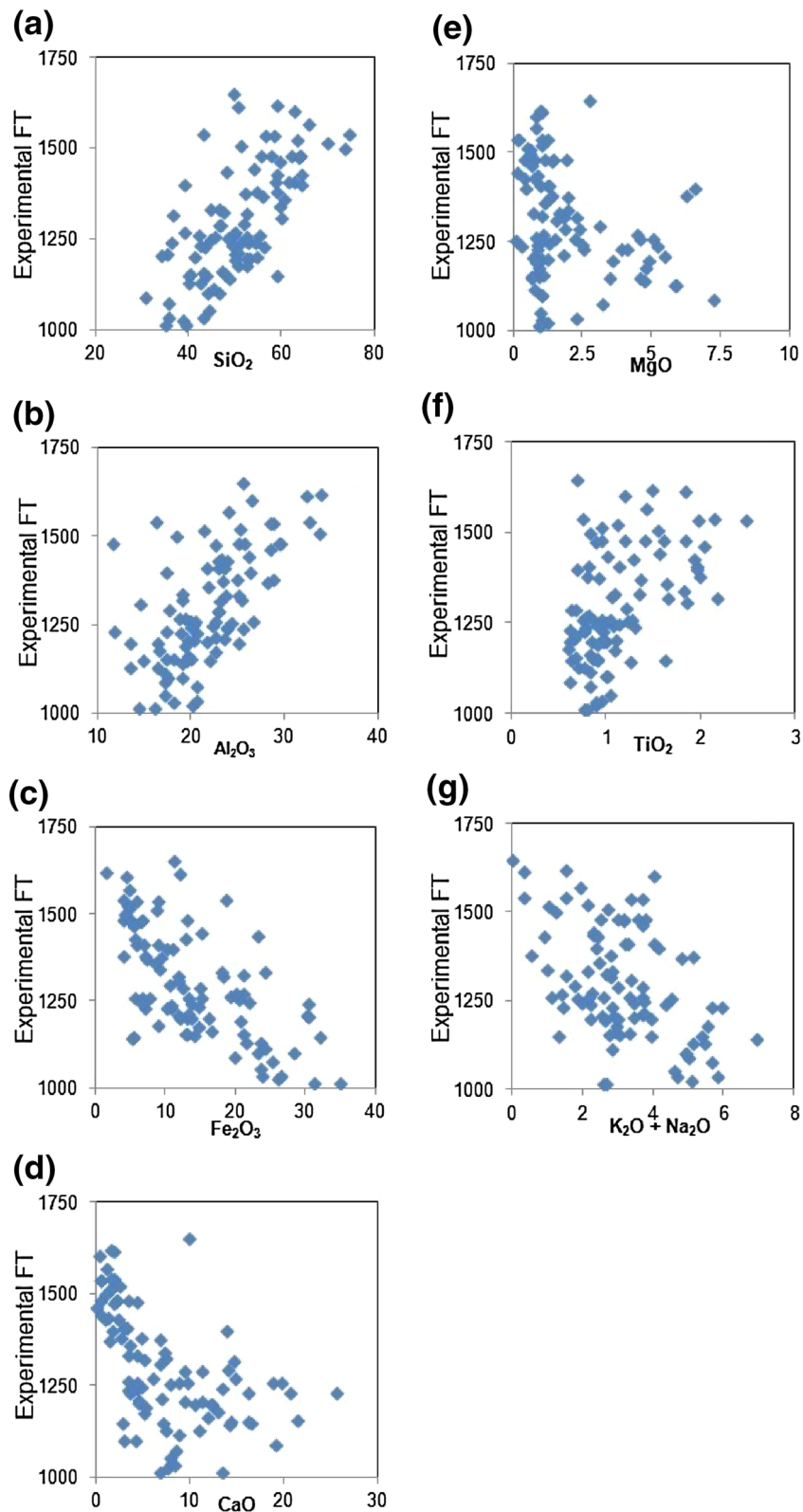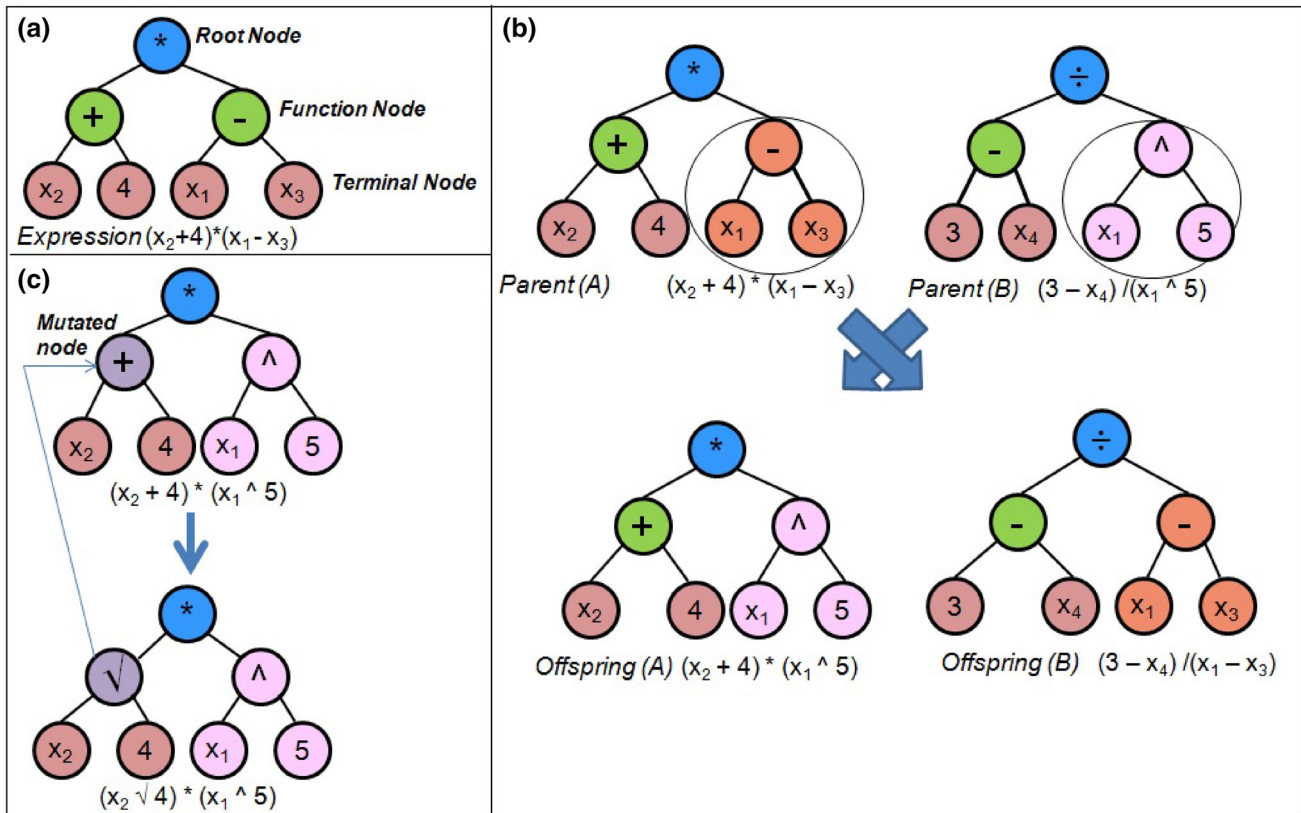
Fig. 5 Schematic diagram of: **a** a typical tree structure, **b** a crossover operation, and **c** a mutation operation
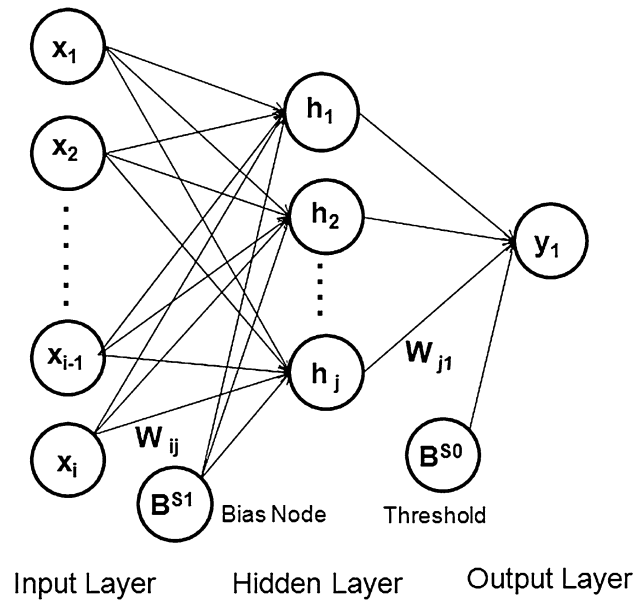

