A Collaborative Design Platform for New Alloy Material Development

Gongzhuang Peng[1], Youzhao Sun[1], Qian Zhang[2], Quan Yang[1], Weiming Shen[3]

1. National Engineering Research Center for Advanced Rolling and Intelligent Manufacturing, University of Science and Technology Beijing, Beijing, China, 100083
2. Department of Electronics and Electrical Engineering, Liverpool John Moores University, Liverpool L3 3AF, U.K
3. The State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, China, 430074

To overcome the shortcomings of the conventional trial and error mode for new material development, a full-process collaborative design platform for steel rolling is developed based on an industrial internet of things (IIoT) system in this study. Equipment, process and product entities are modeled in both the physical domain and the cyber domain. A systematic data-driven Mamdani-type fuzzy modelling methodology is proposed to map the relationship between material chemical compositions, organizational structures, process parameters and mechanical performances. The proposed methodology employs a random forest (RF) algorithm to select important parameters from mechanism models, simulation models and production process variables, utilizes a *K*-means algorithm to merge diverse steel grades into sub-clusters, and implements a multi-objective particle swarm optimization (MOPSO) algorithm to further improve the fuzzy model in terms of both the structure and the membership function parameters. A dataset of 3,500 steel coils collected by the prototype platform built in a large hot rolling mill is used to evaluate the performance of the proposed approach. Experiment results show that the proposed methodology performs well in predicting the yield strength, tensile strength and elongation, with the coverage probability over 90% under 10% deviation and about 70% under 5% deviation on average.

Keywords: new material development, collaborative design platform, mechanical performance prediction, Mamdani-type fuzzy modelling, industrial internet of things (IIoT)

1. Introduction

The development of new materials is to determine the chemical compositions and process parameters of each production step to yield the final mechanical and physical properties, such as the tensile strength, yield strength, elongation, hardness and corrosion behavior [1]. In the field of material science, the conventional product design process includes several stages, namely, requirements collection, discovery, small batch production, property optimization, manufacturing, deployment and so on [2]. This trial and error development mode is not only very time-consuming and costly, but also highly depending upon engineer experience accumulated from previous

projects. Due to the multidisciplinary and highly contextual nature of engineering design, engineering or scientific teams from different institutions require some mechanisms of knowledge sharing and effective collaboration [3][4]. Thus, how to quickly design or adjust chemical compositions and key processes of materials according to customer needs is an urgent problem in both industry and academia.

The booming of modern information technologies, especially high performance computing, big data and artificial intelligence, have greatly accelerated the new material development process [5] [6]. The Materials Genome Initiative (MGI), laid out by US President Barack Obama in 2011, fueled the integration of machine learning methods and material design and discovery [7]. Shen et al. [8] applied the least squares regression method to predict the electrical performance of nanocomposites materials based on high-throughput simulations and proposed a theoretical framework for the design of high-density energy storage materials. Lu et al. [1] established a structure-property mapping relationship between hybrid organic-inorganic perovskites (HOIPs) bandgap and material features, such as iron radii, tolerance factor and so on. Two lead-free HOIPs with proper bandgap and excellent environmental stability are discovered from 5,158 unexplored candidates for solar cells. Wan et al. [9] established a PSO-BP artificial neural network model to predict the hot deformation behaviors of Zr-4 alloy.

Since the development of new materials involves the collaboration of multiple processes and departments, how to integrate mechanism and process knowledge between different disciplines has become a crucial problem. Based on the newly rising cyber physical systems (CPS) [10][11], a collaborative design platform for new alloy material development is introduced in this study, which is composed of the physical domain, the cyber domain and the data domain. In the cyber domain, a multiscale, multidisciplinary digital mapping of physical process is built, and the simulation and optimization of production through data analysis and decision-making are realized. Product-related requirements, production, testing, and usage data are automatically collected into the data domain with the help of an IIoT system. Through design knowledge management, product quality tracking and design optimization, and full-process design process visualization, the digitization and intelligence of material development is realized. As the core part of the product quality tracking, the material mechanical performance prediction based on data, mechanism model and domain knowledge is presented in detail.

At present, there are two main approaches to establish the structure and mechanical performance prediction model [12]. One is to build the strengthening mechanism mathematical models based on the simulation and laboratory experimental results, such as Nitride precipitation and austenite transformation [13]. Li et al. [14] established a temperature-dependent quantitative relationship between yield strength and Young's modulus as well as Poisson's ratio for the metallic materials based on the equivalence between heat energy and distortional strain energy. Bambacha et al. [15] created the grain size calculation and mathematical stability model for the multi-pass hot rolling processes, which determines the product properties from the micro level. Shao et al. [16] studied the changes of microstructures and tensile properties of medium-Mn steel in hot rolling at different critical annealing temperatures by trial production experiments. The other type is to use machine learning algorithms to establish black-box models of the relation between process parameters and performances after collecting a large amount of historical production data [17]. Mohanty et al. [17] designed an online mechanical property prediction system for hot rolled IF steel with a 22-input neural network. Xu et al. [18] converted the chemical compositions and

processing parameters into two-dimension images to extract features and proposed a convolutional neural network (CNN)-based method to predict the mechanical properties of hot rolled steel. However, there are some drawbacks for both the mechanism models and the data-driven black-box models. An accurate first-principles based mathematical model is difficult to establish, since the rolling process is a very complex and dynamic nonlinear system. On the other hand, the black-box NN models lack interpretability and the knowledge behind the models cannot be understood fully. To better utilize the big data and expert knowledge, this paper proposes an optimized fuzzy inference model, which has been widely used for modeling complex systems, nonlinear identification, regression and classification tasks [21], to predict the mechanical properties of materials based on the collaborative product development platform.

The rest of this paper is organized as follows. Section 2 introduces the proposed collaborative design platform for new alloy material development, detailing the physical domain, cyber domain and data domain. The parameters selection of computational models based on a random forest (RF) algorithm is analyzed and described in Section 3. Then in Section 4, the mechanical properties prediction based on an optimized Mamdani fuzzy model is presented, including the grouping of steel grades. Finally conclusions are drawn in Section 5, together with some discussion of potential future work.

2. Proposed collaborative design platform for new alloy material development

In the material collaborative design platform as shown in Figure 1, a full-process digital model is established to interact with the physical entities (such as equipment, process, and product), with the help of production and simulation data. In order to meet users' requirements for material performance and quality, the corresponding relationship between material compositions, organizational structures, process parameters and material performances is explored in the cyber domain, and the experiments and production verification are conducted in the physical domain. Through a few small-batch trials and iterations, the optimized compositions and process parameters are obtained. The modules involved in the collaborative design platform are detailed below.

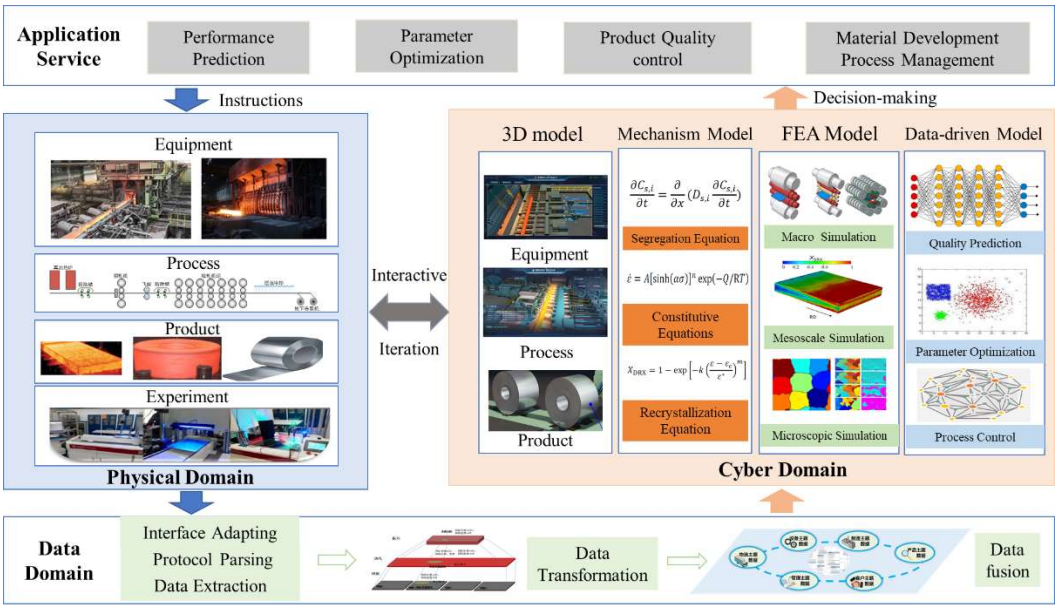


Figure 1 Collaborative design platform for alloy material

2.1 Physical domain

Physical entities are the basis of the system, which include production processes, equipment and various forms of steel products. Iron and steel-making is a typical process production, which is characterized by heterogeneous, dynamic and multi-scale. The production equipment covers dozens of sub-processes, from coking, ironmaking, steelmaking to casting and rolling, and contains a variety of types, including machinery, hydraulics, instrumentation and control equipment. Equipment has attributes such as health statuses, production capacities, and working conditions, which are directly related to the product quality and output.

