Available online at www.sciencedirect.com

ScienceDirect

Journal of Magnesium and Alloys 12 (2024) 750–766

www.elsevier.com/locate/jma

Full Length Article

Accelerated design of high-performance Mg-Mn-based magnesium alloys based on novel bayesian optimization

Xiaoxi Mi[a,b], Lili Dai[a], Xuerui Jing[a], Jia She[a,c], Bjørn Holmedal[d], Aitao Tang[a,c,*], Fusheng Pan[a,c]

[a] College of Materials Science and Engineering, Chongqing University, Chongqing 400044, China
[b] Southwest Technology and Engineering Research Institute, Chongqing 400039, China
[c] National Engineering Research Center for Magnesium Alloys, Chongqing University, Chongqing 400044, China
[d] Department of Materials Science and Engineering, Norwegian University of Science and Technology (NTNU), Trondheim 7051, Norway

Received 23 September 2023; received in revised form 27 November 2023; accepted 2 January 2024
Available online 18 January 2024

Abstract

Magnesium (Mg), being the lightest structural metal, holds immense potential for widespread applications in various fields. The development of high-performance and cost-effective Mg alloys is crucial to further advancing their commercial utilization. With the rapid advancement of machine learning (ML) technology in recent years, the "data-driven'' approach for alloy design has provided new perspectives and opportunities for enhancing the performance of Mg alloys. This paper introduces a novel regression-based Bayesian optimization active learning model (RBOALM) for the development of high-performance Mg-Mn-based wrought alloys. RBOALM employs active learning to automatically explore optimal alloy compositions and process parameters within predefined ranges, facilitating the discovery of superior alloy combinations. This model further integrates pre-established regression models as surrogate functions in Bayesian optimization, significantly enhancing the precision of the design process. Leveraging RBOALM, several new high-performance alloys have been successfully designed and prepared. Notably, after mechanical property testing of the designed alloys, the Mg-2.1Zn-2.0Mn-0.5Sn-0.1Ca alloy demonstrates exceptional mechanical properties, including an ultimate tensile strength of 406 MPa, a yield strength of 287 MPa, and a 23% fracture elongation. Furthermore, the Mg-2.7Mn-0.5Al-0.1Ca alloy exhibits an ultimate tensile strength of 211 MPa, coupled with a remarkable 41% fracture elongation.

Keywords: Mg-Mn-based alloys; High-performance; Alloy design; Machine learning; Bayesian optimization.

1. Introduction

Magnesium (Mg) and its alloys are one of the lightest structural metals, with advantages such as low density, high specific strength and stiffness, good thermal conductivity, damping properties, and environmental friendliness [1,2]. They have enormous potential applications in 3C, automotive, aerospace, biomedical, military, and sports equipment, and have become one of the most popular materials [3,4]. Compared to cast Mg alloys, wrought Mg alloys offer higher strength, improved ductility, and a broader range of mechanical properties, thereby meeting the diverse application requirements of engineering structural components. The development of novel wrought Mg alloys with simple processing techniques, abundant alloying elements, and low cost is crucial to expand the applications of Mg [5].

In recent years, Mg-Mn non-rare earth wrought alloys have attracted mounting focus due to their low elemental costs, simple fabrication processes, and excellent mechanical properties, making them one of the most promising new commercial Mg alloys [6–8]. The main alloying elements in these series

* Corresponding author.
E-mail address: tat@cqu.edu.cn (A. Tang).

https://doi.org/10.1016/j.jma.2024.01.005

alloys include Mn, Al, Zn, Sn, and Ca, offering a more cost-effective alternative compared to rare earth systems. Moreover, the inclusion of Mn in alloys enables the production of ultrafine grain structures through simple extrusion techniques, resulting in improved mechanical properties, particularly in terms of overall performance [9–12]. Despite the substantial development potential and value inherent in non-rare-earth wrought alloys based on Mg-Mn system, there remain hurdles in the formulation and advancement of these innovative alloys. This is primarily due to the need for additional alloying elements to synergistically interact within such alloy systems. The more alloying elements involved, the greater the research difficulty and cost. Furthermore, apart from considering the composition, deformation processing must also be taken into account. Only when the composition and processing are suitable can superior ultrafine-grained Mg alloys be obtained [11]. This undoubtedly poses difficulties when using traditional experimental approaches for research and development. At present, the research and development of new multi-element Mg-Mn series wrought alloys still rely on a trial-and-error method, which not only costly, time-consuming and labor-intensive, but also has unpredictable results, which makes it cannot meet the growing demands of alloy design. Therefore, it is crucial to explore an accurate and efficient means of alloy design to facilitate the design and development of new Mg-Mn-based non-rare earth wrought Mg, which is essential for expanding the applications of Mg alloys.

Over the past decade, machine learning (ML) technology has rapidly advanced and achieved numerous successful applications in materials research, demonstrating tremendous potential [13–15]. Integrating machine learning techniques into the design and development of Mg-Mn-based non-rare earth wrought alloys, leveraging their powerful data processing capabilities to address the multi-element and multi-parameter alloy design challenges, and achieving end-to-end design of new high-performance and low-cost Mg alloys, may offer an effective approach to solving the development of Mg-Mn alloy systems and represents a meaningful exploration for further expanding the commercial applications of magnesium alloys. This also makes it possible to transform material design from the traditional "trial and error" approach to a “data-driven’’ approach [16,17].

Currently, there has been an increasing amount of research focused on utilizing ML for developing innovative materials [18–20]. Among them, Bayesian optimization (BO) has gained significant popularity due to its simplicity, efficiency, accuracy, and alignment with practical design principles. BO algorithms were initially and most commonly used for model hyperparameter optimization but have gradually evolved into a method for active learning, also known as adaptive design [21]. The utilization of BO for the discovery and optimization of novel materials has become a trending topic in the field of materials science. The effectiveness of BO in dealing with extremum problems makes them particularly useful for searching for optimal material properties [22]. For instance, Bowen et al. [23] constructed a BO framework using various surrogate models, which can automatically search for the optimal composition of materials. An BO active learning model was developed by Liu et al. to find the ultra-hard magnesium alloys [24]. Another example comes from Chen et al., who created prediction models for B-C-N super-hard compounds, with their methodology centered around BO [25]. Furthermore, Joshua et al. [26] proposed an BO model for exploring mechanical properties, which can predict properties beyond the original training data of materials and has been validated using a materials database containing 85,707 crystal structures.

In this study, an improved machine learning model, RBOALM, was established for designing high-performance Mg-Mn-based wrought alloys. We innovatively replaced the Gaussian Process (GP) with a well-established regression model as the surrogate function in RBOALM, thereby enhancing the data fitting capability and improving the effectiveness of BO. The extremum solutions of yield strength, ultimate strength, and fracture elongation were computed using RBOALM. Subsequently, several high-performance Mg alloys were designed based on the active learning results, and experimental validations were conducted to evaluate the precision of the RBOALM.

2. Methods

2.1. Design strategy

In this work, an active learning design strategy was proposed to optimize the strength and plasticity of Mg-Mn-based wrought alloys by coupling multiple regression algorithms and BO. The workflow of design strategy is shown in Fig. 1, and mainly steps are as follows:

(1) Dataset and feature. Collected data on Mg-Mn-based extruded alloys from published literatures and previous experimental results to generate a dataset. And determined the input features and output targets of the model by feature analysis and feature screening.
(2) Active learning. Developed the RBOALM using multiple regression algorithms and BO, relating the composition and process parameters to mechanical properties of alloys. In this step, BO algorithm is employed firstly to optimize the hyperparameters of the regression models, thus reducing model errors. Subsequently, leveraging the optimal regression model as a surrogate function for BO, the mechanical performance of alloy is further optimized using BO.
(3) Alloys design and experimental validation. Using active learning, the problem of optimizing alloy properties is converted into a problem of finding the extreme value of a function, followed by the design of novel high-strength and high-plasticity alloys. Subsequently, based on the outcomes of the active learning model, the alloys are synthesized and experimentally validated, and the design errors are computed to evaluate the performance of the model.