Fig. 6 Schematic of a single hidden layer multiple input–single output (MISO) MLP neural network

4 CI-based models for AFT prediction

4.1 Genetic programming (GP)

Genetic Programming formalism belongs to a class known as "evolutionary algorithms" that follow the principal tenet—commonly paraphrased as "survival of the fittest"—of Darwin's theory of evolution along with the genetic propagation of characteristics. Originally, GP was proposed (Koza 1992) to develop automatically the computer programs that would execute the pre-specified tasks. Genetic programming's other application known as "symbolic regression (SR)," is of interest to this study. The novel feature of the GP-based SR is as follows: provided an example dataset containing the function inputs (predictors/independent variables) and the corresponding outputs (dependent variables), it has the ability of searching as also optimising the specific structure (form) of an appropriate linear/nonlinear data-fitting function, and all its associated parameters. And, unlike MLP neural networks and the SVR formalism, the GP-based SR performs the stated search and optimisation without resorting to any assumptions about the structure and/or parameters of the linear/nonlinear data-fitting function. A data-driven modeling problem to be solved by the GP-based SR is explained below.
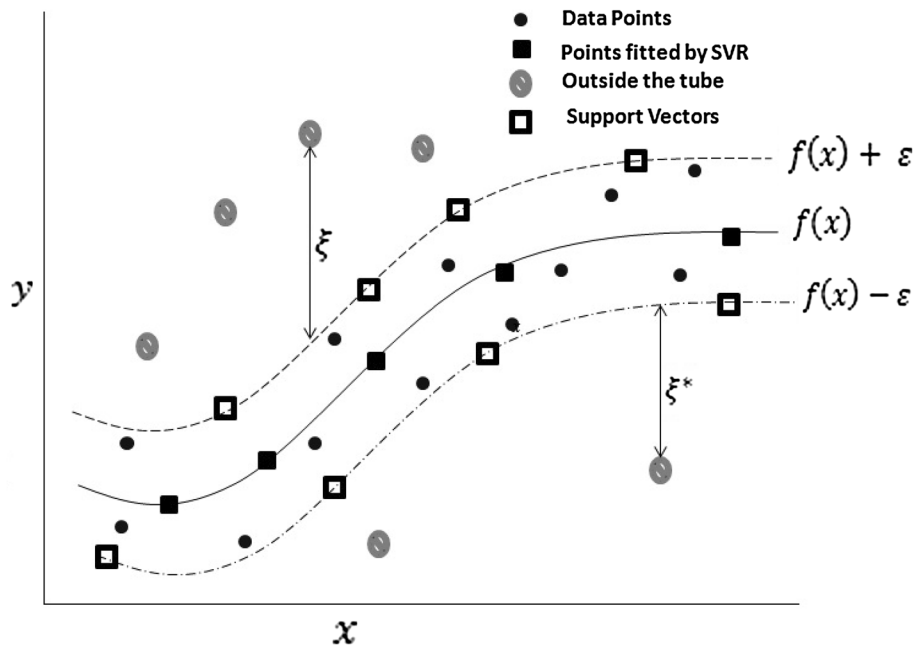
Fig. 7 Schematic of support vector regression using an ε-insensitive loss function

Consider a multiple input–single output (MISO) dataset, $G = (\boldsymbol{x}_i, y_i)_{i=1}^{N_p}$ that consists of N_p patterns ($i = 1, 2, \ldots, N_p$); $\boldsymbol{x}$ (= $[x_1, x_2, \ldots, x_M]^T$) refers to an M-dimensional vector of independent variables/predictors and y is the corresponding scalar output. The task of the GP-based SR is to fit an appropriate linear/nonlinear function (f) that best fits the dataset, G:

$$y = f(x_1, x_2 \ldots, x_M; \beta) \tag{5}$$

where, β represents a K-dimensional vector of function parameters, $\beta = [\beta_1, \beta_2, \ldots, \beta_K]^T$.

To search and optimize the form of f and the associated parameters (β), the GPSR begins by mimicking the Natural evolution in a somewhat abstract form. It randomly generates an initial population of N_{pp} number of probable (candidate) solutions (mathematical expressions/models) to the given data-fitting problem. Commonly, tree structures are used for defining these candidate solutions. Each of

these trees is formed randomly using *function* and *terminal* nodes. While the former represents mathematical operators, the latter defines the predictors (input variables, $\{x_m\}$) and parameters, β, of a data fitting expression. The set of available operators comprises those that perform *addition*, *subtraction*, *multiplication*, *division*, *logarithm*, *exponentiation*, and *trigonometric* operations. Each candidate solution is differentiated on the basis of its ability (fitness) to satisfactorily address the regression task, i.e., how well the solution fits the example set data. In the next step, a pool of relatively fitter candidates termed "parent pool" is formed and its constituents breed among themselves in frequency, which is directly proportional to their fitness. This is known as "crossover" operation that produces two offspring (new candidate solutions) per parent pair and these may replace the existing members of the candidate population. As in natural evolution, the new offspring may undergo mutation, whereby their genetic material is altered randomly, albeit to a small extent. The above-stated process of crossover and mutation is iterative, adaptive and open ended (McConaghy et al. 2010). Over time, the best candidate solutions (i.e., possessing high fitness) will survive. In Fig. 5, panel (a) shows a typical tree structure representing an expression "$(x_2 + 4)*(x_1 - x_3)$". The crossover and mutation are illustrated in panels (5b) and (5c), respectively.

Apart from searching and optimizing both the form and the associated parameters (vector β) of the best possible data-fitting function, GP-based SR method possesses other attractive features as outlined below.

Table 2 Variance captured by the individual principal components (PCs) in respect of *IDT, ST, HT* and *FT* data sets

Dataset	Percentage variance[a]						
	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7
IDT	55.2	14.3	10.8	9.4	6.2	4.1	0
ST	34.1	25.0	17.3	12.5	5.7	4.6	0.8
HT	38.1	20.9	16.8	13.2	6.5	4	0.6
FT	43.1	19.2	13.4	11.4	7.8	5.2	0

[a]PC_j denotes the jth principal component

(1) Generates an initial population of candidate expressions/models in a purely stochastic manner. That is, unlike similar techniques performing data-driven modelling, namely MLP neural network and SVR, the GPSR algorithm does not make any assumptions about the form and parameters of the data-fitting models/expressions.

(2) Invariably, GPSR-searched models are of lesser complexity (i.e. parsimonious) when compared with the corresponding MLP neural network and SVR models. Consequently, these models are easier to grasp and use in practice.

(3) The automatic search and optimization of the form and associated parameters of the linear/nonlinear data-fitting function, performed by the GPSR obviates the trial and error approach associated with the traditional linear/nonlinear regression analysis.

More details of the GP-based SR method and its implementation procedure can be found in, for example, Ghugare et al. (2014), Goel et al. (2015), Poli et al. (2008), and Vyas et al. (2015).

4.2 Multilayer perceptron (MLP) neural network

The *multilayer perceptron neural network* is devised on the basis of the functioning of the naturally occurring network of neurons in the human brain. It is a widely employed data-driven nonlinear function approximation technique. The phenomenal information processing capability of MLP arises from its multilayer architecture housing artificial neurons (processing elements/nodes) that are linked using weighted synaptic connections. An MLP possesses a feed-forward structure meaning the information flow occurs only in the forward direction. Most often, it contains three layers of nodes namely *input, hidden,* and *output* layers (see Fig. 6); multiple hidden layers can also be housed in an MLP architecture. Each node in its hidden layer processes incoming information using a nonlinear transfer function, such as the *logistic sigmoid* or *hyperbolic tangent* (*tanh*), to compute its output. The much desired nonlinear function approximation or equivalently input–output mapping capability of an MLP is due to the said nonlinear processing performed by its hidden layer nodes. Given an example dataset, $G = (x_i, y_i)_{i=1}^{N_p}$, consisting of the model's inputs and the corresponding outputs, an MLP can learn the complex nonlinear input–output relationships therein. This learning (training) is conducted using a suitable learning algorithm [for example, the *generalized delta rule* based *error-back-propagation* (EBP) algorithm (Rumelhart et al. 1986)] that optimizes the interlayer connection weights such that the error between the MLP-computed outputs and their desired (target) magnitudes (known as *prediction*

error) is minimized. A detailed description of the MLP is beyond the scope of this study since it is available at numerous books (e.g. Bishop 1995; Tambe et al. 1996; Zurada 1992), and reviews and research articles (see, for example, Rumelhart et al. 1986; Zhang et al. 1998).