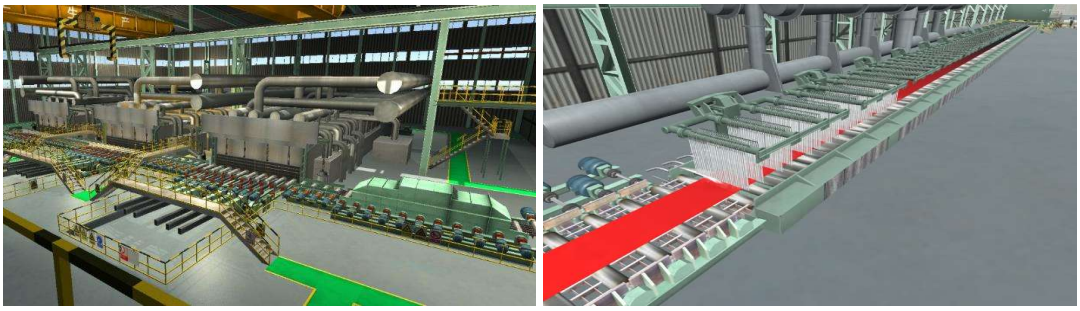
Production processes of iron and steel-making include the interconnection and cooperative operation of material flow, energy flow, and information flow. The material flow such as the iron undergoes a series of complex physical and chemical reactions to achieve the phase transformation and deformation driven by the energy flow, mainly the carbon, ultimately achieving the optimization of dissipation under specific environmental conditions. Information flow is the carrier of characterization and production control of the material flow and energy flow during the manufacturing process.

Products are generated at every stage of the steel production, including continuous casting billets, slabs, hot-rolled strips, and cold-rolled strips. According to different usages and performances, there are hundreds of types of products at each stage. The products of the upstream process will have a genetic influence on the quality of the downstream process. For example, the crown and crack of the hot rolling process have a direct effect on the shape and surface of the cold rolling process. Therefore, it is necessary to establish a digital model of the whole production process at different levels to achieve the synergy and optimal control of products, processes and quality.

2.2 Cyber domain

1) Three-dimensional visualization models

The three-dimensional visualization model can provide multi-dimensional, multi-temporal-scale high-fidelity digital mapping for physical entities, including 3D geometric models, kinematics models, and dynamics models. The 3D geometric models display the static characteristics of the actual physical objects, including the component parts, spatial structures, assembly relationship and other attributes of the equipment. It is developed through 3D modeling software, such as SolidWorks, 3DMAX, Pro, etc. The kinematic models reflect the motion characteristics of the physical objects, including attributes such as control logic and production execution actions. Through the analysis of real-time or historical production data by the motion state observer, combined with the automation control model, the kinematic models are driven to synchronize with the production state of physical objects, including linear motion, rotary motion, and multi-axis synchronized motion. The dynamic models describe the inherent characteristics of the physical objects, including vibration and inertia, elastoplastic deformation (such as shape wave, bending, and deviation), and potential interaction properties between different objects.



a) 3D visualization model of the rolling production line

b) Equipment model - finishing mill c) Product model - digital coil

Figure 2 3D visualization model of an actual rolling factory

Figure 2 shows some 3D visualization models of an actual rolling factory. Figure 2a) displays the rolling production line, including sub-processes from slab storage, heating furnace to rough rolling, finishing rolling, and crimping. In this process, the slab is finally processed into products that meet the quality requirements, mainly the size, temperature, plate shape, surface, etc. In addition to mapping the real-time status of production processes and materials, the rolling rhythm, output, energy consumption of gas, water, and electricity related to the production line are all

synchronized to the 3D visualization model. Figure 2b) shows the visualization model of the finishing roll equipment. Based on the historical data, the accuracy change and state evolution of its service processes are monitored. Through the correlation analysis with process data such as production stability, product quality and control accuracy, the influence between the service quality of equipment and production status are mined. Figure 2c) is a visualization model of the steel coil. Process data, equipment data, operation records, energy consumption data, quality data, cost data, and delivery data in the full life cycle of product production are integrated into the cyber steel coils.

2) Computational models

The computational models are the core of the design platform, including mechanism models, simulation models and data-driven models. Among them, mechanism models are theoretical formulas between the mechanical properties and internal structures of the strip steel as well as the process parameters based on the law of organization evolution and strengthening mechanism in the actual production process. Simulation models are combinations of mathematical modeling and computing solution to simulate various physical and chemical processes, including mechanics, fluids, thermodynamics, dynamics and other fields. The data-driven models build black-box or grey-box models to establish the mapping relationship between inputs and outputs through the learning of historical data.

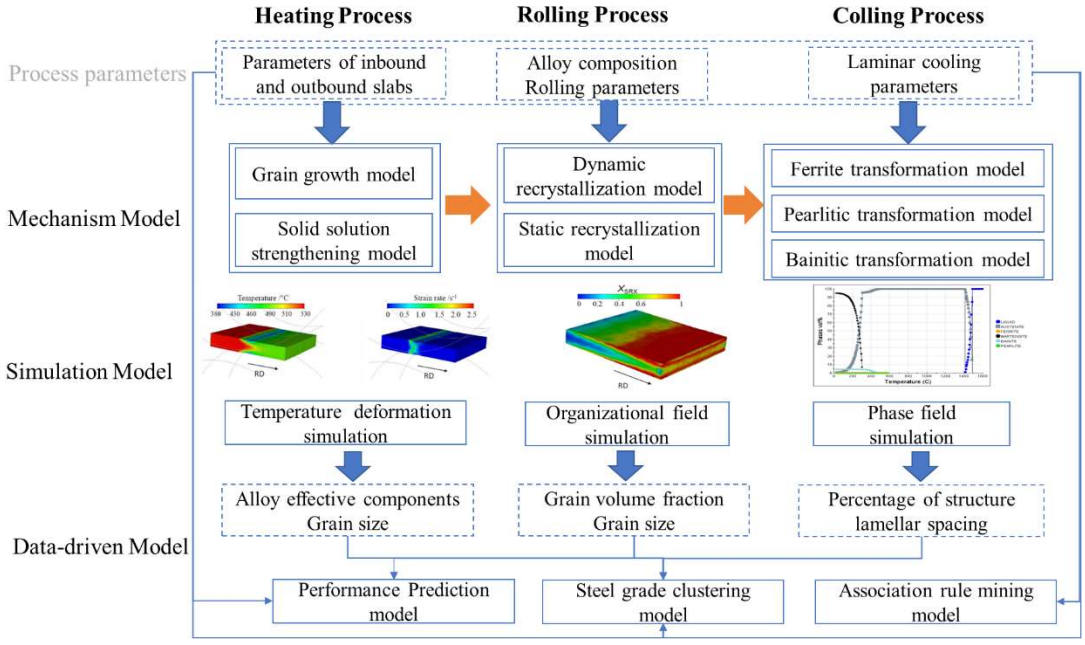


Figure 3 Computational model of the rolling production

a) Mechanism models

Small and uniform crystal grains help stabilize the strength, plasticity and toughness of the metal. As shown in Figure 3, in the heating process, the grain growth model and the microalloying re-dissolution model lead the major organization evolution, where the heating temperature and holding time play a decisive role. In the rolling process, the recrystallization model is mainly considered, where the deformed metal is heated at a high temperature. The elongated and broken crystal grains then re-nucleate and grow into new uniform and fine equiaxed crystals due to the

increased atomic diffusion capacity. After the deformed metal is recrystallized, the strength and hardness of the metal are significantly reduced, while the plasticity and toughness are greatly improved. The cooling process is mainly related to the phase transition model, during which the type, shape and distribution of precipitates have direct impacts on the strength and toughness of the strip.