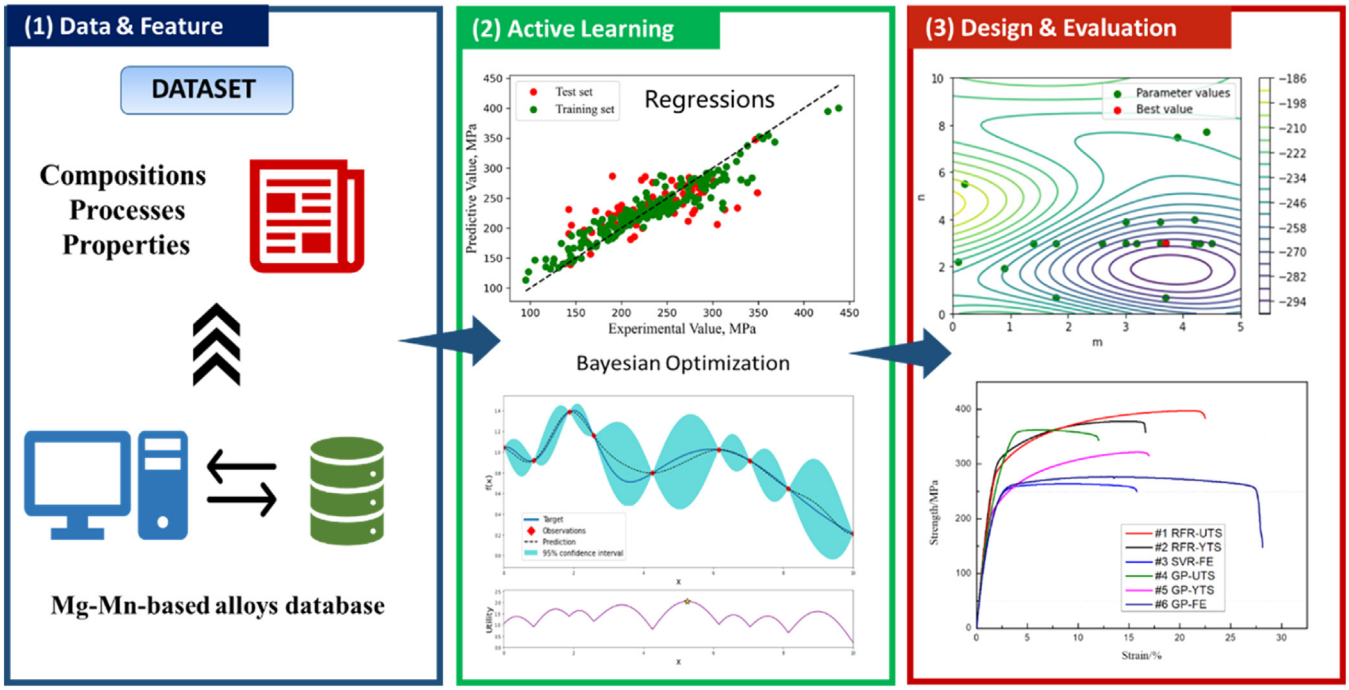

Fig. 1. Schematic diagram of the workflow of the active learning design strategy.

2.2. Features and Dataset

Training and testing data are mainly collected from Refs [2,7,9,10],[27-54] and some previous experiments. All samples exclusively consist of extruded rod materials belonging to the Mg-Mn-based systems. Based on the original data and some research results, 10 items, including alloy composition, extrusion process parameters, and mechanical properties, were selected as the features. The feature selection and establishment of the raw data table primarily adhere to the following principles:

(1) Alloy composition: The alloy composition feature includes five elements, Mn, Al, Zn, Sn, and Ca. The reasons for selecting these five elements are: (i). Al, Zn, Sn, and Ca have been utilized as alloying elements in the development of various high-performance Mg-Mn-based systems, such as AZ, AM, ZM, and AMZ, indicating their potential for further enhancing alloy properties. Extensive experimental studies conducted on these alloy systems provide ample data for machine learning. (ii). Compared to rare earth-containing alloy systems, these systems exhibit alloy properties that are relatively close to rare earth systems while featuring lower element costs, enabling significant reduction in production expenses.

(2) Extrusion process parameter: Extrusion is a crucial process of wrought Mg alloys. During the hot extrusion, dynamic recrystallization and texture formation occurs, which directly impacting the final properties. Additionally, all alloy samples in the dataset are rods, which are directly extruded from ingots after being held at the extrusion temperature. Alloys that underwent preheat-treatment processes (such as homogenization) were not selected in this study. This helps reduce the number of input features and effectively control production costs. Some researches show that extrusion temperature and extrusion speed have significant influence on the quality of Mg alloys [2,31,36,46,47]. Generally, the extrusion ratio (ER) is also an important factor affecting the structure and properties of alloys, but in this work, the distribution of ER is too concentrated to make significant effect on the targets (Among the 310 data, 236 instances had an ER value of 25, and some with 20 or 28). Therefore, in order to improve the learning efficiency and accuracy of model, only extrusion temperature (ET, °C) and extrusion speed (ES, m/min) are selected as the primary extrusion process features.

(3) Mechanical property: Generally, yield tensile strength (YTS, MPa), ultimate tensile strength (UTS, MPa), and fracture elongation (FE, %) are commonly used to reflect the mechanical properties of alloys. Therefore, these three attributes are chosen as the feature parameters for alloy mechanical performance.

Based on the aforementioned principles, the database for RBOALM was established, comprising a total of 310 data sets. Each data set consists of 10 features, as shown in Table 1. In regression models, the inputs are alloy compositions and extrusion process parameters, while the outputs are mechanical properties. In BO process, no inputs need to be set and the model outputs the extremes of performance as well as the corresponding solutions for the composition and extrusion parameters.

Table 1
Definitions and ranges of features in the RBOALM dataset.

Feature		Range
Composition	Mn (wt.%)	0–3.00
	Al (wt.%)	0–9.00
	Zn (wt.%)	0–8.80
	Sn (wt.%)	0–9.70
	Ca (wt.%)	0–3.50
Extrusion parameter	Extrusion Temperature (°C)	175–500
	Extrusion Speed (m/min)	0.2–12.0
Mechanical property	Yield Tensile Strength (MPa)	98–438
	Ultimate Tensile Strength (MPa)	154–457
	Fracture Elongation (%)	2.5–55.0

2.3. Algorithms

Herein, the RBOALM couples four regression and BO algorithms, where the regression algorithms are used to establish the connection between features and property targets, and as well as a surrogate function in RBOALM. The BO algorithm is used to optimize the hyper parameters of regression models and also to search for the best solutions of extreme performance values. All algorithms related in this paper are described below.

2.3.1. Gradient boosting regression

Gradient Boosting Regression (GBR) is a powerful machine learning algorithm that falls under the umbrella of ensemble learning methods. Ensembles combine multiple weak learners, typically decision trees, to create a strong predictive model. GBR is widely used in various domains, including finance, healthcare, and industry, due to its ability to handle complex non-linear relationships and high-dimensional data. Its iterative boosting process allows it to correct errors and improve predictive accuracy. With its ability to handle complex non-linear relationships and missing data, GBR has found widespread use in diverse fields and continues to be a valuable tool for data-driven decision-making [55,56].

2.3.2. Support vector regression

The regression algorithm of Support Vector Machines (SVM) is also known as Support Vector Regression (SVR) and is a supervised machine learning algorithm primarily used for regression problems. Similar to the SVM classification algorithm, the core idea of it is to construct a hyperplane that maximizes the minimum distance between the samples in the training set and the regression line (hyperplane), aiming to achieve regression predictions for new data points. The advantages of SVR lie in its ability to handle high-dimensional data and nonlinear regression problems while avoiding overfitting. However, it is sensitive to parameter selection and requires parameter tuning, and the computational complexity of the algorithm is relatively high. SVR is a powerful regression analysis algorithm widely applied in practical problems [57,58].

2.3.3. Decision tree regression

Decision Tree Regression (DTR) is a variant of the decision tree algorithm and is a regression algorithm based on a tree-like structure. It is used when there is a need to predict a continuous output value based on input features. The DTR algorithm can be employed in such cases. DTR algorithm utilizes a tree-like structure to represent the relationship between input features and outputs, where each non-leaf node represents a feature and each leaf node represents an output value. When constructing a DTR model, the algorithm initially selects a feature and divides the data set into two subsets based on that feature. Then, for each subset, the algorithm recursively repeats this process by selecting the best feature for further division until certain stopping criteria are met. Upon meeting the stopping criteria, the algorithm outputs the average output value of the subset at the leaf node as the prediction result [57,58].

2.3.4. Random forest regression

The Random Forest Regression (RFR) is a machine learning algorithm based on an ensemble of decision trees. Its main idea is to analyze data through the construction of multiple decision trees and combine the results of these trees to improve the accuracy and robustness of regression analysis. The RFR algorithm offers advantages such as high accuracy, robustness, and strong generalization performance. It can be applied to various regression problems, particularly suitable for high-dimensional data and situations with significant noise [57,59].

2.3.5. Bayesian optimization

Bayesian Optimization (BO) is an algorithm that uses Bayesian statistical methods for optimization. It dynamically adjusts the search space based on existing data to find the optimal solution for the target function [60]. Specifically, the BO algorithm constructs a prior distribution for the target function (typically using Gaussian processes or tree structures) and a likelihood function. It then applies Bayesian inference to compute the posterior distribution of the target function, thereby identifying potential optimal points in the search space. Unlike traditional optimization algorithms, BO does not require the use of gradient information of the target function, making it suitable for black-box optimization problems. The advantages of the BO algorithm are its ability to quickly find global optima in high-dimensional, non-convex, and noisy settings [23,61].

2.3.6. Gaussian process

In the BO algorithm, it is necessary to determine the prior distribution and likelihood function of the target function, where the prior distribution is also known as the surrogate model, typically represented using Gaussian process. Gaussian process (GP) is a tool for probabilistic modeling, commonly used to estimate the distribution of unknown function values and provide quantification of uncertainty. In BO, Gaussian processes are employed as surrogate models of the target

function, providing estimates of the target function during the sampling and fitting processes [62,63].

All the regression ML algorithms in this work come from the *scikit-learn* library (http://scikit-learn.sourceforge.net.). And the Bayesian Optimization algorithm was based on *Bayesian-Optimization* module in Python.

3. Modeling

3.1. Feature analysis and screening

In the present work, the correlation coefficients between features and targets were calculated, which are shown in Fig. 2. Fig. 2(a) presents the heatmap of Pearson correlation coefficients, which was used to assess the linear correlation between features and targets. Fig. 2(b) shows the Spearman correlation coefficients, which was used to evaluate the monotonic correlations. From the figures, it can be observed that the absolute values of the majority of Pearson and Spearman coefficients are less than 0.3, indicating weak linear and monotonic correlations between features and targets. Therefore, non-linear and non-monotonic algorithms are more suitable and will be employed for the subsequent modeling tasks in this study.