4.3 Support vector regression (SVR)

Support vector machine (SVM) is a statistical learning based formalism for conducting supervised nonlinear classification (Vapnik 1995). To perform the said classification, SVM first maps the coordinates of objects to be classified into a high-dimensional *feature* space by employing nonlinear functions called *kernels* or *features.* Next, two classes are separated in this high dimensional space using a linear classifier as done customarily. Support vector regression employs same principles, however, for performing a nonlinear regression, which is of interest to this study.

The goal of SVR is to approximate a function $f(\mathbf{x})$ from a given example data set, $G = (x_i, y_i)_{i=1}^{N_p}$. The underlying objective is to map the input data $(x_i)_{i=1}^{N_p}$, nonlinearly into a high dimensional feature space (Φ) and conduct a linear regression in this space as given by:

$$y = f(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) + b \qquad (6)$$

where $\mathbf{w}$ is the function coefficient vector, $\Phi(\mathbf{x})$ refers to a set of nonlinear transformations, and b is a real constant (threshold value). For estimating the quality of regression, a loss function (LF) is utilized, such as the commonly "ε-insensitive" function as given below.

$$|y - f(\mathbf{x})|_\varepsilon = \begin{cases} 0 & if |y - f(\mathbf{x})| \leq \varepsilon \\ |y - f(\mathbf{x})| - \varepsilon & \text{otherwise} \end{cases} \qquad (7)$$

The SVR algorithm tries locating a tube of radius ε that surrounds the regression function (see Fig. 7). The region enclosed by the tube is called "ε-insensitive" zone wherein ε signifies the tolerance to the deviation. According to this formulation, errors are considered to be those deviations which are larger than ε.

The coefficient vector ($\mathbf{w}$) and constant b can be estimated from the training data by minimizing the following empirical risk function:

$$R_{emp}(\mathbf{w}) = \frac{1}{2}||\mathbf{w}||^2 + \frac{C}{N}\sum_{i=1}^{N_p}|y_i - f(\mathbf{x}_i)|_\varepsilon \qquad (8)$$

where, C refers to the regularization constant determining the trade-off between the training data set error and the model complexity. The SVR tries to find coefficients such that the maximum number of data points lie within the epsilon-wide insensitivity tube (Vapnik 1995). A detailed description of the SVR and its implementation can be

found in e.g., Vapnik (1995), and Ivanciuc (2007). The final form of the SVR-based regression function is given as:

$$f(\boldsymbol{x}, \boldsymbol{w}) = f(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\lambda}^*) = \sum_{i=1}^{N_p} f(\lambda_i^* - \lambda_i) K(\boldsymbol{x}_i, \boldsymbol{x}) + b \qquad (9)$$

where, N_p refers to the number of training data points; $K(\boldsymbol{x}_i, \boldsymbol{x})$ denotes the kernel function representing the dot product in the feature space (Φ); λ_i, λ_i^* (> 0) are the coefficients (Lagrange multipliers) fulfilling the condition $\lambda_i \lambda_i^* = 0$ ($i = 1, 2,\ldots, N_p$). The vector $\boldsymbol{w}$ is described in terms of the Lagrange multipliers λ_i and λ_i^*. In Eq. (9), only some of the coefficients, ($\lambda_i^* - \lambda_i$), possess non-zero magnitudes, and the corresponding input vectors, $\boldsymbol{x}_i$, are termed "support vectors (SVs)" signifying the most informative observations that compress the information content of the training set. Since those observations lying close to the prediction by the SVR model and located within the limit defined by the ε-insensitive tube are ignored, the final model is governed only by the SVs.

The salient features of the SVR are given below (Sharma and Tambe 2014).

(1) It minimizes a quadratic function with a single minimum, which avoids the problems associated with finding a solution in the presence of multiple local minima.

(2) Guarantees (a) robustness of the solution, (b) good generalization ability, and sparseness of the regression function, and (c) an automatic control of the regression function's complexity.

(3) An explicit knowledge of the *support vectors,* which play a major role in defining the regression function assists in the interpretation of the SVR-derived model in terms of the training data.

5 Results and discussion

5.1 Principal component analysis (PCA)

It is a requisite—while developing the data-driven models—to avoid correlated inputs (predictors) since these cause redundancy and an unnecessary increase in the computational load. Thus, the seven inputs ($x_1 - x_7$) of *IDT, ST, HT,* and *FT* prediction models were subjected to the *principal component analysis* (PCA) (Geladi and Kowalski 1986). This analysis performs a transformation to obtain linearly uncorrelated variables. Subsequently, only the first few principle components (PCs) that capture the maximum amount of variance in the data are chosen as model inputs (predictors), thus enabling a reduction in the

dimensionality of the model's input space. In the present study, seven PCs were extracted from the wt% values of the seven oxides in the coal ashes listed in *Supplementary Material.*

Prior to performing the PCA, the seven inputs in the example data for *IDT, ST, HT* and *FT* (listed in the *Supplementary material* in Tables S1–S4, respectively) were normalized using "Z-score" technique as given by,

$$\hat{x}_k^i = \frac{x_k^i - \bar{x}_k}{\sigma_k}; \quad i = 1, 2, \ldots, N_p; \ k = 1, 2, \ldots, K \qquad (10)$$

where, x_k^i represents the ith value of the kth un-normalized input variable, x_k; K denotes the number of inputs subjected to PCA ($= 7$); $\bar{x}_k$ refers to the mean of $\{x_k^i\}, i = 1, 2, \ldots, N_p$, and σ_k represents the standard deviation of $\{x_k^i\}, i = 1, 2, \ldots, N_p$. Similar to the predictor variables, the magnitudes of the four model outputs, namely *IDT* (y_1), *ST* (y_2), *HT* (y_3) and *FT* (y_4) were normalized as given below.

$$\hat{y}_s^i = \frac{y_s^i - \bar{y}_s}{\sigma_S}, \quad i = 1, 2, \ldots, N_p; \ s = 1, 2, \ldots, S \qquad (11)$$

where, $\hat{y}_s^i$, and y_s^i respectively represent the ith normalized and un-normalized values of the sth output variable (i.e., AFT), y_s; S refers to the number of outputs ($=$ the number of AFTs $= 4$); $\bar{y}_s$ refers to the mean of $\{y_s^i\}, i = 1, 2, \ldots, Np$ values in the example set, and σ_s represents the standard deviation of $\{y_s^i\}, i = 1, 2, \ldots, Np$. The mean and standard deviation values pertaining to weight percentages of seven oxides (SiO_2, Al_2O_3, Fe_2O_3, CaO, MgO, TiO_2, and $K_2O + Na_2O$) and all four ash fusion temperature phases (*IDT, ST, HT,* and *FT*) used in the normalization procedure are given in Table S5 in *Supplementary material.* The expressions of the PCs derived from the input data pertaining to the *IDT, ST, HT* and *FT* models are provided in the next section.

Table 2 lists the magnitudes of the variance captured by the seven PCs in the experimental data. From Table 2, it is seen that the first five PCs have cumulatively captured large percentages of variance, viz. $\approx$ 96%, 96%, and 95% in the oxides data for *IDT, HT* and *FT,* respectively. In the case of oxides data for *ST* the corresponding variance magnitude is $\approx$ 92% for the first four PCs. This result suggests that the first four PCs can be considered as the predictors in the model predicting *ST,* whereas for *IDT, HT* and *FT* prediction models first five PCs can be regarded as inputs.