Taking the recrystallization during rough rolling as an example, dynamic recrystallization occurs when the deformation in a single pass is greater than the critical strain. The percentage of dynamic recrystallization X_d can be expressed by the following formula.

$$X_d = 1 - \exp ¿¿ \quad (1)$$

where ε is the single pass strain,ε_c is the critical strain for dynamic recrystallization, ε_p is the peak strain at which the stress reaches the peak value, k, n, a_3 are material constants related to the steel type.

The grain size after dynamic recrystallization is:

$$d_{drx} = a Z^b \quad (2)$$

where a and b are material-related constants, Z is Zener-Hollomon parameter which is calculated as (3), in which Q_1 is dynamic recrystallization activation energy, T_{drx} is deformation temperature.

$$Z = \varepsilon \exp (\frac{\dot{Q}_1}{R T_{drx}} ¿) ¿ \quad (3)$$

b) Simulation models

As shown in Figure 3, simulation models in the rolling production mainly include the temperature field, stress field, microstructure field and phase field model. Among them, the temperature field and stress field are mainly simulated by the finite element method (FEM). The principle of FEM is to discretize the solution domain of the continuum into several nodes and connect them to each other through the nodes on their boundaries. According to the variational principle, finite element equations for solving the unknown variables of the nodes are established. Commonly used simulation software tools in material design include ABAQUS and MARC. The simulation of microstructure field can be realized by FEM or cellular automata (CA). CA is a local dynamic system with discrete time and space. By simply determining the interaction rules between adjacent cells, complex evolutionary phenomena can be simulated. In the process of recrystallization, the cell transformation rules include thermal activation mechanism, grain boundary migration mechanism and energy dissipation mechanism. Phase field simulation (PFS) reflects the phase transitions, transformations and microstructures evolution. It bridges the gap between atomic simulations such as density functional theory (DFT), molecular dynamics (MD), Monte Carlo (MC), etc., and macro continuum methods. Except for some commercial software packages, such as Thermo-calc, more and more open-source software packages that have integrated PFS capabilities emerged, including MOOSE and PRISMS [25].

c) Data-driven models

The expressions involved in the mechanism models and simulation models are usually complex, and it is difficult to establish an accurate mathematical model of the organization evolution under different working conditions. Furthermore, steel production has significant nonlinearity, time-

varying, strong coupling and multi-parameter characteristics. Since most variables in the model cannot be detected in real time, necessary assumptions and simplifications are required, which will inevitably lead to the modeling error. A large number of running logs and production data has accumulated with big data characteristics such as multi-source heterogeneous, multi-time scale, and multi-space-time time series correlation. A hybrid modeling method that combines mechanism models and data models has become the frontier development direction of complex industrial quality modeling and optimization control system.

As shown in Figure 3, the data models of the product design platform include performance prediction model, steel grades merging model and association rule mining model. Among them, the prediction model refers to the use of machine learning algorithms to establish a mathematical model between performances and organizations or process parameters. There are hundreds of steel grades in actual production, and some steel grades have less measured performance data. The merging model is used to cluster different steel grades to improve the applicability of the prediction model. The association rule model is to mine the parameters with the greatest impact on product quality from thousands of process parameters, which improves the performance of the prediction algorithm, and also facilitates the reverse optimization of the parameters.

2.3 Data domain

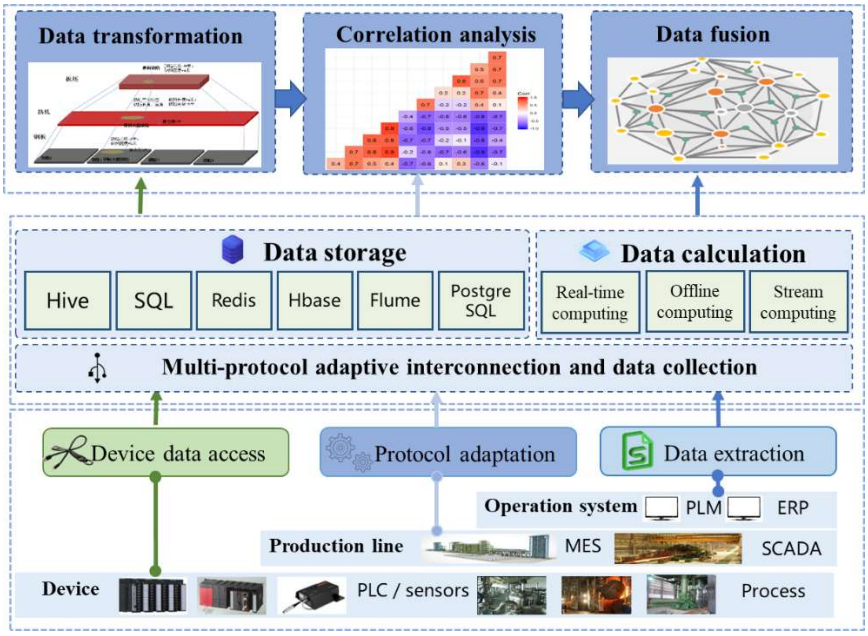


Figure 4 Data domain of the rolling factory

After decades of development, steel industries have generally formed a mature five-level information system, i.e., the equipment control system (L1), the process control system (L2), the workshop level manufacturing execution (MES) system (L3), enterprise resource planning (ERP) system (L4) and the inter-enterprise management decision support system (L5). In the production and operation processes, the five-level systems continuously generate data, which differ greatly in data protocols, data types, and real-time performance. To solve the problem of protocol

incompatibility and intercommunication difficulty, an End-Edge-Cloud collaborative IIoT platform (see Figure 4) to provide services for data collection and storage is established, which supports highly reliable equipment interoperability and multi-source heterogeneous data integration. Industrial field devices are connected to the edge layer of the platform through industrial communication protocols such as industrial ethernet and buses, and wireless protocols such as 4G/5G and NB-IoT. Regarding the multi-source, heterogeneous and multi-granularity feature of industrial data, different interface protocols are adapted through plug-in drivers. Protocol parsing middleware technology is compatible with various protocols such as Modbus, OPC, CAN, Profibus, and software communication interfaces are applied to achieve data format conversion and unification. HTTP, MQTT and other protocols are used to transmit data from the edge to the cloud to achieve remote data access. Multi-source heterogeneous data are then integrated into a hybrid cloud-based storage platform that supports stream computing, real-time computing and offline computing, creating a global data space for the collaborative design platform.

3. Parameters selection from computational models

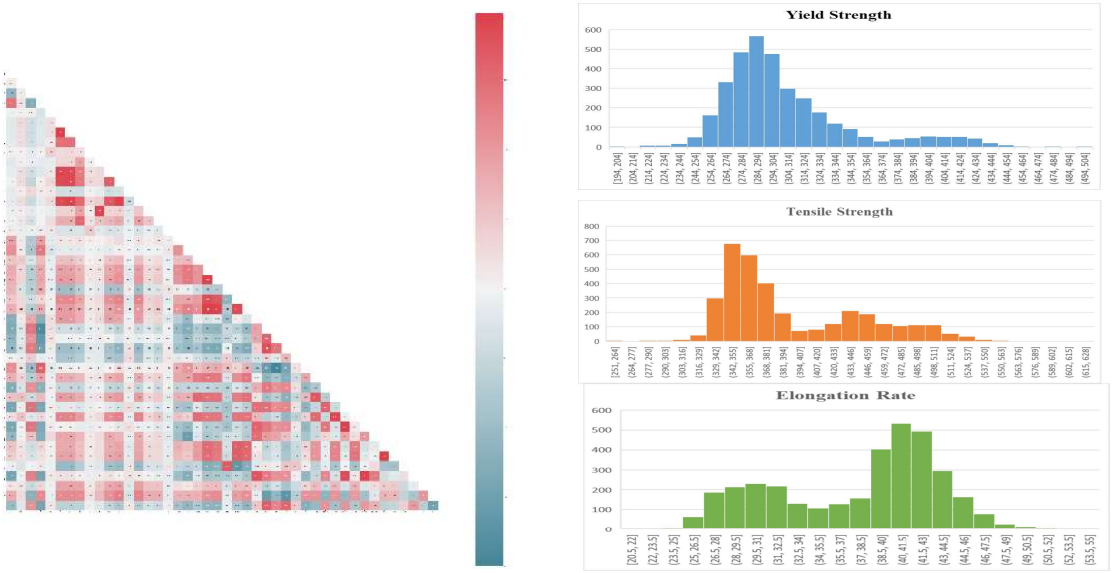
This paper takes the CSP (Compact Strip Production) production line of a steel plant as an example, which is a cutting-edge rolling technology with the advantages of compact process, low investment and low energy consumption. The chemical compositions, material parameters as well as the process parameters are regarded as the design variables, and the mechanical properties are optimized including yield strength (YS), tensile strength (TS) and elongation (ER). In addition to the main chemical compositions - Fe, the steel product contains dozens of standing elements, incidental elements, cryptic elements and alloying elements. Among them, C, N and Nb, Ti, V will form compounds that inhibit the recrystallization of austenite, while P, Ca, and Si have the effect of solid solution strengthening. As a good deoxidizer and desulfurizer, Mn will coarsen crystal grains. The material parameters mainly include the slab thickness and the final strip thickness, which are related to the reduction rate of finishing mill stand and indirectly affect the crystal transformation. To meet the final properties performance, the chemical compositions has to cooperate with the appropriate process parameters. The reduction ratio, rolling temperature, and rolling velocity of each pass are the key process parameters to ensure complete recrystallization, while the cooling rate is significant for austenite transformation.