Moreover, some insights can be gleaned from Fig. 2. For instance, in Fig. 2(a), the correlations of Al, Zn, and Sn with UTS appear notably similar, suggesting a potential positive synergistic interaction among these three elements. This implies that the UTS of an alloy can be boosted by increasing these elements, especially in the form of compounds. Similarly, a negative synergistic interaction is indicated between Mn and Al, implying that the performance of the alloys could be improved by increasing the ratio of Al and Mn (Al/Mn). Furthermore, as observed in Fig. 2(b), features such as Ca and ES exhibit closely aligned Spearman coefficients, indicating a consistent monotonic influence of Ca and ES on alloy performance. In alloy design, simultaneous consideration of these two features may be beneficial.

In this work, a random forest model was employed to calculate the importance of features with the aim of identifying the key features that affect the mechanical properties of the alloys. Additionally, principal component analysis (PCA) was conducted to achieve effective dimensionality reduction of the dataset. By using the selected features, ML models were built, leading to significant improvements in model accuracy and fit. The feature importance was computed using the Mean decrease impurity (MDI) method, which evaluates the extent to which each feature reduces the errors in regression problems. The sum of feature importance is equal to 1, and the value of each feature importance represents the weight of its impact on the output of the regression model [64–66]. After parameter optimization, the random forest model was configured with 'n_estimators=200′ and 'max_depth=10′. The computed results are shown as Fig. 3. As mentioned before,

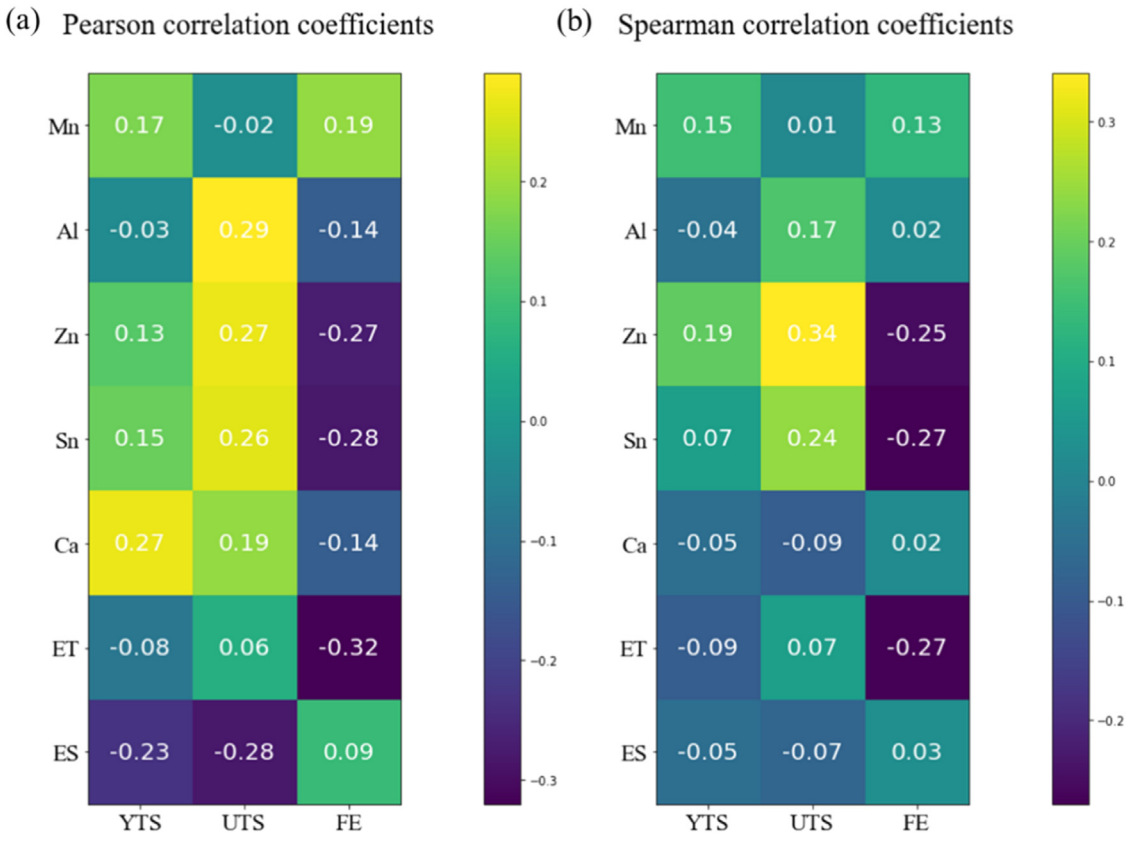

Fig. 2. The correlation coefficients between features and targets. (a) Pearson correlation coefficients. (b) Spearman correlation coefficients.

the extrusion ratio (ER) is also an important parameter that affects the mechanical properties of the alloys. Therefore, in the initial modeling stage of this study, we chose the extrusion ratio as one of the input features and calculated the corresponding feature importance. The computational results illustrated that the feature importance of the extrusion ratio for YTS, UTS and FE were 0.03, 0, and 0.01, which indicated that the extrusion ratio is not a main factor affecting properties in this work, due to its over-concentrated distribution, and thus, in this paper the extrusion ratio feature was eliminated so as to improve the model accuracy and generalization capability.

Fig. 3 shows the calculated results of feature importance, revealing that the main factors affecting YTS, UTS, and FE

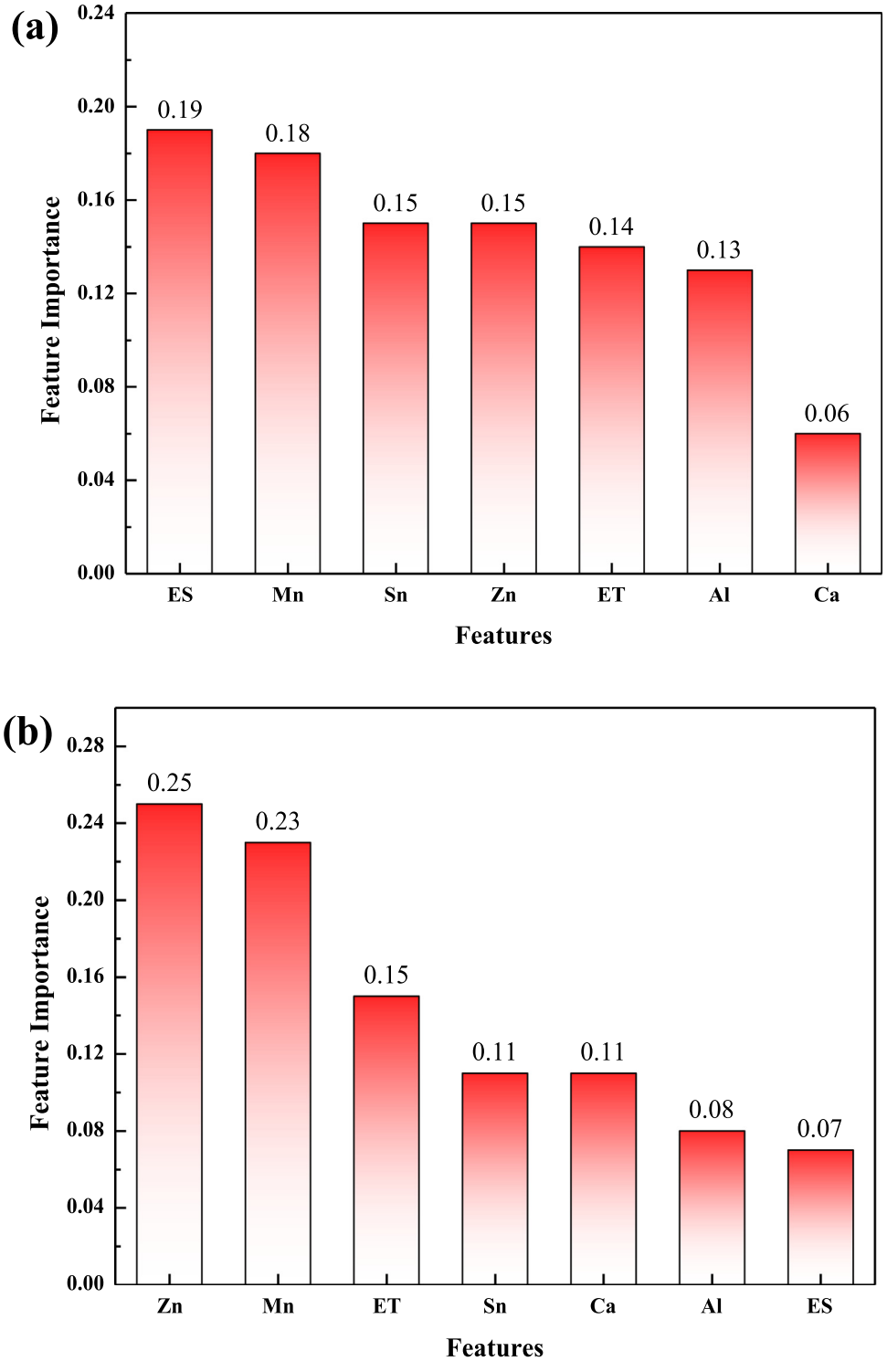

Fig. 3. Feature importance of mechanical properties. (a) Feature importance for YTS. (b) Feature importance for UTS. (c) Feature importance for FE.