The prediction accuracy and generalization capability of a CI-based model were examined using three statistical metrics, namely *coefficient of correlation* (CC), *root mean square error* (RMSE), and *mean absolute percent error* (MAPE); these were calculated using the experimental and

Table 3 Statistical analysis of the prediction accuracy and generalization performance of the GP-, MLP-, and SVR based models predicting *IDT, ST, HT,* and *FT*

Type of AFT	Dataset[a]	Computational intelligence based models								
		GP			MLP			SVR		
		CC	*RMSE*	*MAPE* (%)	*CC*	*RMSE*	*MAPE* (%)	*CC*	*RMSE*	*MAPE* (%)
IDT	Training	0.843	59.60	3.96	0.876	53.22	3.39	0.850	58.29	4.18
	Test	0.846	70.48	4.60	0.824	77.93	4.73	0.776	84.13	5.23
	Validation	0.803	33.79	2.18	0.764	38.78	2.66	0.774	38.83	2.65
ST	Training	0.833	102.16	6.35	0.841	127.20	7.78	0.832	103.6	5.32
	Test	0.921	68.81	4.83	0.841	121.60	6.47	0.836	93.31	4.97
	Validation	0.823	82.20	5.86	0.878	90.96	5.09	0.949	48.35	3.04
HT	Training	0.803	96.63	5.68	0.926	70.10	4.12	0.813	95.42	4.89
	Test	0.929	75.07	4.09	0.894	107.30	6.77	0.800	110.68	5.74
	Validation	0.953	81.01	5.26	0.851	109.59	5.80	0.804	118.02	5.76
FT	Training	0.915	62.53	3.59	0.962	44.70	2.81	0.941	53.21	2.91
	Test	0.941	56.94	3.75	0.901	71.38	4.38	0.857	82.61	4.74
	Validation	0.886	79.33	4.95	0.900	71.18	4.53	0.891	75.18	4.76

[a] The number of patterns in training, test and validation sets are: (a) *IDT* model: 129, 37, 18, (b) *ST* model: 68,19,10, (c) HT model: 57,17,8, (d) *FT* model: 66,19,9

Table 4 Details of the MLP-based optimal models for the prediction of *IDT, ST, HT,* and *FT*

MLP model	Number of nodes in input layer (*L*)	Number of nodes in the hidden layer (*M*)	Error back propagation algorithm parameter		Transfer function for hidden nodes	Transfer function at output node
			Learning rate (η)	Momentum (μ)		
IDT	5	6	0.3	0.8	Logistic sigmoid	Identity
ST	4	6	0.31	0.23	Logistic sigmoid	Identity
HT	5	6	0.271	0.1809	Logistic sigmoid	Identity
FT	5	6	0.27	0.16	Logistic sigmoid	Identity

Table 5 Details of the SVR-based models predicting *IDT, ST, HT,* and *FT*

SVR Model	Kernel gamma (γ)	Tube radius (ε)	Regularization parameter (*C*)	Kernel degree (δ)	Number of support vectors	Kernel function
IDT	2	0.6	2	1	48	ANOVA
ST	0.011	0.2	10	–	45	RADIAL
HT	0.098	0.24	0.5	2.7	41	ANOVA
FT	1	0.09	1.1	1	60	ANOVA

the corresponding model predicted AFT magnitudes. The *MAPE* was evaluated according to the following expression:

$$MAPE_j(\%) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left| \frac{y_i - \hat{y}_{i,j}}{y_i} \right| \times 100; \quad j = 1, 2, \ldots, N_{pp}$$

(12)

where, *j* denotes the index of the candidate solution in a population; *MAPE$_j$* refers to the *MAPE* pertaining to the *j*th
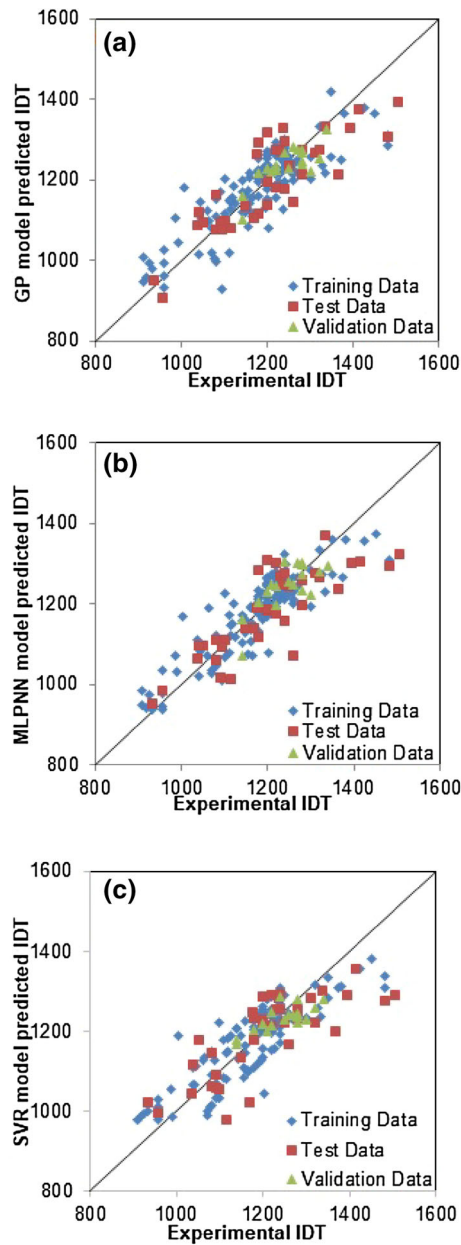
Fig. 8 Plots of the experimental *IDT* values versus those predicted by the GP (panel **a**), MLP (panel **b**), and SVR based (panel **c**) models

candidate solution in the population containing N_{pp} candidate solutions; y_i is the desired (target) output value corresponding to the ith input data pattern in the training/test data set; and $\hat{y}_{i,j}$ is the magnitude of the model predicted AFT when the ith input pattern is used to compute the output of the jth candidate solution. The *RMSE* was evaluated as follows:

$$RMSE_j = \sqrt{\frac{\sum_{i=1}^{N_p}\left(y_i - \hat{y}_{i,j}\right)^2}{N_p}}; \quad j = 1, 2, \ldots, N_{pp} \qquad (13)$$

where, $RMSE_j$ refers to the *RMSE* pertaining to the jth candidate solution.

The PCA-transformed variables were used as the inputs in developing the GP-, MLP-, and SVR-based *IDT, ST,* HT, and *FT* prediction models. For constructing and examining the generalization ability of these models, the experimental data set for each AFT was randomly partitioned in 70:20:10 ratio into *training, test,* and *validation* sets. While the first set was used in training the CI-based models, the test and the validation sets were respectively used in testing and validating the generalization capability of models.

5.2 GP-based modelling

The four GP-based AFT models were developed using *Eureqa Formulize* software package (Schmidt and Lipson [2009]). A notable feature of this package is that it is tailored to search and optimize models with a low complexity and, thereby, possessing the much-desired generalization ability. There are multiple procedural attributes that affect the final solution provided by the GP. These include sizes of the training, test and validation sets, choice of the operators, and input normalization schemes. To secure parsimonious models endowed with a good AFT prediction accuracy and generalization capability, several GP runs were conducted by imparting variations in each of the stated attributes. The best solution (possessing maximum fitness) secured in each run was recorded. From multiple such solutions, the ones fulfilling the following criteria were screened to choose an overall optimal model (Sharma and Tambe [2014]): (a) high and comparable magnitudes of *CC*s, and small and comparable magnitudes of *RMSE,* and *MAPE*, pertaining to the model predictions in respect of the training, test, and validation set data, and (b) model should possess a low complexity (i.e., containing a small number of terms and parameters in its structure).

The GP-based overall best models for the four AFTs are given below wherein $\widehat{IDT}, \widehat{ST}, \widehat{HT},$ and $\widehat{FT}$, respectively refer to the normalized values of *IDT, ST, HT,* and *FT* (see Eqs. [14], [20], [25] and [31]).

(a)   Model for predicting Initial deformation temperature (IDT)

$$\begin{aligned}\widehat{IDT} = {} & 0.096PC_1PC_2 + 0.309PC_4^2 \\ & + \left(\frac{PC_1 + PC_5}{PC_3 - 1.426PC_5 - 3.252}\right) - 0.096PC_1PC_4^2 \\ & - 0.121PC_3 - 0.199\end{aligned}$$

$$(14)$$

Fig. 9 Plots of experimental *ST* values versus those predicted by the GP (panel **a**), MLP (panel **b**), and SVR based (panel **c**) models
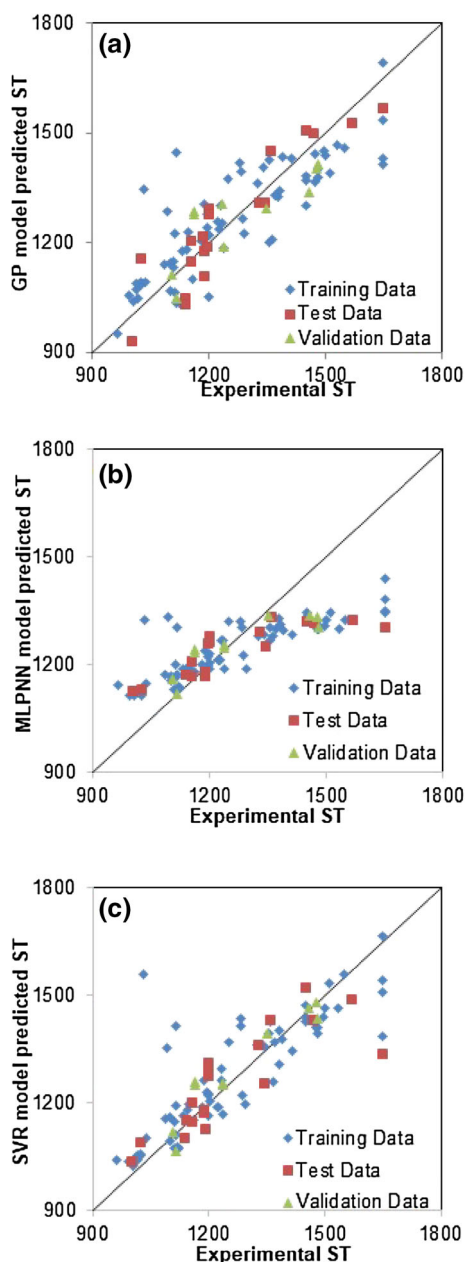