Table 1 A part of the selected variables

	Variables	Unit	Range		Mean	Description
Chemical composition	C	wt-%	0.022	0.204	0.073	
	N	wt-%	0	0.053	0.005	
	Mn	wt-%	0.086	1.5	0.157	
	S	wt-%	0	0.018	0.003	
	Si	wt-%	0.00	0.27	0.045	

Material parameters	THK	mm	2 54.12	90	67.82	Slab thickness
	IRT	℃	973	1203	1120	Initial rolling temperature
	RV-F1	m/s	0.333	0.826	0.576	Rolling velocity of F1
	RT-F1	℃	943	1098	1051	Rolling temperature of F1
Process parameters	RDR-F4	%	16.1	42.3	32.6	Rack drop rate
	CR	℃/s	5.34	65.4	28.9	Cooling Rate
	AGS-F1	μm	20.3	81.7	29.4	Austenite grain size of F1
	AGR-F4	%	20.8	54.6	43.1	Austenite grain growth rate – F4
	RR-F7	%	0.1	1	0.5	Recrystallization rate
Model parameter	FPTT	℃	815	898	865	Ferrite phase transition temperature
	FVF	%	5	100	90	Ferrite volume fraction
	FAG	μm	1.6	21	11	Ferrite grain size
Physical properties	YS	MPa	194	504	307	Yield Strength
	TS	MPa	251	626	394	Tensile Strength
	ER	%	20.5	54.5	38	Elongation Rate

About 4,000 samples with more than 500 dimensions of parameters are collected by the IIoT platform. Forty-two parameters are selected as inputs based on mechanism models and process knowledge. Table 1 lists the value range and descriptions of key parameters. Figure 5 a) illustrates the closeness of the correlation between various variables through the correlation matrix, and Figure 5 b) shows the distribution histogram of the sample's mechanical properties.



a) Variable correlation matrix b) Histogram of physical properties

Figure 5 Correlation analysis and properties distribution

In order to further reduce the dimensionality of the data-driven model, this paper uses a random forest (RF) algorithm to calculate the importance of features, which is measured by the impurity index of each variable. The impurity of the dataset can be evaluated by the Gini coefficient, which describes the uncertainty degree of a random variable, and has been used as an index for selecting the optimal feature in a decision tree for classification problems. The Gini coefficient is expressed by the following formula.

$$GI_m = \sum_{k=1}^{|K|} \sum_{k' \neq k} p_{mk} p'_{mk} = 1 - \sum_{k=1}^{|K|} p_{mk}^2 \tag{4}$$

where GI_m is the Gini index of the feature m, K is the number of categories of feature m, and p_{mk} is the proportion of category k in feature m.

The importance of the *j*-th feature of the sample in the dataset can be represented by the amount of change in the Gini index before and after the *n*-th node branch of the decision tree.

$$VIM_{jn}^{(Gini)} = GI_n - GI_l - GI_r \tag{5}$$

where GI_l is the Gini index of the left node, and GI_r is the Gini index of the right node after the branch.

Figure 6 shows the feature importance of 42 parameters for YS, TS and ER. It can be seen that different parameters have different effects on the three mechanical properties. In the performance prediction algorithm, the corresponding high-importance parameters will be selected for model training according to the prediction index.

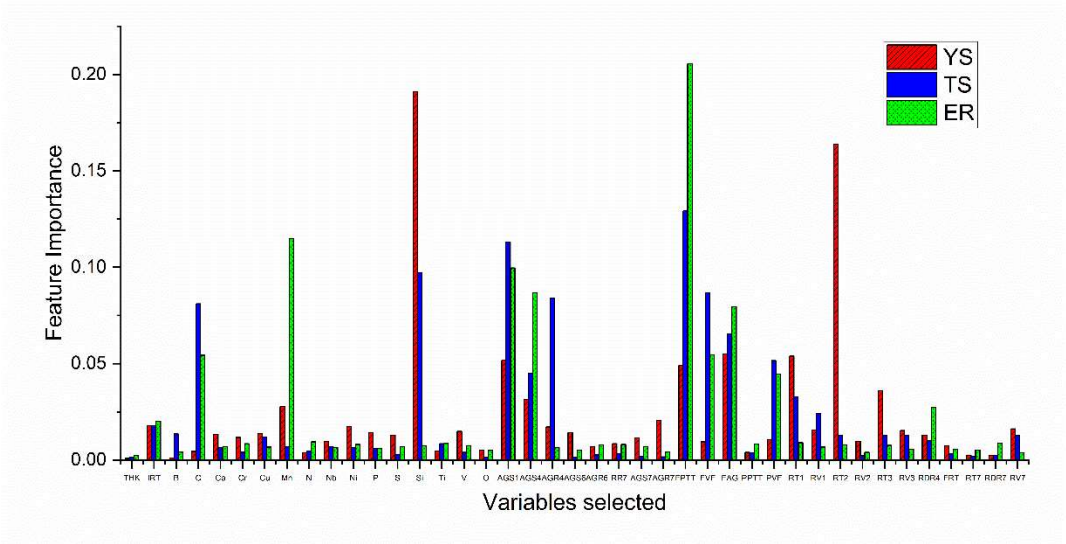


Figure 6 The feature importance for YS, TS and ER

4. Mechanical properties prediction based on the optimized fuzzy model

4.1 Grouping of steel grades

As described above, the establishment of a mapping relationship between mechanical properties and chemical compositions as well as process parameters is the core of new material development. A data-driven mechanical performance prediction model is designed by combining the mechanism models and simulation models in the collaborative design platform. There are often hundreds of different grades of steel in a steel industry, some of which have similar chemical compositions and properties. We first take advantages of the clustering method to automatically merge diverse grades into sub-clusters, which effectively reduces the search space of design variables and facilitates flexible production of multiple varieties and small scales. In this paper, the *K*-means algorithm is used to conduct steel merging on 4,000 samples covering 35 different grades of steel. Calinski-Harabasz (CH) and Silhouette Coefficient (SC) are chosen as the evaluation indices to determine the number of clusters.

The CH index is the ratio of dispersity to closeness, as shown in (6). $tr(B)$ represents the closeness within clusters and is calculated by the sum of distances between individual points in a cluster and the center of the cluster, while $tr(W)$ represents the dispersity between clusters and is calculated by the sum of the distances between the center point of each cluster and the center point of the dataset. From the definition it can be drawn that a larger CH index means higher similarity within the cluster and lower similarity between clusters, resulting in good clustering performance.

$$CH(K) = \frac{tr(B)/(K-1)}{tr(W)/(N-K)} \tag{6}$$

where $tr(B) = \sum_{j=1}^{k} \left\| z_j - z \right\|^2$, $tr(W) = \sum_{j=1}^{k} \sum_{x^i \in z_k} \left\| x_i - z_j \right\|^2$, z is the mean of the entire dataset, z_j is the mean of the *j*-th cluster, *N* is the number of samples, and *K* represents the number of clusters.

The SC index is calculated as (7):

$$SC = \frac{1}{N} \sum_{i=1}^{N} \frac{b_i - a_i}{\max(a_i, b_i)} \tag{7}$$

where a_i represents the average distance between sample *i* and other samples in its cluster, while b_i is the average distance between sample *i* and samples in other clusters.