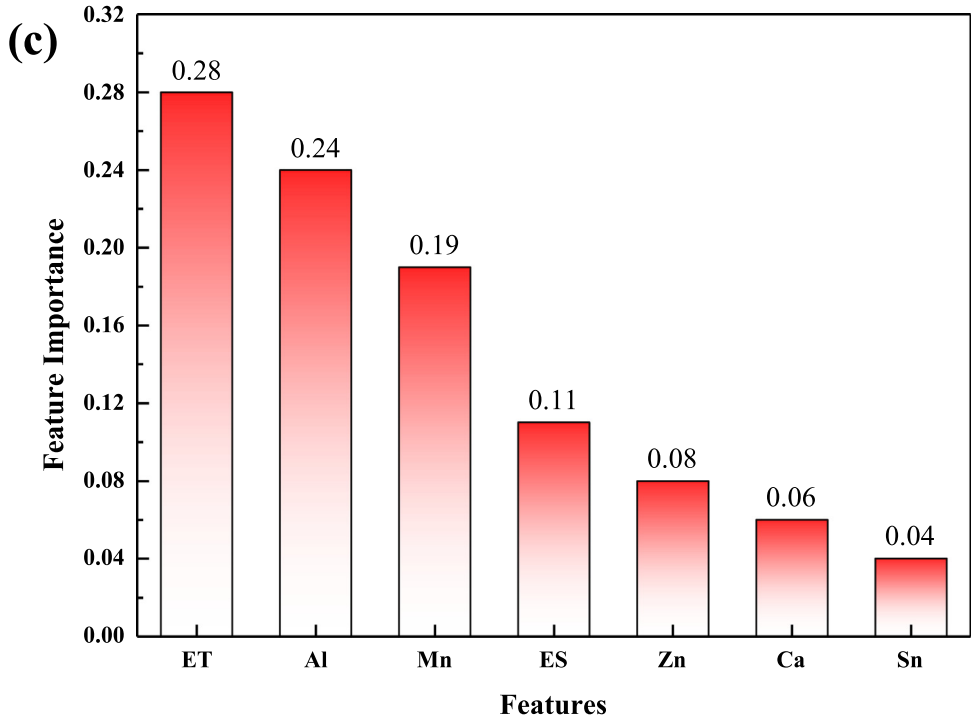

Fig. 3. Continued

are not the same. As shown in Fig. 3(a), in Mg-Mn-Al-Zn-Sn-Ca system alloys, ES is the most significant influencing factor concerning YTS with a feature weight of 0.19. Following that, the Mn, Sn, Zn, ET, Al and Ca contribute with feature weights of 0.18, 0.15, 0.15, 0.14, 0.13 and 0.06, respectively. It suggests that when designing the tensile yield strength of Mg-Mn-Al-Zn-Sn-Ca alloys, priority should be given to extrusion speed and the Mn content in the alloys, followed by considering factors such as Sn, Zn, ET, and Al content. In Fig. 3(b), the weights of Zn and Mn both exceed 0.2, suggesting that the second-phase behavior of Mn and Zn in the system significantly impacts the ultimate tensile strength of the alloys. Furthermore, the feature importance of ES noticeably decreases, indicating that the effect of extrusion speed on ultimate tensile strength is considerably weaker compared to yield strength. Fig. 3(c) reveals that related to FE, the impact of ET is the highest with a weight of 0.28, followed by Al and Mn, both around 0.2. The influence of Sn and ER is negligible and can be disregarded. This implies that the Mg-Mn-Al system holds the potential for developing high ductility wrought Mg but requires optimal matching of extrusion process parameters, particularly extrusion temperature.

3.2. Model building and optimization

In the construction of the RBOALM, the BO algorithm was employed twice for different requirements. The first application was for optimizing the hyperparameters of the regression models, while the second was coupled with the regression models to design high-performance alloys. In a typical BO model, it is common to use the Gaussian process (GP) as a surrogate function. However, in this work, we innovatively used a well-established regression model as the surrogate function to search for extrema in BO. The main reason is a well-established regression model can describe the mapping relationship of inputs and outputs more accurate than GP, which can make the BO process more efficient.

Firstly, three regression models were constructed, with each model designed to specifically describe the relationship between the components, process parameters, and YTS, UTS, FE, respectively. And then BO was employed to optimize the hyperparameters of all regression models for better accuracy. The modeling and optimization process of the three regression models is as follows:

3.2.1. Regression models of YTS

Based on the feature analysis results from the previous section, it has been determined that the principal factors for the YTS of Mg-Mn-based alloys are ES, Mn, Sn, Zn, ET, and Al. As a result, a new YTS dataset has been constructed by extracting these features from the original dataset for each data entry. The input features of the YTS regression model include ES, Mn, Sn, Zn, ET, and Al, while the output feature is YTS. The dataset size and the range of feature parameter values are consistent with those presented in Table 1.

Gradient boosting regression (GBR), support vector regression (SVR), decision tree regression (DTR), and random forest regression (RFR) algorithms were used to establish the YTS prediction models. The performance of these models was evaluated using mean absolute error (MAE) and mean absolute percent error (MAPE) as quantitative metrics after

conducting cross-validation, which are defined as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_i'\right| \tag{1}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - y_i'\right|}{|y_i|} \tag{2}$$

In Eqs. (1) and (2), y_i and y_i' denote the actual value and the predicted value of YTS, n is the size of the training or test set. The training and testing set ratio for all algorithm models was set to 7:3. Subsequently, the BO algorithm was employed with the optimization criterion of minimizing the MAPE to optimize the hyperparameters of the YTS regression model. In the computational results of this study, MAE and MAPE exhibit a clear positive linear correlation. Therefore, it is sufficient to choose one as the hyperparameter BO criterion, and in this paper, we have selected 'Minimize MAPE' as the criterion. The initial observation points for the BO model were set to 10, and the number of iterations was set to 100.

In the GBR model, the key hyperparameters for optimization were "n_estimators'' and "max_depth''. The optimization range for the "n_estimators'' parameter was set to 50-500, while the optimization range for the "max_depth'' parameter was set to 5-20. For the SVR model, the "kernal''='rbf', and the optimization range for both the "C'' and "gamma'' parameters was set to 0.001-1000. The main optimization parameter for the DTR algorithm was the maximum tree depth, with the optimization range for the "max_depth'' parameter set to 5-20. In the case of RFR, the primary optimization parameters were "n_estimators'' and "max_depth'', with the optimization range for "n_estimators'' set to 50-500, and for "max_depth'' set to 5-20. The results before and after BO are presented in Table 2, and the corresponding MAPE results are illustrated in Fig. 4 From the figures, it can be observed that the DTR model had the highest error, followed by GBR and SVR, while the RFR model exhibited relatively smaller errors. The optimized MAPE after BO was found to be 16.6%.

Fig. 5 shows the fitting plots of the four ML models. The points on the dashed line 'y=x' represent the predictions of the models that are identical to the actual values. Therefore, the closer the distribution of points is to the dashed line, the higher the degree of model fitting and the stronger the generalization ability. Among them, the SVR and DTR models exhibit relatively low fitting (R^2) scores, both below 0.50. In contrast, the GBR and RFR models demonstrate higher degrees of fitting. The GBR model achieves a fitting score of 0.75, while the RFR model performs the best with an R^2 score of 0.80. This finding is consistent with the MAPE error calculations. Therefore, considering all factors, the RFR model is selected as the surrogate model (prior distribution function) for subsequent active learning model work, denoted as RFR-YTS.

3.2.2. Regression models of UTS

From Fig. 3(b), it can be observed that for Mg-Mn-based wrought alloys, the main influential factors affecting UTS are Zn, Mn, ET, Sn, Ca, Al, and ES. Therefore, the corresponding features of each data group were extracted from the dataset to reconstruct the UTS dataset. The input features for the UTS regression model are Zn, Mn, ET, Sn, Ca, Al, and ES, while the output feature is UTS. The data size and feature parameter range remain the same as in Table 1. The modeling method and initial modeling parameters for the UTS regression model are identical to those of the YTS regression model, and the BO algorithm is also employed to optimize the hyperparameters. The optimized hyperparameters and errors before and after BO are shown in Table 3.

The optimized MAPE is depicted in Fig. 6. From the graph, it can be observed that the DTR model exhibits the highest error, while the SVR model has the lowest error. The optimized MAPE after BO is 12.5%. Fig. 7 shows the fitting results of the four UTS regression models. The R^2 values of GBR, SVR, DTR and RFR models for the test set are 0.75, 0.77, 0.48, and 0.73, respectively. Additionally, since the SVR model exhibits the minimum MAPE, it is selected as the surrogate function for subsequent active learning model work, denoted as SVR-UTS.

3.2.3. Regression models of FE

From Fig. 3(c), it can be observed that the main features influencing FE are ET, Al, Mn, ES, Zn, and Ca. Therefore, similar to the YTS and UTS models, when constructing the FE model, the input features selected are ET, Al, Mn, ES, Zn, and Ca, while the output feature is FE. The modeling method and initial modeling parameters for the FE model are the same as the YTS and UTS prediction models, and the hyperparameters of the model are optimized using the BO algorithm. The optimized hyperparameters and errors before and after BO are shown in Table 4. From Fig. 8, it can be observed that the RFR model exhibits the lowest MAE and MAPE errors, with optimized values of 3.7 and 32.3%, respectively. It is worth noting that although the MAE values for all models are less than 5, the MAPE values exceed 30%.

Table 2
The hyperparameters and errors of YTS models optimized by BO.