Fig. 10 Plots of experimental *HT* values versus those predicted by the GP (panel **a**), MLP (panel **b**), and SVR based (panel **c**) models

where,

$$PC_1 = -0.423\hat{x}_1 - 0.375\hat{x}_2 + 0.343\hat{x}_3 + 0.427\hat{x}_4 \\ + 0.366\hat{x}_5 - 0.384\hat{x}_6 + 0.315\hat{x}_7 \tag{15}$$

$$PC_2 = 0.406\hat{x}_1 - 0.119\hat{x}_2 - 0.709\hat{x}_3 + 0.163\hat{x}_4 + 0.474\hat{x}_5 \\ - 0.192\hat{x}_6 + 0.171\hat{x}_7 \tag{16}$$
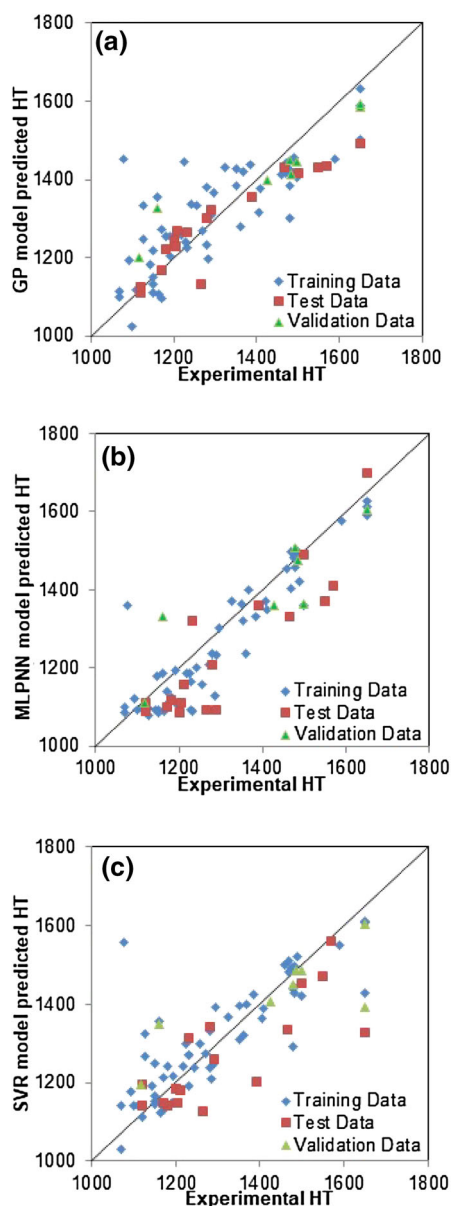
$$PC_3 = 0.373\hat{x}_1 - 0.479\hat{x}_2 + 0.120\hat{x}_3 - 0.389\hat{x}_4 - 0.405\hat{x}_5 \\ - 0.250\hat{x}_6 + 0.489\hat{x}_7 \tag{17}$$

$$PC_4 = 0.236\hat{x}_1 - 0.459\hat{x}_2 + 0.085\hat{x}_3 + 0.122\hat{x}_4 - 0.078\hat{x}_5 \\ - 0.314\hat{x}_6 - 0.779\hat{x}_7 \tag{18}$$

$$PC_5 = 0.064\hat{x}_1 - 0.539\hat{x}_2 - 0.003\hat{x}_3 + 0.252\hat{x}_4 + 0.023\hat{x}_5 \\ + 0.799\hat{x}_6 + 0.052\hat{x}_7 \tag{19}$$
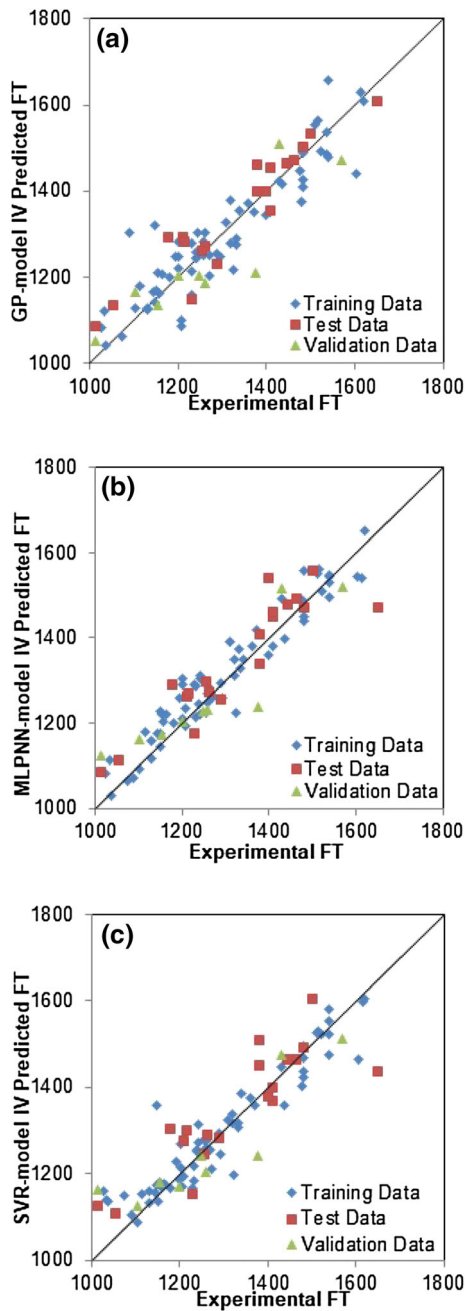
Fig. 11 Plots of experimental *FT* values versus those predicted by the GP (panel **a**), MLP (panel **b**), and SVR based (panel **c**) models

The magnitudes of $\hat{x}_k$; $k = 1,2,\ldots,7$ were evaluated using Eq. (10); the mean and standard deviation values pertaining to the *IDT* data are given in Table S5.

(b) Model for predicting softening temperature (*ST*)

$$\widehat{ST} = 0.261PC_2 + 0.230PC_1PC_2 + 0.142PC_1PC_3 \\ - 0.329PC_4 - 0.527PC_1 - 0.261PC_3PC_4 \quad (20)$$

where,

$$PC_1 = -0.494\hat{x}_1 - 0.394\hat{x}_2 + 0.356\hat{x}_3 + 0.535\hat{x}_4 \\ + 0.325\hat{x}_5 - 0.237\hat{x}_6 + 0.165\hat{x}_7 \quad (21)$$

$$PC_2 = -0.307\hat{x}_1 + 0.457\hat{x}_2 - 0.316\hat{x}_3 + 0.237\hat{x}_4 \\ + 0.399\hat{x}_5 + 0.596\hat{x}_6 + 0.161\hat{x}_7 \quad (22)$$

$$PC_3 = 0.291\hat{x}_1 - 0.134\hat{x}_2 - 0.426\hat{x}_3 + 0.25\hat{x}_4 + 0.478\hat{x}_5 \\ - 0.228\hat{x}_6 - 0.611\hat{x}_7$$

$$(23)$$

$$PC_4 = 0.328\hat{x}_1 - 0.362\hat{x}_2 - 0.523\hat{x}_3 + 0.153\hat{x}_4 - 0.012\hat{x}_5 \\ - 0.066\hat{x}_6 + 0.678\hat{x}_7$$

$$(24)$$

In Eqs. (21) to (24), $\hat{x}_k$; $k = 1,2,\ldots,7$ were computed using Eq. (10); the mean and standard deviation values pertaining to the *ST* data are given in Table S5.