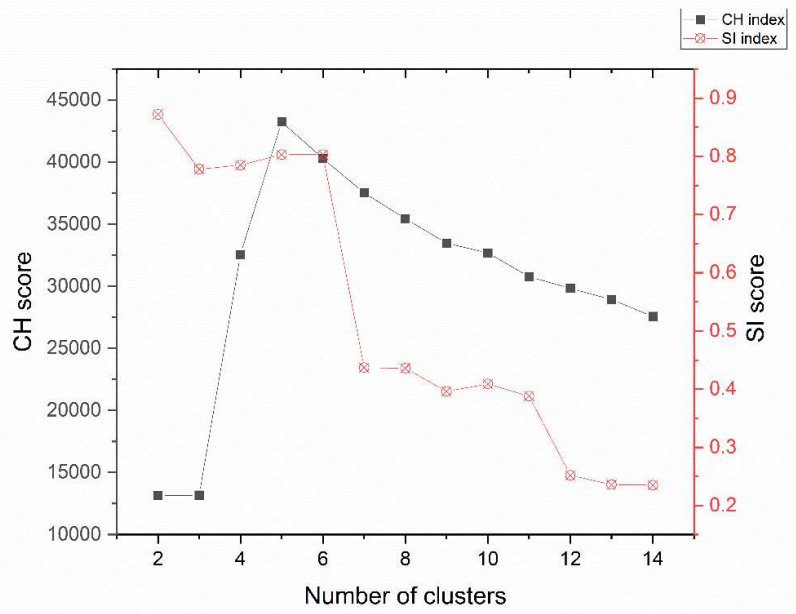


Figure 7 Index values of different cluster number

Figure 7 shows the change curve of the two index values when the number of clusters ranges from 2 to 14. It can be seen from the figure that when the number of clusters is 5, the clustering performance corresponding to the two indexes are both optimal. Table 2 lists the representative steel types and the number of steel grades in individual clusters when 35 steel types are grouped into 5 clusters.

Table 2 Steel grades merging results

Group ID	Number of steel grades	Representative steel grade
A	10	SAE1006B，SPHC2, SPHC-YH
B	5	SPA-H,
C	8	Q235B，
D	9	Q345B-L, SAPH440-P1
E	3	Q235B-L

4.2 Mamdani fuzzy model optimized by MOPSO

1) Mamdani based fuzzy inference

The Mamdani fuzzy inference is the most commonly used fuzzy modeling method, in which the fuzzy implication is carried out by the minimum operator and the rule aggregation is implemented by the maximum operator.

The Mamdani rules can be represented as follows:

$$\text{R1: if } x_1 \text{ is } \tilde{A}_1 \text{ and } x_2 \text{ is } \tilde{B}_1 \text{, then } y \text{ is } \tilde{C}_1$$

$$\text{R2: if } x_1 \text{ is } \tilde{A}_2 \text{ and } x_2 \text{ is } \tilde{B}_2, \text{ then } y \text{ is } \tilde{C}_2$$

Where R1 and R2 are different rules, x_1 and x_2 are inputs, y is the output, $\tilde{A}_1$, $\tilde{A}_2$, $\tilde{B}_1$, $\tilde{B}_2$, $\tilde{C}_1$,

$\widetilde{C}_2$ is fuzzy subset of the domain of input and output variables, respectively. The fuzzy subset F is described by its membership function $\mu_F(U)$, which maps each element in the domain U to a value on [0,1]. The membership function can be in different forms such as trigonometric function, sigmoid function, and Gaussian function.

Assume that there is a precondition: if x_1 is $\widetilde{A}^i$ and x_2 is $\widetilde{B}^i$, then we can get the new fuzzy subset of the domain y $\widetilde{C}^i$ according to (8). The reasoning process of Mamdani fuzzy inference is shown in Figure 8.

$$\mu_{\widetilde{C}}(z)=\bigvee_{i=1}^{n}\left[\omega_i\wedge\left[\mu_{\widetilde{A}}(x)\wedge\mu_{\widetilde{B}}(y)\right]\wedge\left[\mu_{\widetilde{R}_{M1}}(x,y,z)\vee\mu_{\widetilde{R}_{M2}}(x,y,z)\right]\right]=\bigvee_{j=1}^{n}\omega_j \tag{8}$$

where

$$\mu_{\widetilde{R}_{M1}}(x,y,z)=\mu_{\widetilde{A}_1}(x)\wedge\mu_{\widetilde{B}_1}(x)\vee\mu_{\widetilde{C}_1}(z)$$

$$\mu_{\widetilde{R}_{M2}}(x,y,z)=\mu_{\widetilde{A}_2}(x)\wedge\mu_{\widetilde{B}_2}(x)\vee\mu_{\widetilde{C}_2}(z)$$

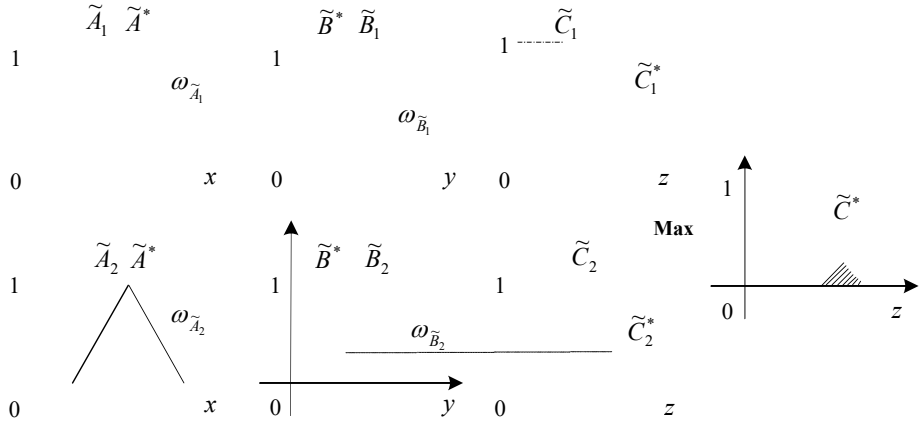


Figure 8 Example of the Mamdani fuzzy reasoning process

2) Model structure and parameters optimized by MOPSO

It can be seen from the foregoing that the number of rules obtained by the fuzzy inference system is positively correlated with the dimensionality of the input variables and the number of fuzzy sets. The increase in the number of rules can improve the accuracy of model prediction. However, excessive amounts of rules and the number of antecedents in each rule will reduce the interpretability for human beings and increase the inconvenience for system control. In addition,

the exponentially increasing model complexity requires extremely high computing resources and cost, which also limits the applicability of the model in industrial sites. Therefore, how to balance the accuracy and complexity of the model is a difficult problem that the fuzzy inference system needs to solved [26]. We use the MOPSO algorithm to simultaneously optimize the structure and parameters of the Mamdani fuzzy system, of which the former includes the number of rules and the antecedents of the rules, while the latter is mainly the mean and variance of the Gaussian membership function.

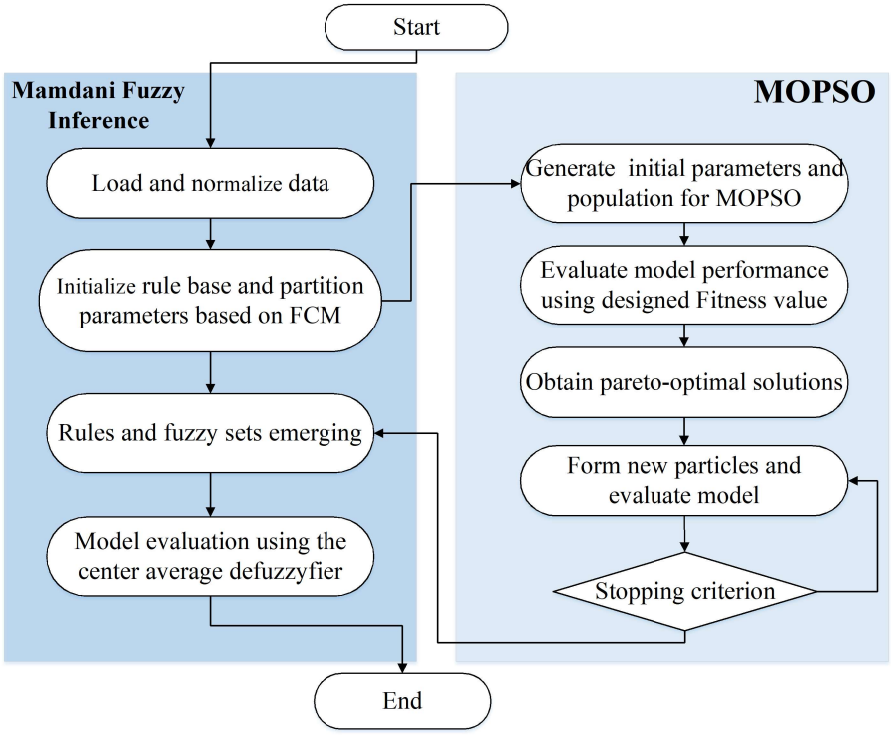


Figure 9 Flowchart of the optimized Mamdani fuzzy model

Figure 9 shows the optimization process of Mamdani fuzzy inference system using the MOPSO algorithm. The position of each particle in MOPSO contains structure and parameter information of the fuzzy model. Each particle will determine the direction and adjust the corresponding speed and position according to the fitness value until the predetermined accuracy is reached. In the multi-objective algorithm, the particles will update iteratively based on the dominance of different fitness values, resulting in a set of Pareto solutions approaching the optimal frontier. The objective function of this subject includes the following:

$$obj1 = RMSE(Y,n) = \sqrt{\frac{\sum_{i=1}^{n}(Y(i)-y(i))^2}{n}} \tag{9}$$

$$obj2 = \frac{Nrule'}{Nrule} + \frac{Nset'}{Nset} \tag{10}$$

where obj1 is the root mean square error of the fuzzy model, $Nrule'$ and $Nset'$ in obj2 represent the number of fuzzy rules and the number of fuzzy sets, respectively.