Model	GBR	SVR	DTR	RFR
Hyperparameters before BO	n_estimators=100, max_depth=8	C=100, gamma=1	max_depth=6	n_estimators=100, max_depth=6
Hyperparameters after BO	n_estimators=427, max_depth=11	C=904, gamma=0.24	max_depth=8	n_estimators=206, max_depth=9
MAEs before BO	23.2	25.7	32.9	27.2
MAPEs before BO	24.7%	21.7%	28.5%	18.9%
MAEs after BO	20.9	21.2	26.7	22.5
MAPEs after BO	21.1%	19.6%	24.9%	16.6%

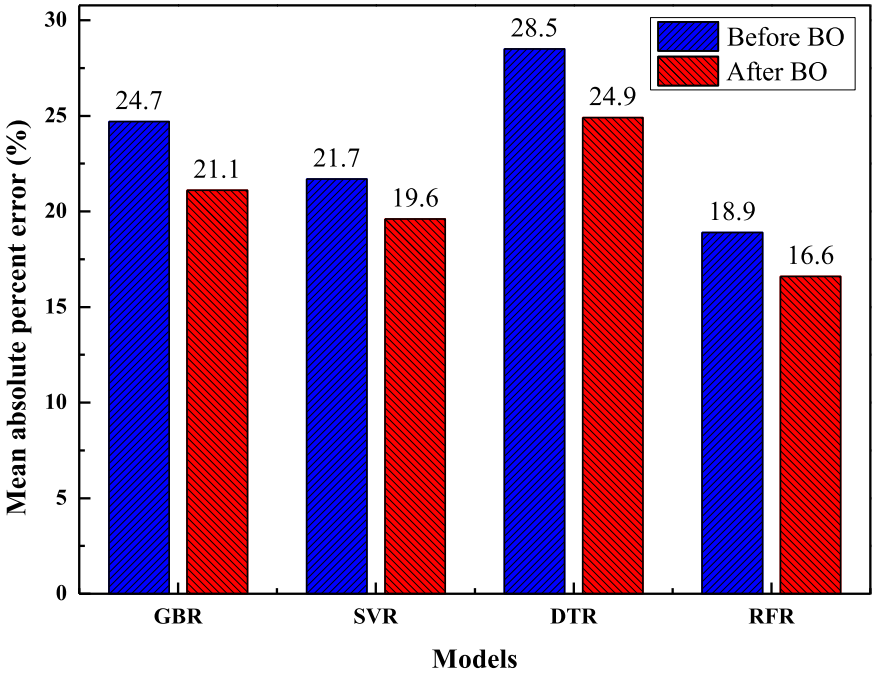

Fig. 4. Comparison of errors (MAPE) before and after using BO in YTS regression models.

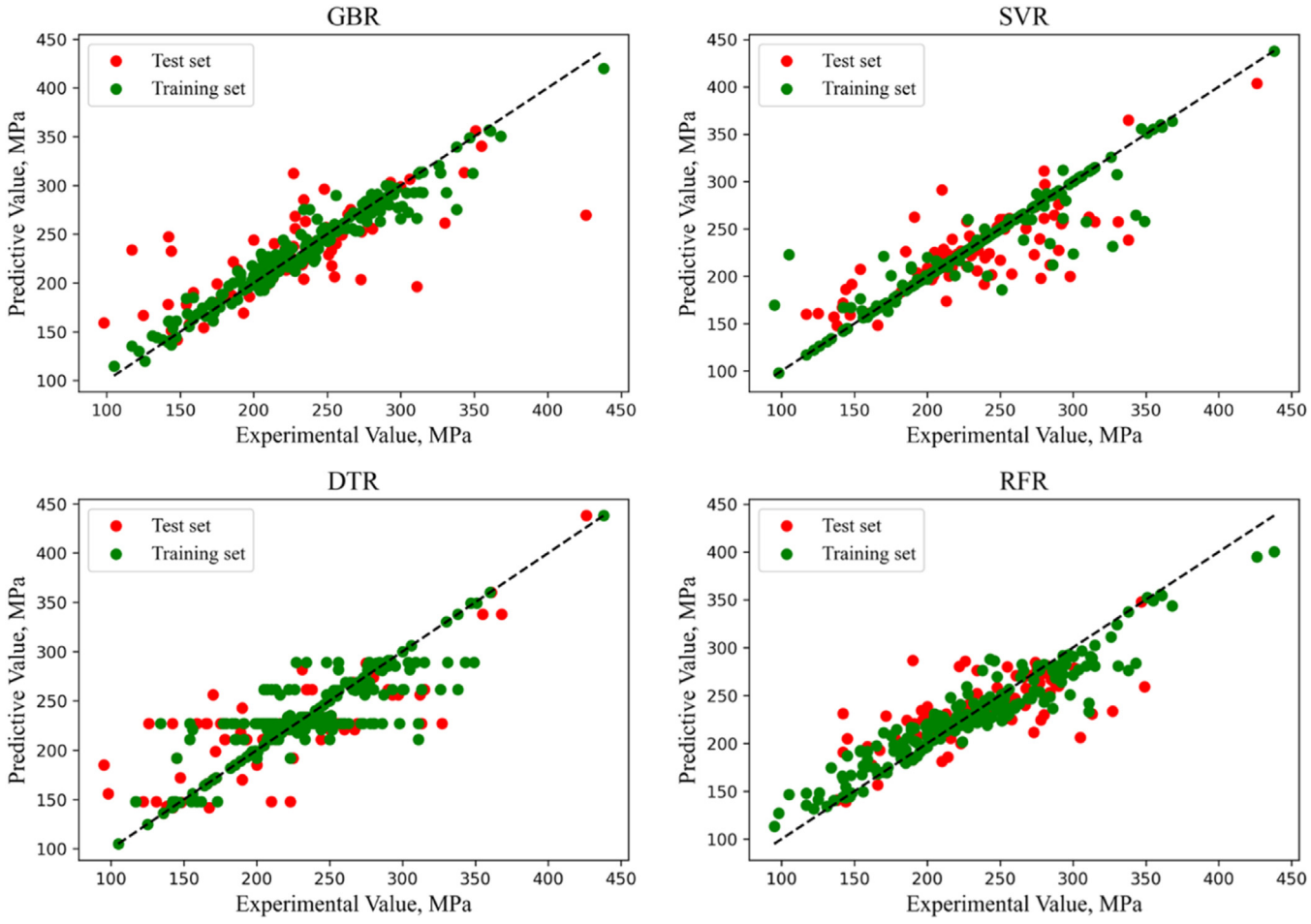

Fig. 5. Fitting performance of four YTS regression models.

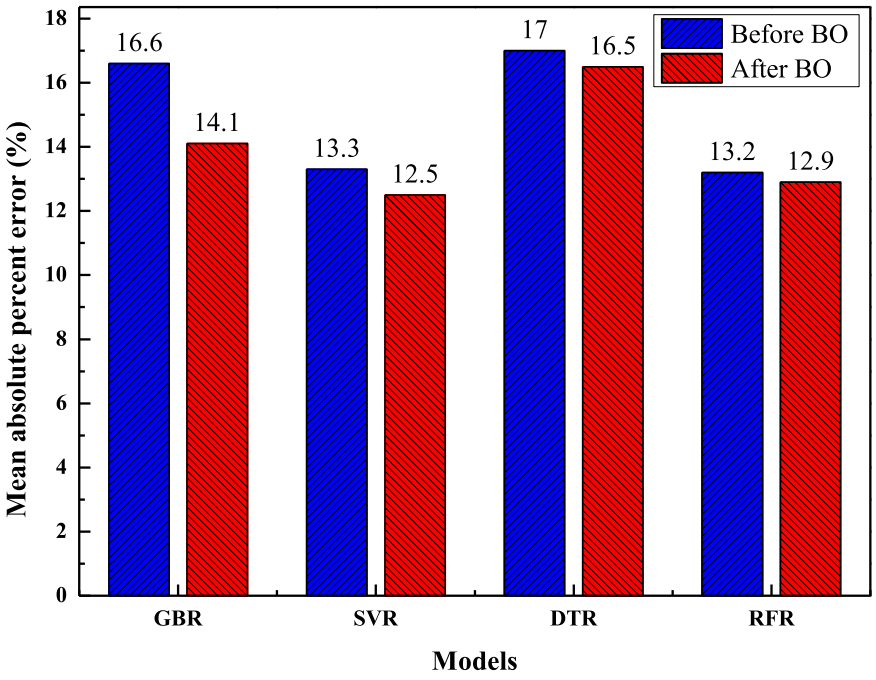

Fig. 6. Comparison of errors (MAPE) before and after using BO in UTS regression models.