(c) Model for predicting hemispherical temperature (*HT*)

$$\widehat{HT} = 0.529 \\ + \left(\frac{0.736PC_4^2 - PC_1 - 1.678 - 0.592PC_5 - 0.578PC_4PC_5}{\exp(0.397PC_2PC_3 + 0.358PC_1PC_2 + 0.358PC_1^2)} \right) \quad (25)$$

where,

$$PC_1 = -0.44\hat{x}_1 - 0.486\hat{x}_2 + 0.375\hat{x}_3 + 0.421\hat{x}_4 + 0.25\hat{x}_5 \\ - 0.303\hat{x}_6 + 0.314\hat{x}_7$$

$$(26)$$

$$PC_2 = 0.202\hat{x}_1 - 0.312\hat{x}_2 + 0.402\hat{x}_3 - 0.426\hat{x}_4 - 0.396\hat{x}_5 \\ - 0.522\hat{x}_6 - 0.299\hat{x}_7$$

$$(27)$$

$$PC_3 = -0.549\hat{x}_1 + 0.277\hat{x}_2 + 0.54\hat{x}_3 - 0.25\hat{x}_4 - 0.26\hat{x}_5 \\ + 0.447\hat{x}_6 - 0.014\hat{x}_7$$

$$(28)$$

$$PC_4 = -0.129\hat{x}_1 + 0.047\hat{x}_2 + 0.13\hat{x}_3 - 0.061\hat{x}_4 + 0.662\hat{x}_5 \\ - 0.017\hat{x}_6 - 0.723\hat{x}_7$$

$$(29)$$

$$PC_5 = 0.206\hat{x}_1 - 0.043\hat{x}_2 + 0.159\hat{x}_3 - 0.629\hat{x}_4 + 0.517\hat{x}_5 \\ + 0.054\hat{x}_6 + 0.515\hat{x}_7$$

$$(30)$$

The magnitudes of $\hat{x}_k$; $k = 1,2,\ldots,7$ in Eqs. (26)–(30) were calculated using Eq. (10); the mean and standard deviation values pertaining to the HT data are given in Table S5.

(d) Model for predicting flow temperature (*FT*)

Table 6 Results of the Steiger's Z-test (testing the null hypothesis H_0, $CC_{AB} = CC_{AC}$)

Performance variable	Model pair (B–C)	CC_{AB}	CC_{AC}	CC_{BC}	df[a]	z[a]	p value[a]	H_0
IDT	GP-MLP	0.848	0.860	0.942	181	− 0.993	0.321	Accept
	MLP-SVR	0.860	0.832	0.917	181	1.869	0.06	Accept
	SVR-GP	0.832	0.848	0.925	181	− 1.07	0.286	Accept
ST	GP-MLP	0.848	0.843	0.903	94	0.221	0.825	Accept
	MLP-SVR	0.843	0.841	0.964	94	0.148	0.883	Accept
	SVR-GP	0.841	0.848	0.914	94	− 0.332	0.74	Accept
HT	GP-MLP	0.844	0.909	0.862	79	− 2.681	0.007	Reject
	MLP-SVR	0.909	0.802	0.820	79	3.723	0.0002	Reject
	SVR-GP	0.802	0.844	0.892	79	− 1.518	0.113	Accept
FT	GP-MLP	0.916	0.944	0.933	91	− 2.354	0.019	Reject
	MLP-SVR	0.944	0.919	0.945	91	2.258	0.024	Reject
	SVR-GP	0.916	0.919	0.945	91	0.265	0.791	Accept

[a] df degrees of freedom, z test statistics, p significance level

$$\widehat{FT} = 0.133PC_1PC_2 + 0.133PC_4^2$$
$$+ 0.0577PC_1^2\left(0.058PC_1^2\right)^{\left(PC_5 + 0.133PC_1 + 0.133PC_4^2\right)}$$
$$- 0.354 - 0.133PC_3 - 0.407PC_1$$
$$(31)$$

$$PC_1 = -0.45\hat{x}_1 - 0.389\hat{x}_2 + 0.34\hat{x}_3 + 0.449\hat{x}_4 + 0.318\hat{x}_5$$
$$- 0.41\hat{x}_6 + 0.244\hat{x}_7$$
$$(32)$$

$$PC_2 = 0.391\hat{x}_1 - 0.082\hat{x}_2 - 0.668\hat{x}_3 + 0.209\hat{x}_4 + 0.547\hat{x}_5$$
$$- 0.222x_6 + 0.05\hat{x}_7$$
$$(33)$$

$$PC_3 = -0.013\hat{x}_1 + 0.176\hat{x}_2 - 0.007\hat{x}_3 - 0.3721\hat{x}_4$$
$$+ 0.109\hat{x}_5 + 0.055\hat{x}_6 + 0.903\hat{x}_7 \qquad (34)$$

$$PC_4 = -0.474\hat{x}_1 + 0.726\hat{x}_2 - 0.055\hat{x}_3 + 0.152\hat{x}_4 + 0.41\hat{x}_5$$
$$+ 0.183\hat{x}_6 - 0.146\hat{x}_7$$
$$(35)$$

$$PC_5 = -0.056\hat{x}_1 - 0.33\hat{x}_2 - 0.1\hat{x}_3 + 0.378\hat{x}_4 + 0.049\hat{x}_5$$
$$+ 0.841\hat{x}_6 + 0.162\hat{x}_7$$
$$(36)$$

In Eqs. (32)–(36), $\hat{x}_k$; $k = 1,2,\ldots,7$ were computed using Eq. (10); the mean and standard deviation values pertaining to the FT data are given in Table S5.

As can be seen, all the four GP-based models (Eqs. 14, 20, 25, 31) respectively predicting IDT, ST, HT and FT values have nonlinear forms. It is also noticed that through the four principal components, these models contain terms corresponding to the concentrations of all eight principal metal oxides contained in the coal ashes. This result suggests that there indeed exists nonlinear relationships between the magnitudes of the four AFTs and the weight percentages of the eight metal oxides. Consequently, it can be inferred that nonlinear models are ideally suited for the AFT prediction than the linear ones. The CC, RMSE and MAPE magnitudes in respect of the AFT predictions made by the four GP-based models (Eqs. 14, 20, 25, 31) for the training, test, and validation datasets are listed in Table 3.

5.3 Multilayer perceptron (MLP) neural network based AFT models

The MLP-based four AFT models were trained using *error-back-propagation* (EBP) algorithm in *RapidMiner* data-mining suite (Mierswa et al. 2006; RapidMiner 2007). To secure an optimal MLP model possessing high prediction accuracy and generalization and validation performance, its structural and EBP algorithm-specific parameters, namely the number of hidden layers, number of hidden nodes in each layer, the *learning rate* (η), and the *momentum coefficient* (μ), were varied in a systematic manner. The details of the MLP architecture, EBP-specific parameter values (η, μ), and the type of transfer functions used in obtaining the optimal IDT, ST, HT and FT prediction models are provided in Table 4. The four input nodes in the MLP-based models represent as many principal components computed using the respective data sets given in four Tables (S1–S4) in *Supplementary Material*. The prediction accuracy and the generalization performance of the four optimal MLP-based AFT model have been evaluated in terms of CC, RMSE and MAPE magnitudes (see Table 3); these were computed using the target and the corresponding model predicted values of the four AFTs.

5.4 SVR-based modelling

The SVR based AFT prediction models were also developed using *RapidMiner* data-mining suite (Mierswa et al. 2006; RapidMiner 2007). Specifically, the models were constructed using the widely used ε-SVR algorithm and the kernel functions employed were *ANOVA* and *radial basis function* (RBF). The ε-SVR training algorithm employs three parameters, viz. *kernel gamma* (γ), *regularization constant* (C), and, the *radius of the tube* (ε) for ANOVA and RBF kernels; in the case of ANOVA kernel, an additional parameter namely *kernel degree* is considered. These parameters were varied systematically to obtain optimal SVR models possessing high AFT prediction accuracy and generalization capability. The magnitudes of the stated ε-SVR parameters that produced optimal SVR models and the corresponding number of support vectors are given in Table 5. Table 3 lists the *CC*, *RMSE* and *MAPE* magnitudes pertaining to the SVR model predictions of the four AFTs for both training, test, and validation set data.

5.5 Comparison of AFT prediction models

5.5.1 Initial deforming temperature prediction models

It is noticed in Table 3 that the *CC* magnitudes corresponding to the *IDT* predictions by the GP model are sufficiently high (range: ∼ 0.80 to ∼ 0.85); also the related *MAPE* (range: ∼ 2.20 to ∼ 4.0) magnitudes are low. These magnitudes indicate that the GP-based model possesses good prediction accuracy and the much-desired generalization capability. It is also observed that the *CC*, *RMSE,* and *MAPE* magnitudes in respect of the *IDT* predictions made by the GP-model for training, test and validation set data are superior (high *CC* and low *RMSE* and *MAPE* values) to the corresponding values in respect of the IDT predictions made by the MLP and SVR based models. Figure 8 portrays three parity plots displaying the experimental *IDT* values and those predicted by the GP-, MLP-, and SVR-based models, respectively. In all the three plots, a good match is seen between the experimental and model-predicted *IDT* values. It is also noticed in Fig. 8a that taken together the GP-model predictions for the training, test and validation data exhibit lower scatter compared to the predictions by the MLP and SVR based models thus further supporting the inference of its superior performance based on the *CC*, *RMSE* and *MAPE* values.