Fuzzy rules and fuzzy sets merging are also key steps in the model optimization. It is necessary to

merge similar fuzzy rules and delete contradictory fuzzy rules, in which the similarity of two fuzzy rules is defined as the multiplicative of the similarity of the antecedents.

$$SIM(R_1, R_2) = \prod_{i=1}^{m} SIM(A_{1i}, A_{2i}) = \prod_{i=1}^{m} \frac{1}{1+\sqrt{(c_{1i}-c_{2i})^2+(d_{1i}-d_{2i})^2}} \quad (11)$$

where $SIM(R_1, R_2)$ is the similarity between rule R1 and R2, $SIM(A_{1i}, A_{2i})$ is the similarity between fuzzy set A_{1i} and A_{2i}, m is the antecedent number of the fuzzy rule, c_{1i}, c_{2i}, d_{1i}, d_{2i} is the mean value and variance of the two fuzzy subsets, respectively.

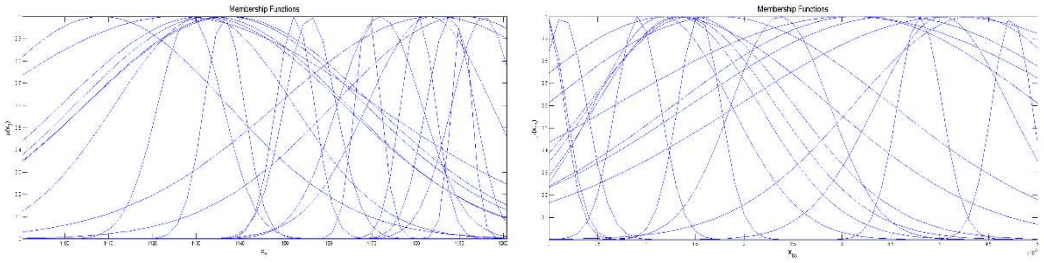
4.3 Experimental results

In order to verify the effectiveness of the proposed approach, the data of 3,500 coils of steel collected by the collaborative design platform built in a large hot rolling mill is used for training and testing. As mentioned above, each roll of steel has 45-dimensional variables, of which 42-dimensional is input, including 15-dimensional chemical compositions, 14-dimensional mechanism model parameters, and 13-dimensional process parameters. First, we use the *K*-means clustering algorithm described in Section 4.1 to merge steel types. Among them, 1,127 coils belong to class A, 735 coils belong to class B, 863 coils belongs to class C, 420 coils belongs to class D, and 355 coils belong to class E. 75% of the data is used for training, 10% of the data is used for validation and the remaining 15% is used for final testing. To verify the improvement effect of the optimization algorithm on fuzzy system, the original Mamdani-type fuzzy modeling method (FM) and the optimized fuzzy system (FM-MOPSO) are both implemented for the performance prediction of YS, TS and ER. To compare the prediction accuracy improvement of the steel type merging approach, the mechanical performance prediction model is also established in each cluster of the steel group. Since the collaborative design platform can supplement the data model with the help of mechanism and simulation models, we also establish a 28-dimensional fuzzy prediction model in each cluster without the mechanism model. The parameter configuration of MOPSO is illustrated in Table 3, which was inspired from suggestions included in [26]. The experiments were carried over for 30 runs. One set of results out of the 30 runs is randomly selected and shown in the following figures.

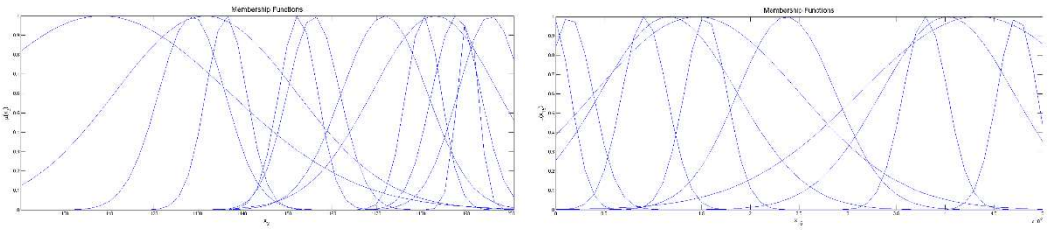
Table 3 Parameters configuration of MOPSO

Parameter	Value
Number of particles	Nump=30
Acceleration factor	c_1=1.5,c_2=1.5
Scaling factor	k_1=0.5,k_2=1.5
Velocity boundary	Vmax=0.5*L
Error limitation	Error=1e-4

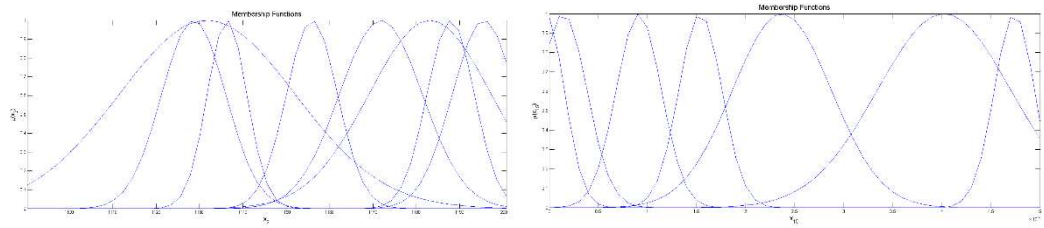
Figure 10 shows the distribution of the membership functions of the two inputs (x_1 and x_{15}) of the initial model and the two optimized models. It is obvious that the initial model has 20 fuzzy sets, some of which has a high degree of similarity. The mean and variance of the optimized membership function are selected from all Pareto solutions with 12 rules and 8 rules, respectively. It can be seen that the membership function of x_1 is mainly concentrated in the two intervals [120 130] and [180 190], while x_{15} is [1 1.5] and [3.5 4]. For different optimized models, more rules and fuzzy sets will bring higher accuracy, while models with fewer rules and parameters have a simpler structure and better interpretability.