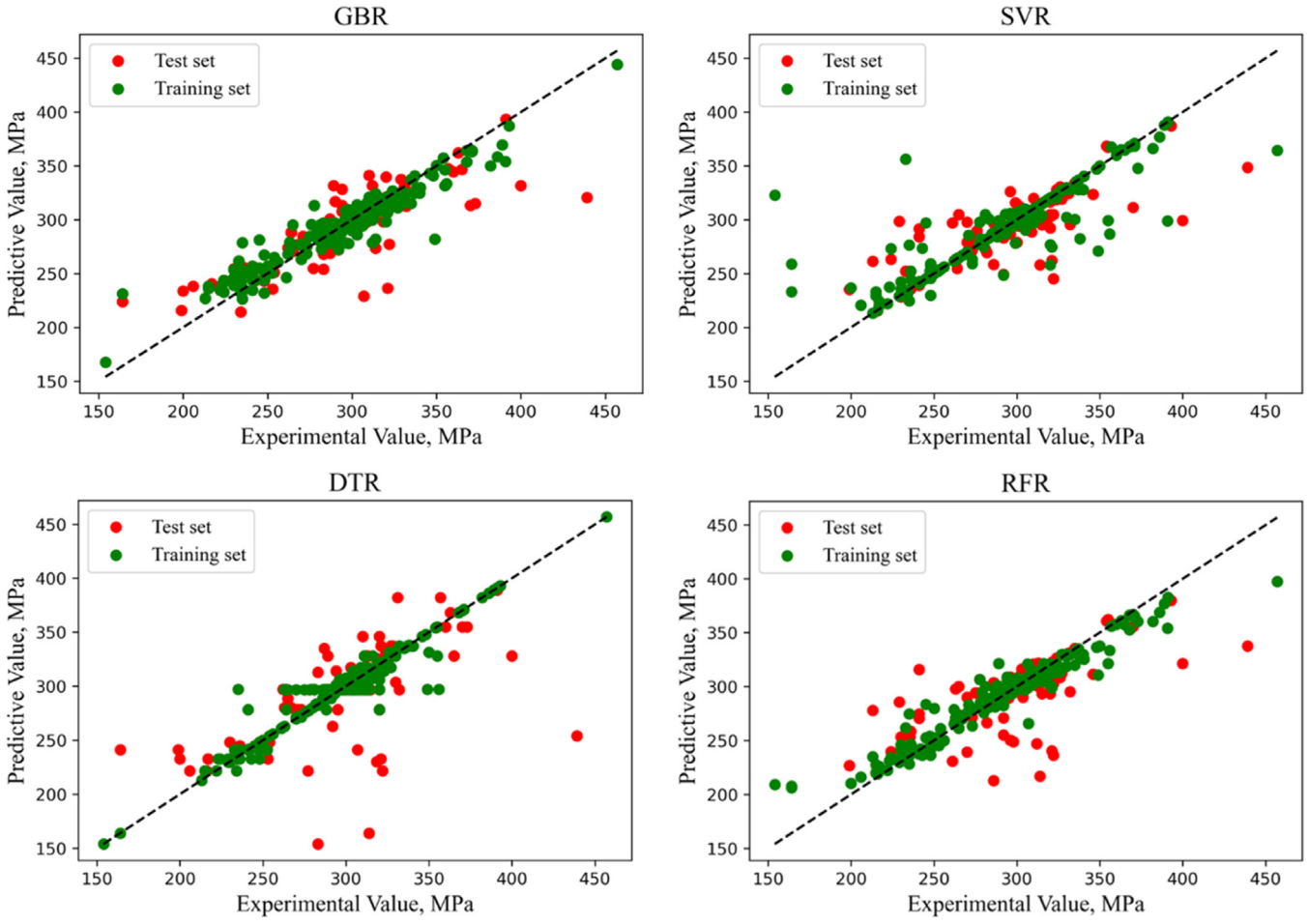

Fig. 7. Fitting performance of four UTS regression models.

Table 3
The hyperparameters and errors of UTS model optimized by BO.

Model	GBR	SVR	DTR	RFR
Hyperparameters before BO	n_estimators=100, max_depth=8	C=100, gamma=1	max_depth=6	n_estimators=100, max_depth=6
Hyperparameters after BO	n_estimators=301, max_depth=10	C=904, gamma=0.24	max_depth=8	n_estimators=206, max_depth=9
MAEs before BO	17.9	21.1	26.1	21.1
MAPEs before BO	16.6%	13.3%	17.0%	13.2%
MAEs after BO	15.3	17.2	22.3	20.5
MAPEs after BO	14.1%	12.5%	16.5%	12.9%

Table 4
The hyperparameters and errors of FE model optimized by BO.

Model	GBR	SVR	DTR	RFR
Hyperparameters before BO	n_estimators=100, max_depth=8	C=100, gamma=1	max_depth=6	n_estimators=100, max_depth=6
Hyperparameters after BO	n_estimators=227, max_depth=10	C=904, gamma=0.24	max_depth=8	n_estimators=206, max_depth=9
MAEs before BO	3.7	4.5	4.3	3.8
MAPEs before BO	46.3%	47.4%	48.1%	36.9%
MAEs after BO	3.2	4.3	4.1	3.7
MAPEs after BO	37.2%	35.3%	39.2%	32.3%

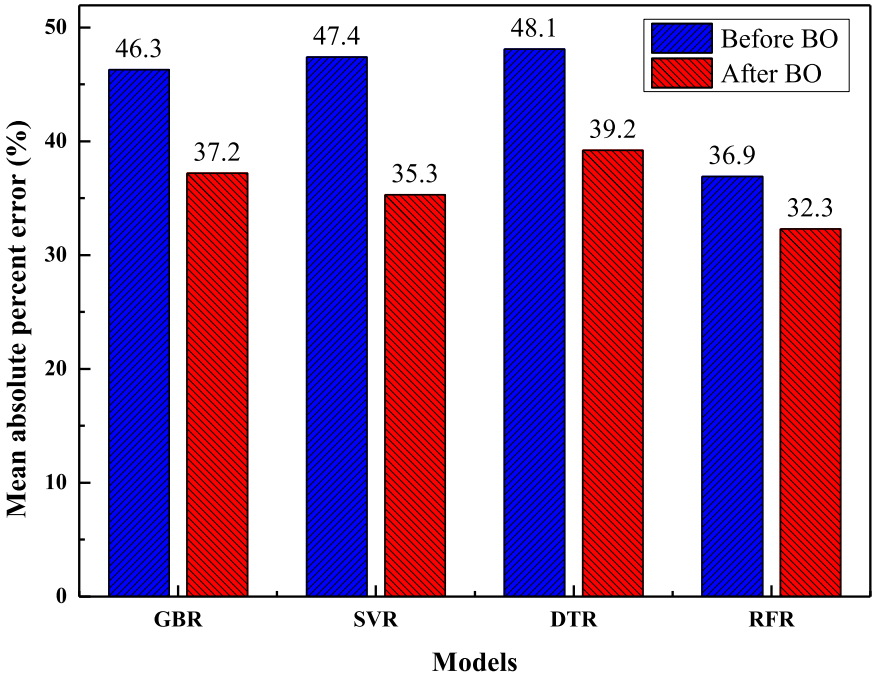

Fig. 8. Comparison of errors (MAPE) before and after using BO in FE regression models.

This indicates that these models cannot accurately describe the relationship between the input features and FE. The main reasons for this are: (i). The dataset size is relatively small, and it is not sufficient to establish a comprehensive relationship for FE using only this limited data. (ii). The selection of features is inappropriate, and some features have become noisy points, interfering with the establishment of the relationship function.

Fig. 9 displays the fitting results of the FE prediction models. From the graph, it can be observed that the SVR model exhibits the lowest fit, with a test set R^2 value of 0.58, indicating poor performance. On the other hand, GBR and DTR models show similar levels of fit, with test set R^2 values of 0.68 and 0.67, respectively. However, the RFR model demonstrates the best fitting performance, with a test set R^2 value of 0.70. Therefore, the RFR model is chosen as the prior function distribution (surrogate model) for subsequent Bayesian optimization of alloy plasticity models, denoted as RFR-FE.

4. Active learning and alloys design

Theoretically, the regression models established previously can predict the tensile yield strength, ultimate tensile strength, and fracture elongation of any Mg alloy with

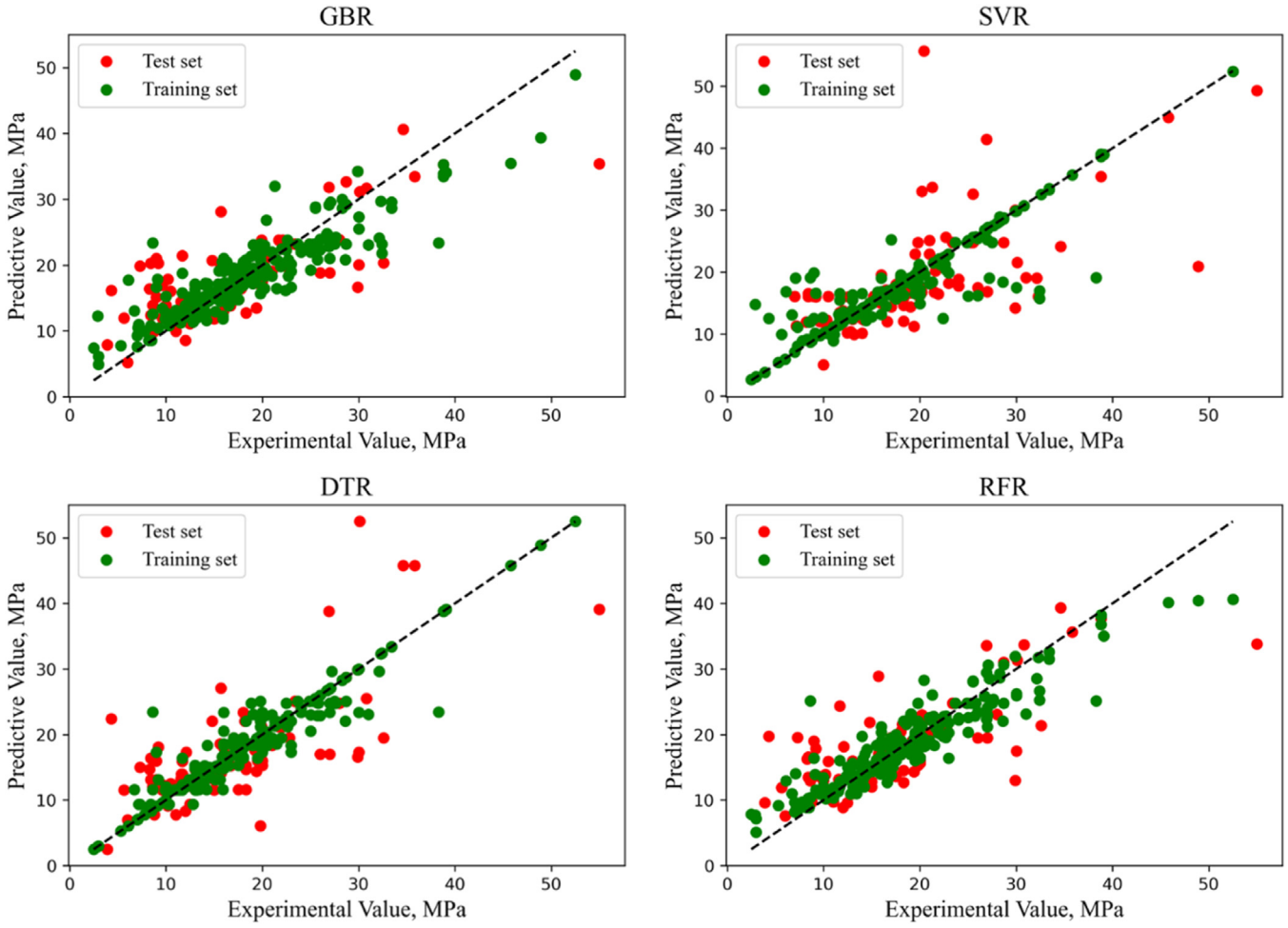

Fig. 9. Fitting performance of four FE regression models.

specific compositions and extrusion processes. However, due to the limited training data, which only covered a range of knowledge domains for the entire Mg-Mn-Al-Zn-Se-Ca system. When the model is applied outside of this domain, there is significant uncertainty. Furthermore, from practical application perspectives, researchers are often more interested in preparing alloys with higher strength or ductility. Machine learning models are generally black boxes, but if one is only interested in obtaining detailed information about the best-performing point within this black box, it is not necessary to traverse all the solutions in the entire ML model. It is sufficient to find the extreme value of this function, and the BO algorithm can effectively solve this problem.