5.5.2 Softening temperature prediction models

The *CC*, *RMSE* and *MAPE* values pertaining to the *ST* predictions made by the three CI-based methods listed in Table 3 indicate the following.

(1) There exists a minor variation in the *ST* prediction accuracies and generalization capabilities of the three CI-based models.

(2) Among the CI-based models, the overall prediction and generalization performance of the GP-based model is marginally better than the MLP- and SVR-based models. This observation is unambiguously supported by the lower scatter in the predictions of the GP-based model (see Fig. 9a) when compared with the predictions of the other two models.

5.5.3 Hemispherical temperature predicting models

The *CC*, *RMSE* and *MAPE* magnitudes pertaining to the *HT* predictions by the three CI-based models provide following insights.

(1) The *CC* magnitude corresponding to the predictions of the training set outputs (termed "recall" ability) by the GP-model is lower (0.803) than that of the corresponding MLP (0.926) and SVR (0.813) based models. However, the *CC* magnitudes in respect of the GP-model predictions for the test and validation sets (0.929, 0.953) are higher than that of the MLP (0.894, 0.851) and SVR (0.800, 0.804) models. It may be noted that higher *CC* values pertaining to the test and validation set outputs are indicative of the better generalization ability possessed by the model, which is critically important in correctly predicting the *HT* values for an entirely new set of inputs. The above observations suggest that the GP model possesses better generalization ability than the MLP and SVR based models. This inference is also supported by the parity plots depicted in Fig. 10 where it is seen that although the predictions of the training set outputs by the GP model (shown by "diamond" symbol) exhibit a higher scatter relative to the *HT* predictions by the MLP and SVR models, the GP-model predictions pertaining to the test and validation set data exhibit lower scatter (better generalization) than the predictions by the other two models.

(2) The parity plots in respect of the hemispherical temperature predictions depict that the MLP-based model has yielded better prediction and generalization performance than the GP- and SVR-based models. The MLP model predictions of *HT* have also yielded higher (lower) magnitudes of the

correlation coefficient (*RMSE/MAPE*) relative to the two other CI-based models.

5.5.4 Fluid Temperature predicting models

The *CC*, *RMSE* and *MAPE* trends pertaining to the *FT* predictions made by the three CI-based models are similar to those observed in the prediction of the hemispherical temperature, *HT*. Specifically, it is seen that although the recall ability of the GP-based model is marginally inferior to that of the MLP and SVR models, its crucial generalization capability is better than the stated two models. This trend of better generalization by the GP-model is also witnessed in the parity plots shown in Fig. 11 where it is clearly seen that the scatter corresponding to the predictions in respect of the test and validation set outputs by the GP-based model is lesser than that for the corresponding predictions by the MLP- and SVR- models.

The rigorous Steiger's Z-test (Steiger 1980) was also employed to statistically compare the prediction and generalization performance of the CI-based models. This test examines whether the two correlation coefficients evaluated using the predictions of two competing models are statistically equal. In particular, it tests the null hypothesis (H_0) that two *CC* magnitudes are equal in the statistical sense; that is, $CC_{AB} = CC_{AC}$, where for the present study subscripts *A*, *B*, and *C*, respectively denote the experimental AFT values and those predicted by the models *B* and *C*. In this study, the Steiger's Z-test has examined the validity of the null hypothesis ($CC_{AB} = CC_{AC}$) in respect of the following three pairs of AFT values.

(1) [Experimental (*A*)—GP model predicted (*B*)] and [Experimental (*A*)—MLP model predicted (*C*)]
(2) [Experimental (*A*)—MLP model predicted (*B*)] and [Experimental (*A*)—SVR model predicted (*C*)]
(3) [Experimental (*A*)—SVR model predicted (*B*)] and [Experimental (*A*)—GP model predicted (*C*)]

The results of The Steiger's Z-test test are tabulated in Table 6 and they are indicative of the following.

(1) The performances of the GP, MLP, and SVR models in predicting the *IDT* and *ST* magnitudes are comparable.
(2) In the case of *HT* and *FT* predictions, the performance of the MLP based models is better than that of the the GP and SVR models.

The overall inferences that can be drawn from the results presented above are as follows.

(1) All the four GP-based models (Eqs. 14, 20, 25, and 31) are nonlinear. As stated earlier, depending upon the relationship that exists between the inputs and the output, the GP method can search and optimize an appropriate linear or a nonlinear model from the example input–output data. The nonlinear forms fitted by the GP for predicting all four AFTs are indicative that the relationships between the AFTs and concentrations of seven oxides in coal ashes are nonlinear.

(2) A comparison of the performance of the CI-based models with the existing high performing ones with relatively wider applicability (Seggiani and Pannocchia 2003) indicate that (a) the GP and MLP based models predicting all four AFTs have outperformed the existing ones in terms of possessing better generalization capability, and (b) the GP and MLP based models require lower number of predictors (= 7) than those needed by the models proposed in Seggiani and Pannocchia (2003) that consider 13, 11, 11 and 12 predictors, respectively for *IDT, ST, HT,* and *FT* prediction models. The lower number of predictors used by the GP-, MLP-, and SVR-based models have reduced the effort and cost involved in compiling the predictor data.

(3) The *RMSE* which has the same units as the model predicted output is a measure of how close the model predicted values are to the corresponding experimentally measured ones. It is an absolute measure of the data fit and can be interpreted as the standard deviation of the unexplained variance. It has been observed that the reproducibility of the AFT magnitudes measured by the same analyst and using the same instrument varies between 30 and 50 °C. The corresponding variation between the measurements done at different laboratories is ∼ 50–80 °C (Seggiani and Pannocchia 2003). In Table 3, it is seen that the *RMSE* magnitudes in respect of the test and validation set predictions by the GP-based models predicting *IDT, ST, HT,* and *FT* vary between 33.79 and 82.20. These magnitudes are a measure of the generalization ability of the models. Considering the extent of the inherent variability in the experimental measurements of the AFT values, the stated *RMSE* magnitudes can be considered as reasonable and, therefore, indicative of good prediction and generalization performance of the GP-based models.

(4) Among the three types of CI-based models, the GP-based ones due to their compact size, and ease of evaluation are more convenient to use and deploy in the practical applications. However, in situations when the highest AFT prediction accuracy is required then the MLP based models should be used preferentially for the prediction of *HT* and *FT*.

6 Conclusions

Coal comprises mineral matter, which varies in the form and composition. In the coal-based combustion and gasification processes this mineral matter, predominantly consisting of oxides, is left as *ash* (an incombustible residue). Very often, in coal combustion and gasification processes, ash deposits are formed on the heat absorbing surfaces of the exposed process equipment. These deposits give rise to the undesirable slagging and/or fouling phenomenon and their deleterious effects such as corroding of the furnace components, channel burning, pressure drop in the heat transfer equipment, and unstable process operation. *Ash fusion temperatures* (AFTs) are the important characteristics of the coal ashes and signify the temperature range over which the ash deposits are formed on the heat absorbing surfaces of the process equipment. The currently available models for AFT prediction have following characteristics.

(1)   They are predominantly linear models although a detailed scrutiny of the data indicates that the relationships between the AFTs and the weight percentages of some of the mineral matter constituents could be nonlinear.

(2)   The models are developed using data of coals belonging to a limited number of geographical regions and, therefore, do not have wider applicability since coal properties differ widely depending upon coal's geographical origin,

To address the above-stated issues pertaining to the existing AFT prediction models, this study has used three computational intelligence (CI) based exclusively data-driven formalisms, namely, genetic programming (GP), multi-layer perceptron (MLP) neural network, and support vector regression (SVR) for developing the nonlinear models. Also, a large number of data pertaining to the ashes of coals from multiple geographies have been utilized in the model development. These characteristics have imparted a wider applicability to the CI-based models. Among the three CI-based methods, the GP formalism has a unique capability that depending upon the relationships between the magnitudes of the oxides and AFTs in the example data, it can search and optimize a linear or a nonlinear data-fitting function. All the four best fitting GP-based models developed in this study for the prediction of four ash phase temperatures possess nonlinear forms. This result clearly indicates that the relationships between the weight percentages of the eight oxides (predictor variables) in the coal ashes and the corresponding AFTs are indeed nonlinear. A comparison of the prediction and generalization performance of the three CI-based models indicates that (a) the performance of the GP, MLP and SVR models

in predicting *IDT* and *ST* magnitudes is comparable, and (b) in the case *HT* and *FT* predictions, the performance of the MLP based models is better than the GP and SVR models. Owing to their parsimonious (less complex) nature, GP-based models are easy to understand and use. Since they are endowed with good AFT prediction and generalization performance and wider applicability, the CI-based models developed in this study possess a potential to be the preferred ones for predicting AFT magnitudes of coal ashes from different geographies in the world.