(a) the initial model



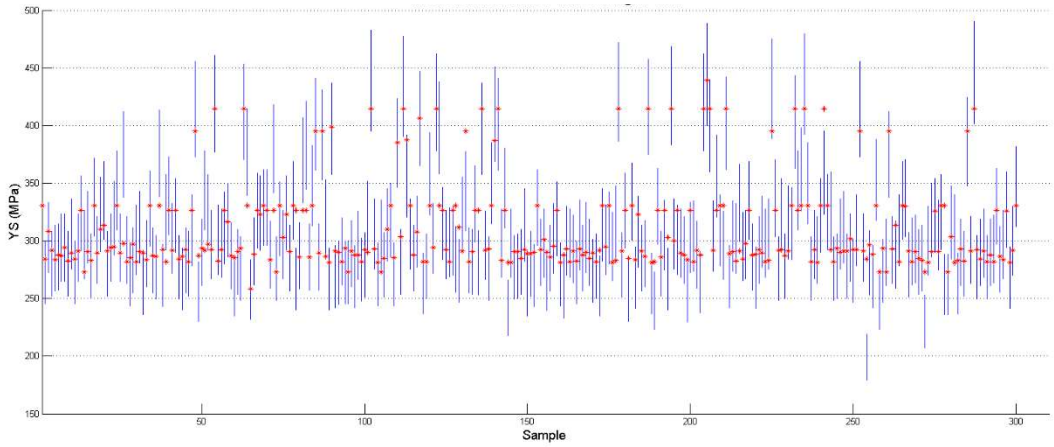
(b) an optimized model with 12 rules



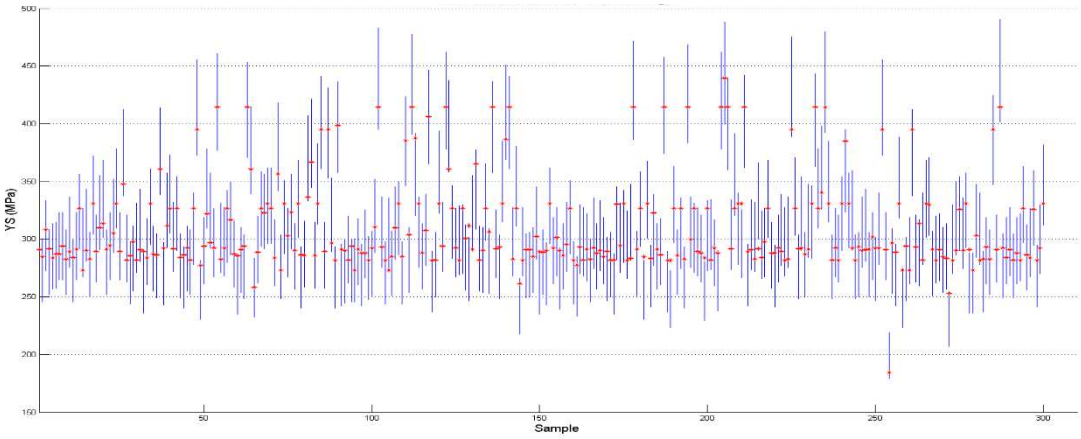
(c) an optimized model with 8 rules

Figure 10 The fuzzy models' membership functions

Since customers' requirements for mechanical properties of steel products are usually put forward in the form of intervals, we take prediction interval coverage probability (PICP) as one of the performance evaluation indicators for prediction algorithm, which is the probability that the targets lie within the constructed intervals. We calculated the PICP of the 10% and 5% deviation intervals respectively. Figure 11 illustrates the 10% deviation intervals of testing samples' actual value and the predicted values of the FM-MOPSO and the FM algorithm. It can be seen that most of the predicted values of the two algorithms lie the 10% deviation intervals, and the FM-MOPSO algorithm has obvious advantages over the FM algorithm.



(a) the FM model



(b) the FM-MOPSO model

Figure 11 The predicted values and the 10% deviation intervals

Table 4 lists the evaluation index results of the four types of algorithms to predict the three mechanical properties. It can be seen that the fuzzy model after steel classification and parameter optimization has the best prediction performance. The root mean square error (RMSE) of YS is reduced by 10.2% on average compared with the FM-MOPSO algorithm without steel type merging, and reduced by 30.6% compared with the FM algorithm, and reduced by 24.9% compared with the FM-MOPSO algorithm without mechanism model data. Among different steel grade types, the prediction effect of class C is the best, with the 10% PICP of YS reaches 98%, and the 10% PICP of TS reaches 99%. This is mainly attributed to the fact that class C is mainly composed of Q235B steel grade, which is the most stable product produced by steel mills and has a large number of sample data. On the contrary, due to the small number of samples, the prediction effect of class D and class E are worse than other steel grades, with most of the 5% PICP less than 70%. At the same time, we can see that the mechanism model data has improved the algorithm more than the steel grade merging, which can illustrate the effectiveness of the collaborative design platform. In terms of different mechanical properties, TS has the best predictive effect, YS the second, and ER the worst.

Table 4 Performance evaluation indicators for different prediction algorithms

		YS			TS			ER		
		RMSE	10%	5%	RMSE	10%	5%	RMSE	10%	5%
FM-MOPSO	Cluster A	18.89	0.93	0.72	19.23	0.94	0.83	2.80	0.90	0.78
	Cluster B	17.65	0.96	0.74	19.11	0.96	0.82	2.71	0.91	0.79
	Cluster C	15.90	0.98	0.78	17.62	0.99	0.86	2.46	0.93	0.76
	Cluster D	21.04	0.89	0.63	22.87	0.91	0.69	3.13	0.85	0.64
	Cluster E	20.50	0.90	0.71	21.14	0.93	0.74	2.99	0.88	0.67
FM-MOPSO without mechanism data	Cluster A	25.36	0.83	0.65	24.09	0.89	0.68	3.49	0.79	0.59
	Cluster B	23.64	0.84	0.67	23.76	0.90	0.69	3.55	0.76	0.58
	Cluster C	18.91	0.94	0.75	19.93	0.97	0.82	2.75	0.91	0.73
	Cluster D	28.24	0.80	0.60	28.79	0.86	0.68	3.31	0.81	0.60
	Cluster E	29.05	0.77	0.61	30.12	0.85	0.69	4.09	0.81	0.60
FM-MOPSO without clustering		20.92	0.92	0.71	21.45	0.94	0.75	2.98	0.88	0.69
FM		27.08	0.81	0.63	26.67	0.88	0.70	3.45	0.83	0.61

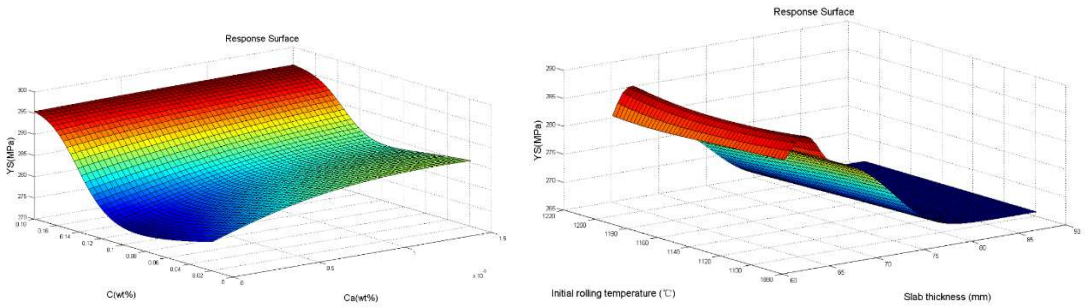


Figure 12 Response surfaces of the optimized 12-rule YS prediction model

In order to verify the effectiveness of the developed model from the perspective of physical interpretation, Figure 12 shows the three-dimensional response surface (RS) of the YS prediction model under optimized 12-rule model. The RS is drawn where the output variable changes to some input variables while keeping the others unchanged. The left figure is the RS of YS with the chemical content C and Ca, while the right is the RS between YS and the thickness of the slab and

the initial heating temperature. The figures are consistent with those variable effect graphs in, which have been confirmed to follow the expected behavior predicted by theory or expert knowledge.

5. Conclusion and future work

In this study, a full-process collaborative design platform is established to accelerate the development of new alloy materials. Based on the mechanism models, simulation models and data-driven models in the platform, a systematic data-driven Mamdani-type fuzzy modelling methodology is proposed to map the relationship between material chemical compositions, organizational structures, process parameters and material mechanical performances. Forty-two parameters are selected as input from more than 500 dimensional variables using a random forest (RF) algorithm, while the YS, TS and ER are chosen as the output parameters. The *K*-means algorithm is utilized to merge 35 different steel grades into 5 groups. Taking both the RMSE and the number of fuzzy rules and fuzzy sets as objective values, the MOPSO algorithm is developed to further improve the fuzzy model in terms of both the structure and the membership function parameters. The effectiveness of the proposed approach was verified on the data of 3,500 coils of steel collected from a large hot rolling mill. Results show that the proposed 12-rules FM-MOPSO with mechanism data and steel grades merging performs the best in the coverage probability and RMSE. Among the results, the prediction effect of Q235B is the best, with the 10% PICP of YS reaches 98%, and the 10% PICP of TS reaches 99%.

Since this study focuses on the prediction of the mechanical properties of steel products and the analysis of influencing factors, the reverse design and optimization of the components and parameters are still lacking. In future research, we will conduct research on process design and adjustment based on the digital twin platform, combining experimental results and user feedback information. In addition, we will also combine the quality control system to further analyze the factors that affect product quality from different dimensions, such as surface defects, size, and shape.

Acknowledgement

This research is supported by the National Natural Science Foundation of China under the grant number 61903031, National Key Research and Development Program of China under grant number 2020YFB1713600, and the Fundamental Research Funds for the Central Universities under the grant number FRF-TP-18-035A1 and FRF-TP-20-050A2.