In this section, an active learning model (RBOALM) based on Bayesian optimization algorithm was developed to guide the design of novel high-performance Mg alloys. The core thought of the RBOALM is to transform the problem of finding the optimal performance of alloys into a problem to search an extreme solution to a regression model. Generally, Gaussian Process (GP) is used as the prior distribution function in BO, also known as the surrogate model (or surrogate function). In this paper, in addition to using GP, the previously established regression models RFR-YTS, SVR-UTS, and RFR-FE aforementioned were also employed. This was done to explore whether using the established regression models as replacements for the GP function could provide a more appropriate description of the relationship between features and targets. The optimization process of RBOALM consists of four main steps:

Step 1: Establishing a surrogate function from the existing material dataset, which in this paper refers to the previously established regression models. These models can predict the values of target properties (YTS, UTS, and FE) based on the characteristics of any new material, such as compositions and process parameters.

Step 2: Apply the surrogate functions to the entire predefined search space to predict the performance of unexplored alloys.

Step 3: Constructing a utility function to determine the candidate materials for the next synthesis step. The utility function aims to balance the search for high-performance new alloys, the utilization of these alloys to expand the existing knowledge domain, and the reduction of overall uncertainty in the entire search space.

Step 4: Taking experimental tests of candidate materials, which are then added to the dataset. The surrogate model is

Table 5
Bayesian optimization of the search space.

Feature	Search range	Interval	Total
Mn	0-5.0 wt.%	0.1	50
Al	0-10.0 wt.%	0.1	100
Zn	0-10.0 wt.%	0.1	100
Sn	0-10.0 wt.%	0.1	100
Ca	0-4.0 wt.%	0.1	40
ET	175-550 °C	5	75
ES	0.5-15.0 m/min	0.5	29

updated using the enhanced dataset. The four steps mentioned above are repeated multiple times until the target properties reach their limits.

Using the RBOALM, we searched the designated composition and process parameter space to find wrought Mg alloys with the highest YTS, UTS, and FE values, respectively. The original ranges of each parameter in the dataset (Table 1) were as follows: Mn: 0-3.0 wt.%; Al: 0-9.0 wt.%; Zn: 0-8.8 wt.%; Sn: 0-9.7 wt.%; Ca: 0-3.5 wt.%; ET: 175-500 °C; ES: 0.2-12 m/min. When defining the search space for composition and process parameters, and appropriately expanded the ranges of the features to broaden the existing knowledge domain. Table 5 shows the search ranges of BO space, from which it can be inferred that it would require a total of 4.35×10^{12} computations to compute all feature combinations and select the extreme values by using conventional grid, enumeration, or iterative methods. Such a vast computational burden would significantly consume computing and time resources. Opting for the BO approach, we gradually obtained the global maximum through sampling, and the entire computation process took only a few minutes [67].

In the alloy design process, the BO method was employed, with GP, RFR-YTS, SVR-UTS, and RFR-FE used as surrogate functions, to explore the compositions and process parameters of high-performance alloys. Among these, three sets of alloys were designed using GP as the surrogate function in a control experiment to compare the improvement in model design accuracy before and after replacing GP with regression models. When designing alloys using the GP function, all feature parameters were chosen to be the same as the corresponding regression models. In the active learning process, the maximum number of iterations for BO was set to 100, and the process was terminated once this iteration limit was reached. The acquisition function used in this model was the Efficient global optimization (EGO) utility function [68], as shown in Eq. (3). The Expected Improvement (*EI*) value for each model was calculated based on the formula, representing the expected improvement upon incorporating new data. Here, μ and δ denote the mean and standard deviation of the predicted performance of the selected new materials using the current surrogate model, $z = (\mu - \mu^*) / \delta$, where μ^* is the maximum performance value in the current dataset. Additionally, $\varphi(z)$ and $\Phi(z)$are the standard normal density and cumulative distribution function [24].

$$EI(\mu, \delta) = \delta[\varphi(z) + Z(\Phi(z))] \quad (3)$$

Generally, ML models are considered as 'black boxes', making it challenging to observe the internal mechanisms and data flow of the ML system using conventional modeling techniques. In this paper, we innovatively present a visualization of the BO process with Al, Mn and YTS. This visualization allows us to gain a clear understanding of the proposed regression-Bayesian active learning process, as depicted in Fig. 10. Fig. 10(a) illustrates the heat map of the simulated relationship between Al and Mn contents and YTS using the RFR model. Each point in the graph represents a combination of Al and Mn components, with the color indicating the corresponding YTS value. It is evident from the graph that three distinct regions exhibit a light-yellow shade, indicating higher yield strength within these areas. Fig. 10(b) showcases the optimization process using BO within the compositional space simulated by the RFR model. The green points represent observation points for each iteration, while the red point indicates the globally optimal solution found by the active learning model after completion of the iterations, representing the point with the highest YTS value in the entire space. From Fig. (b), it is evident that when considering only Al and Mn as features, the global maximum value of YTS is 327 MPa, corresponding to Al and Mn contents of 1.1 wt.% and 2.5 wt.%, respectively.

In the actual alloy design of this work, both the YTS and FE regression models have six input features, while the UTS model has seven input features. This presents a significantly more complex scenario than the one depicted in Fig. 10, making it impossible to visualize the high-dimensional feature space of the BO workflow effectively. Nonetheless, the optimization principles and procedures remain identical to those illustrated in Fig. 10.

After optimizing each property of the alloy 10 times using the RBOALM, the top one sets with the highest *EI* values for each model were selected as the final results for alloy design. After multiple iterations, the alloy design results are shown in Table 6, represented by six alloys named as GP-YTS, GP-UTS, GP-FE, RFR-YTS, SVR-UTS and RFR-FE. The name of each alloy represents the type of surrogate function used and the optimized performance objective. From the table, it can be observed that when optimizing different alloy performance, the designed compositions and extrusion process parameters exhibit significant variations. However, when optimizing the same performance, whether employing GP or regression models as surrogate functions, the design outcomes show relatively minor discrepancies. This observation suggests that within the BO process, the optimization directions of regression-Bayesian and Gaussian- Bayesian are aligned. Furthermore, it also can be observed that the extrusion temperature for high-strength alloys is higher than that for highly ductile alloys, while there is no apparent pattern in extrusion speed.

5. Experimental validation and model evaluation

After completing the active learning design, six alloys were experimentally verified based on the design schemes provided

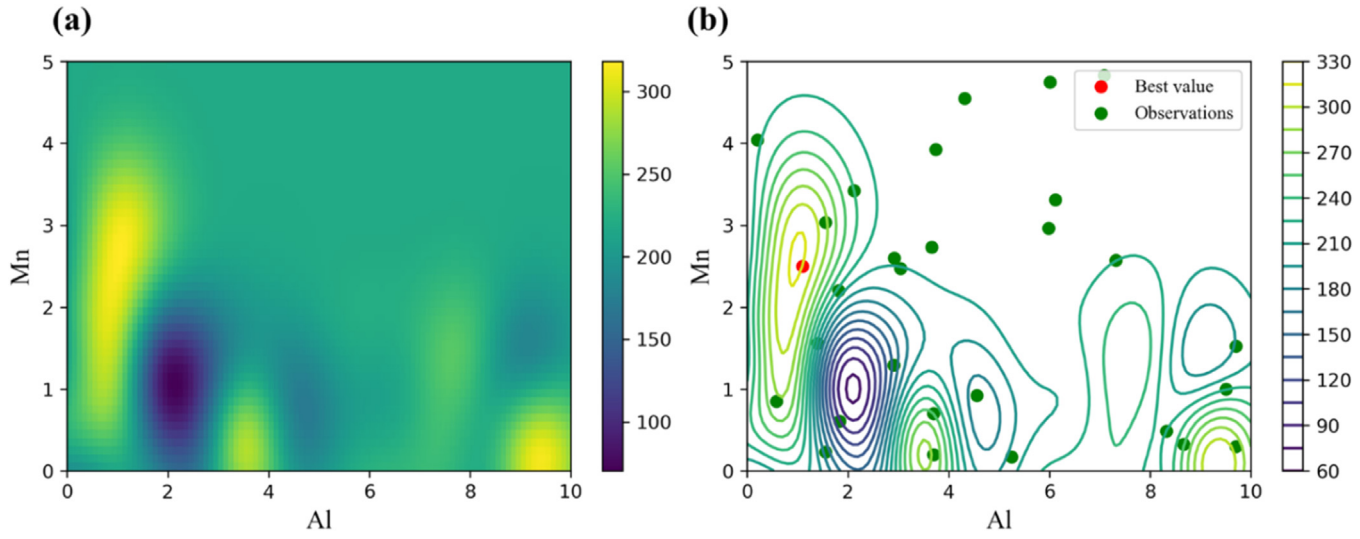

Fig. 10. Bayesian optimization process diagram for YTS performance with dual features Al and Mn. (a) Heatmap of YTS performance simulated by RFR model. (b) Bayesian optimization process diagram for YTS performance.