Compliance with ethical standards

References

Alpern B, Nahuys J, Martinez L (1984) Mineral matter in ashy and non-washable coals—its influence on chemical properties. Comun Serv Geol Portugal 70(2):299–317

Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press, Oxford

Bryers RW (1996) Fireside slagging, fouling, and high-temperature corrosion of heat-transfer surface due to impurities in steam raising fuels. Prog Energy Combust Sci 22:29–120

Gao F, Han P, Zhai Y-J et al (2011) Application of support vector machine and ant colony algorithm in optimization of coal ash fusion temperature. In: Proceedings of the 2011 international conference on machine learning and cybernetics, Guilin, 10–13 July 2011

Geladi P, Kowalski BR (1986) Partial least squares regression (PLS): a tutorial. Anal Chim Acta 85:1–17

Ghugare SB, Tiwary S, Elangovan V et al (2014) Prediction of higher heating value of solid biomass fuels using artificial intelligence formalisms. Bioenergy Res 7:681–692. https://doi.org/10.1007/s12155-013-9393-5

Goel P, Bapat S, Vyas R, Tambe A et al (2015) Genetic programming based quantitative structure-retention relationships for the prediction of Kovats retention indices. J Chromatogr A 1420:98–109. https://doi.org/10.1016/j.chroma.2015.09.086

Goni C, Helle S, Garcia X, Gordon A et al (2003) Coal blend combustion: fusibility ranking from mineral matter composition. Fuel 82:2087–2095

Gray VR (1987) Prediction of ash fusion temperature from ash composition for some New Zealand coals. Fuel 66:1230–1239

Huggins FE, Kosmack DA, Huffman GP (1981) Correlation between ash-fusion temperatures and ternary equilibrium phase diagrams. Fuel 60:577–584

Hurst HJ, Novak F, Patterson JH (1996) Phase diagram approach to the fluxing effect of additions of $CaCO_3$ on Australian coal ashes. Energy Fuels 10:1215–1219

Ivanciuc O (2007) Applications of support vector machines in chemistry. Rev Comput Chem 23:291–400. https://doi.org/10.1002/9780470116449

Jak E (2002) Prediction of coal ash fusion temperatures with the F*A*C*T thermodynamic computer package. Fuel 81:1655–1668

Kahraman H, Bos F, Reifenstein A et al (1998) Application of a new ash fusion test to Theodore coals. Fuel 77:1005–1011

Karimi S, Jorjani E, Chelgani SC et al (2014) Multivariable regression and adaptive neurofuzzy inference system predictions of ash fusion temperatures using ash chemical composition of us coals. J Fuels. https://doi.org/10.1155/2014/392698

Koza JR (1992) Genetic programming: on the programming of computers by means of natural selection. A Bradford Book. MIT Press, Cambridge

Kucukbayrak S, Ersoy-Mericboyu A, Haykiri-Acma H et al (1993) Investigation of the relation between chemical composition and ash fusion temperatures for soke Turkish lignites. Fuel Sci Technol Int 1(9):1231–1249. https://doi.org/10.1080/08843759308916127

Liu YP, Wu MG, Qian JX (2007) Predicting coal ash fusion temperature based on its chemical composition using ACO-BP neural network. Thermochim Acta 454:64–68

Lloyd WG, Riley JT, Zhou S, Risen MA (1995) Ash fusion temperatures under oxidizing conditions. Energy Fuels 7:490–494

Lolja SA, Haxhi H, Dhimitri R et al (2002) Correlation between ash fusion temperatures and chemical composition in Albanian coal ashes. Fuel 81:2257–2261

McConaghy T, Vladislavleva E, Riolo R (2010) Genetic programming theory and practice 2010: an introduction, vol VIII. Genetic programming theory and practice. Springer, New York

Miao S, Jiang Q, Zhou H et al (2016) Modelling and prediction of coal ash fusion temperature based on BP neural network. In: MATEC web of conferences, vol 40, pp 05010. EDP Sciences. https://doi.org/10.1051/matecconf/20164005010

Ozbayoglu G, Ozbayoglu ME (2006) A new approach for the prediction of ash fusion temperatures: a case study using Turkish Lignites. Fuel 85:545–562. https://doi.org/10.1016/j.fuel.2004.12.020

Patterson JH, Hurst HJ (2000) Ash and slaq qualities of Australian bituminous coals for use in slagging gasifiers. Fuel 79:1671–1678

Poli R, Langdon W, Mcphee N (2008) A field guide to genetic programming. http://www0.cs.ucl.ac.uk/staff/wlangdon/ftp/papers/poli08_fieldguide.pdf. Accessed 5 Sept 2018

RapidMiner (2007). https://rapidminer.com/products/studio. Accessed 5 Sept 2018

Rhinehart RR, Attar AA (1987) Ash fusion temperature: a thermodynamically-based model. Am Soc Mech Eng 8:97–101

Rumelhart D, Hinton G, Williams R (1986) Learning representations by back propagating error. Nature 323:533–536

Schmidt M, Lipson H (2009) Distilling free-form natural laws from experimental data. Science 324(5923):81–85

Seggiani M (1999) Empirical correlations of the ash fusion temperatures and temperature of critical viscosity for coal and biomass ashes. Fuel 78:1121–1125

Seggiani M, Pannocchia G (2003) Prediction of coal ash thermal properties using partial least-squares regression. Ind Eng Chem Res 42:4919–4926. https://doi.org/10.1021/ie030074u

Sharma S, Tambe SS (2014) Soft-sensor development for biochemical systems using genetic programming. Biochem Eng J 85:89–100

Skrifvars B-J, Lauren T, Hupa M et al (2004) Ash behavior in a pulverized wood fired boiler-a case study. Fuel 83:1371–1379

Slegeir WA, Singletary JH, Kohut JF (1988) Application of a microcomputer to the determination of coal ash fusibility characteristics. J Coal Qual 7:248–254

Song WJ, Tang LH, Zhu XD (2010) Effect of coal ash composition on ash fusion temperatures. Energy Fuels 24:182–189

Steiger JH (1980) Tests for comparing elements of a correlation matrix. Psychol Bull 87(2):245–251

Tambe SS, Kulkarni BD, Deshpande PB (1996) Elements of artificial neural networks with selected applications in chemical engineering, and chemical and biological sciences. Simulation & Advanced Controls Inc., Louisville

Van Dyk JC, Keyser MJ, Van Zyl JW (2001) Suitability of feedstocks for the Sasol–Lurgi fixed bed dry bottom gasification process. In: Gasification technology conference, Gasification Technologies Council, Arlington, Paper 10-8

Vapnik V (1995) The nature of statistical learning theory. Springer, New York

Vassilev SV, Kitano K, Takeda S et al (1995) Influence of mineral and chemical composition of coal ashes on their fusibility. Fuel Process Technol 45:27–51

Vorres KS (1979) Effect of composition on melting behaviour of coal ash. J Eng Power 101:497–499

Vyas R, Goel P, Tambe SS (2015) Genetic programming applications in chemical sciences and engineering, chapter 5. In: Gandomi AH, Alavi AH, Ryan C (eds) Handbook of genetic programming applications. Springer, Switzerland, pp 99–140. https://doi.org/10.1007/978-3-319-20883-1

Wall TF, Creelman RA, Gupta RP et al (1998) Coal ash fusion temperatures—new characterization techniques, and implications for slagging and fouling. Prog Energy Combust Sci 24:345–353

Winegartner EC, Rhodes BT (1975) An empirical study of the relation of chemical properties to ash fusion temperatures. J Eng Power 97:395–406

Mierswa I, Wurst, M, Klinkenberg, R et al (2006) Rapid prototyping for complex data mining tasks, In: Proceeding of the 12th ACM SIGKDD international conference on knowledge discovery and data mining, 2006, pp 935–940. https://doi.org/10.1145/1150402.1150531

Yin C, Luo Z, Ni M et al (1998) Predicting coal ash fusion temperature with a back-propagation neural network model. Fuel 77:1777–1782

Zhang G, Patuwo BE, Hu MY (1998) Forecasting with artificial neural networks: the state of the art. Int J Forecast 14(1):35–62

Zhao B, Zhang Z, Wu X (2010) Prediction of coal ash fusion temperature by least squares support vector machine model. Energy Fuels 24:3066–3071

Zurada JM (1992) Introduction to artificial neural systems, vol 8. West, St. Paul