Reference

[1] S. Lu, Q. Zhou, Y. Ouyang, Y. Guo, Q. Li, J. Wang, Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning, Nat. Commun. 9 (2018) 1–8. https://doi.org/10.1038/s41467-018-05761-w.

[2] Y. Liu, T. Zhao, W. Ju, S. Shi, S. Shi, S. Shi, Materials discovery and design using machine learning, J. Mater. 3 (2017) 159–177. https://doi.org/10.1016/j.jmat.2017.08.002.

[3] G. Peng, H. Wang, H. Zhang, K. Huang, A hypernetwork-based approach to collaborative retrieval and reasoning of engineering design knowledge, Adv. Eng. Informatics. 42 (2019) 100956. https://doi.org/10.1016/j.aei.2019.100956.

[4] G. Peng, H. Wang, H. Zhang, Y. Zhao, A.L. Johnson, A collaborative system for capturing and reusing in-context design knowledge with an integrated representation model, Adv. Eng. Informatics. 33 (2017) 314–329. https://doi.org/10.1016/j.aei.2016.12.007.

[5] J.E. Gubernatis, T. Lookman, Machine learning in materials design and discovery: Examples from the present and suggestions for the future, Phys. Rev. Mater. 2 (2018) 1–15. https://doi.org/10.1103/PhysRevMaterials.2.120301.

[6] M. Picklum, M. Beetz, MATCALO: Knowledge-enabled machine learning in materials science, Comput. Mater. Sci. 163 (2019) 50–62. https://doi.org/10.1016/j.commatsci.2019.03.005.

[7] The White House, About the Materials Genome Initiative. https://obamawhitehouse.archives.gov/mgi, 2011.

[8] Z.H. Shen, J.J. Wang, J.Y. Jiang, S.X. Huang, Y.H. Lin, C.W. Nan, L.Q. Chen, Y. Shen, Phase-field modeling and machine learning of electric-thermal-mechanical breakdown of polymer-based dielectrics, Nat. Commun. 10 (2019) 1–10. https://doi.org/10.1038/s41467-019-09874-8.

[9] P. Wan, H. Zou, K. Wang, Z. Zhao, Research on hot deformation behavior of Zr-4 alloy based on PSO-BP artificial neural network, J. Alloys Compd. 826 (2020) 154047. https://doi.org/10.1016/j.jallcom.2020.154047.

[10] D.G.S. Pivoto, L.F.F. de Almeida, R. da Rosa Righi, J.J.P.C. Rodrigues, A.B. Lugli, A.M. Alberti, Cyber-physical systems architectures for industrial internet of things applications in Industry 4.0: A literature review, J. Manuf. Syst. 58 (2021) 176–192. https://doi.org/10.1016/j.jmsy.2020.11.017.

[11] S. Aheleroff, X. Xu, R.Y. Zhong, Y. Lu, Digital Twin as a Service (DTaaS) in Industry 4.0: An Architecture Reference Model, Adv. Eng. Informatics. 47 (2021). https://doi.org/10.1016/j.aei.2020.101225.

[12] T. Thankachan, K. Soorya Prakash, M. Kamarthin, Optimizing the Tribological Behavior of Hybrid Copper Surface Composites Using Statistical and Machine Learning Techniques, J. Tribol. 140 (2018). https://doi.org/10.1115/1.4038688.

[13] Q. Xie, M. Suvarna, J. Li, X. Zhu, J. Cai, X. Wang, Online prediction of mechanical properties of hot rolled steel plate using machine learning, Mater. Des. 197 (2021) 109201. https://doi.org/10.1016/j.matdes.2020.109201.

[14] W. Li, X. Zhang, H. Kou, R. Wang, D. Fang, Theoretical prediction of temperature dependent yield strength for metallic materials, Int. J. Mech. Sci. 105 (2016) 273–278. https://doi.org/10.1016/j.ijmecsci.2015.11.017.

[15] M. Bambach, S. Seuren, On instabilities of force and grain size predictions in the simulation of multi-pass hot rolling processes, J. Mater. Process. Technol. 216 (2015) 95–113. https://doi.org/10.1016/j.jmatprotec.2014.07.018.

[16] C. Shao, W. Hui, Y. Zhang, X. Zhao, Y. Weng, Microstructure and mechanical properties of hot-rolled medium-Mn steel containing 3% aluminum, Mater. Sci. Eng. A. 682 (2017) 45–53. https://doi.org/10.1016/j.msea.2016.11.036.

[17] I. Mohanty, S. Sarkar, B. Jha, S. Das, R. Kumar, Online mechanical property prediction

system for hot rolled if steel, Ironmak. Steelmak. 41 (2014) 618–627. https://doi.org/10.1179/1743281214Y.0000000178.

[18] Z.W. Xu, X.M. Liu, K. Zhang, Mechanical Properties Prediction for Hot Rolled Alloy Steel Using Convolutional Neural Network, IEEE Access. 7 (2019) 47068–47078. https://doi.org/10.1109/ACCESS.2019.2909586.

[19] Z.H. Wang, D.Y. Gong, X. Li, G.T. Li, D.H. Zhang, Prediction of bending force in the hot strip rolling process using artificial neural network and genetic algorithm (ANN-GA), Int. J. Adv. Manuf. Technol. 93 (2017) 3325–3338. https://doi.org/10.1007/s00170-017-0711-5.

[20] A.H. Orta, I. Kayabasi, M.G. Senol, Prediction of mechanical properties of cold rolled and continuous annealed steel grades via analytical model integrated neural networks, Ironmak. Steelmak. 47 (2020) 596–605. https://doi.org/10.1080/03019233.2019.1568000.

[21] M. Asadi, Optimized Mamdani fuzzy models for predicting the strength of intact rocks and anisotropic rock masses, J. Rock Mech. Geotech. Eng. 8 (2016) 218–224. https://doi.org/10.1016/j.jrmge.2015.11.005.

[22] A. Berber, K. Dincer, Y. Yilmaz, D.N. Ozen, Rule-based Mamdani-type fuzzy modeling of heating and cooling performances of counter-flow Ranque-Hilsch vortex tubes with different geometric construction for steel, Energy. 51 (2013) 297–304. https://doi.org/10.1016/j.energy.2013.01.005.

[23] M.H. Eghbal Ahmadi, S.J. Royaee, S. Tayyebi, R. Bozorgmehry Boozarjomehry, A new insight into implementing Mamdani fuzzy inference system for dynamic process modeling: Application on flash separator fuzzy dynamic modeling, Eng. Appl. Artif. Intell. 90 (2020) 103485. https://doi.org/10.1016/j.engappai.2020.103485.

[24] M. Tosun, K. Dincer, S. Baskaya, Rule-based Mamdani-type fuzzy modelling of thermal performance of multi-layer precast concrete panels used in residential buildings in Turkey, Expert Syst. Appl. 38 (2011) 5553–5560. https://doi.org/10.1016/j.eswa.2010.10.081.

[25] Z. Hong, V. Viswanathan, Open-Sourcing Phase-Field Simulations for Accelerating Energy Materials Design and Optimization, ACS Energy Lett. 5 (2020) 3254–3259. https://doi.org/10.1021/acsenergylett.0c01904.

[26] Q. Zhang, M. Mahfouf, A hierarchical Mamdani-type fuzzy modelling approach with new training data selection and multi-objective optimisation mechanisms: A special application for the prediction of mechanical properties of alloy steels, Appl. Soft Comput. J. 11 (2011) 2419–2443. https://doi.org/10.1016/j.asoc.2010.09.004.

[27] M. Antonelli, P. Ducange, B. Lazzerini, F. Marcelloni, Learning concurrently partition granularities and rule bases of Mamdani fuzzy systems in a multi-objective evolutionary framework, Int. J. Approx. Reason. 50 (2009) 1066–1080. https://doi.org/10.1016/j.ijar.2009.04.004.

[28] M.A. Kacimi, O. Guenounou, L. Brikh, F. Yahiaoui, N. Hadid, New mixed-coding PSO algorithm for a self-adaptive and automatic learning of Mamdani fuzzy rules, Eng. Appl. Artif. Intell. 89 (2020) 103417. https://doi.org/10.1016/j.engappai.2019.103417.

[29] J.E. Moreno, M.A. Sanchez, O. Mendoza, A. Rodríguez-Díaz, O. Castillo, P. Melin, J.R. Castro, Design of an interval Type-2 fuzzy model with justifiable uncertainty, Inf. Sci. (Ny). 513 (2020) 206–221. https://doi.org/10.1016/j.ins.2019.10.042.