Table 6
The optimal results of properties and alloy design plans predicted by RBOALM.

Alloy	Predicted maximum	Alloy compositions(wt.%)	ET (°C)	ES (m/min)
GP-YTS	YTS=203 MPa	Mg-1.0Mn-0.1Al-5.0Zn-4.0Sn	350	1.0
GP-UTS	UTS=387 MPa	Mg-1.5Mn-1.5Al-1.0Sn-0.2Ca	300	3.0
GP-FE	FE=49%	Mg-0.5Mn-0.8Zn	270	2.0
RFR-YTS	YTS=253 MPa	Mg-0.8Mn-0.3Al-5.3Zn-3.6Sn	365	6.0
SVR-UTS	UTS=408 MPa	Mg-2.0Mn-2.1Zn-0.5Sn-0.1Ca	300	1.5
RFR-FE	FE=55%	Mg-2.7Mn-0.5Al-0.1Ca	270	2.5

in Table 6. The alloy ingots used in this study were prepared through a melting process. Prior to casting, the raw materials were cut into uniform small pieces, and the surface oxide layer was removed to ensure a smooth surface. The ingredients were then weighed and mixed accordingly. During the casting process, pure magnesium was first placed in a stainless-steel crucible (with dimensions of φ90 mm × 300 mm) and completely melted by heating at approximately 720 °C. Once the pure magnesium was molten, the alloying elements were sequentially added to the melt, and the alloy was stirred to ensure a homogeneous composition. Throughout the entire process, a mixed gas of CO_2 (99 vol.%) and SF_6 (1 vol.%) was used as the protective atmosphere to prevent oxidation. The melt was left undisturbed and kept at a constant temperature for 15 minutes to ensure complete alloy melting. After melting, the molten alloy undergoes water cooling, cutting, and cleaning, resulting in ingots with a diameter of 80 mm. Subsequently, the ingots labeled as RFR-YTS, SVR-UTS, RFR-FE, GP-YTS, GP-UTS and GP-FE were obtained and further processed through extrusion.

Extrusion was carried out using an XJ-500 horizontal extrusion machine at temperatures of 365 °C, 300 °C, 270 °C, 350 °C, 300 °C, and 270 °C for ingots, respectively. The ingots were preheated for 30 minutes before extrusion at each temperature. The extrusion speeds were set at 6.0 m/min, 1.5 m/min, 2.5 m/min, 1.0 m/min, 3.0 m/min, and 2.0 m/min, respectively, to obtain four rods with a final diameter of 16 mm. Tensile tests were performed on the extruded samples using an electronic universal testing machine (CMT6305-300 KN) with a strain rate of 1×10^{-3} s^{-1}. The engineering stress-strain curves for the extruded alloys are shown in Fig. 11.

After measurement and statistical analysis, the mechanical properties of each alloy are presented in Table 7, comparing the predictive performance of using regression models as surrogate functions versus using GP functions. The optimization performance predicted using SVR-UTS, RFR-YTS, and RFR-FE as surrogate functions all outperformed the GP model, with a notable enhancement in YTS performance, exceeding 50 MPa. When contrasting the mechanical performance metrics of the data set, the use of regression models in place of GP functions resulted in predictive values that closely approximated, and even surpassed, the extremal values present in the data set. This observation underscores the evident improvement in the ability of the active learning model to compute extrapolated boundaries upon adopting regression models to replace GP functions. The computed results indicate that the performance extrapolated boundaries for tensile yield strength, ultimate tensile strength, and fracture elongation are 253 MPa, 408 MPa, and 55%, respectively.

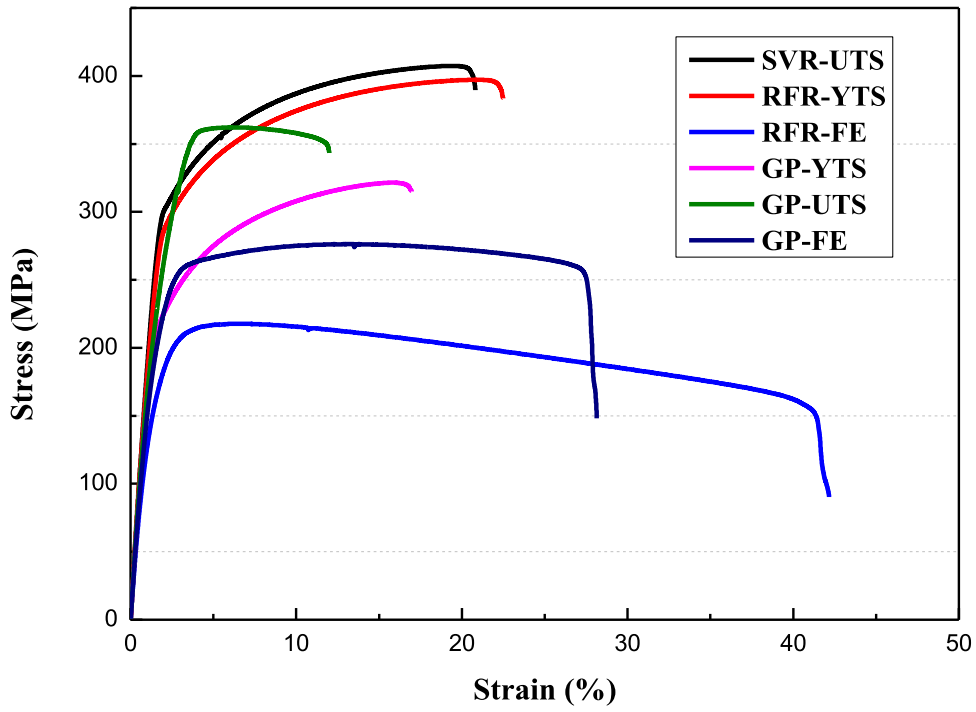

Fig. 11. The tensile engineering stress-strain curves of verification alloys.

Table 7
Predicted and experimental properties of verification alloys.

Target	Alloy	Predictive property	Tested property	Error
YTS	GP-YTS	YTS=203 MPa	YTS=183 MPa	10.9%
	RFR-YTS	YTS=253 MPa	YTS=202 MPa	25.2%
UTS	GP-UTS	UTS=387 MPa	UTS=361 MPa	7.2%
	SVR-UTS	UTS=408 MPa	UTS=406 MPa	0.5%
FE	GP-FE	FE=49%	FE=28%	75.0%
	RFR-FE	FE=55%	FE=41%	34.1%

In terms of performance design, concerning the ultimate tensile strength of the alloys, whether using regression models or GP function models, the actual UTS of the designed alloys closely matched the predicted UTS, with errors not exceeding 10.0%. In fact, the validation error for the SVR-UTS alloy was as low as 0.5%. This demonstrates the effectiveness of BO in developing high UTS magnesium alloys, and the SVR-UTS alloy exhibits outstanding comprehensive mechanical properties, with a ultimate tensile strength of 406 MPa, yield strength of 287 MPa, and fracture elongation of 23%. Regarding the yield tensile strength of the alloys, the validation error for the regression-Bayesian model was higher than that of the GP-Bayesian model. However, the validated yield strength was 24 MPa higher, indicating that the regression Bayesian model was more effective in computing extremal points and yielding better practical alloy designs. As for the fracture elongation of the alloys, the substitution of the GP function with the RFR model reduced the design error from 75.0% to 34.1%, leading in a 13% improvement of the designed performance. Overall, the RBOALM established in this study demonstrated superior performance in designing high-performance alloys, with the most optimal results achieved when designing high UTS magnesium alloys, and the design accuracy not falling below 90%.

6. Conclusions

This paper presents RBOALM, a novel machine learning-based alloy design model that combines active learning with regression and Bayesian optimization algorithms. RBOALM was developed specifically for designing high-performance yet cost-effective Mg-Mn-based wrought Mg alloys. Both regression models and GP functions were employed as surrogate functions for Bayesian optimization to optimize the mechanical properties of alloys, leading to the development of several new high-performance Mg alloys. The main conclusions of this study are as follows:

(1) Four regression algorithms were separately employed to construct predictive models. The random forest regression (RFR) and support vector regression (SVR) models had the highest accuracy, with goodness of fit close to 0.8. Utilizing these regression models, the extrapolated limits for yield tensile strength, ultimate tensile

strength, and fracture elongation were calculated as 253 MPa, 408 MPa, and 55%, respectively. These calculated results provide valid confidence intervals for alloy design.

(2) In RBOALM, Bayesian optimization algorithm was employed firstly to optimize the hyperparameters of the regression models, resulting in improved precision. Then, Bayesian optimization was innovatively utilized to optimize the mechanical properties of the alloys. Substituting the GP function with regression models as surrogate functions led to decreased design errors. The accuracy of the alloy design before and after replacing GP as surrogate function using the regression models were compared through experimental validation. The SVR-UTS model exhibits the highest accuracy at 99.5%, which is much higher than the 92.8% of GP-UTS. The design accuracy for fracture elongation was improved by 40.9% compare with RFR-FE and GP-FE, which is the most significant performance improvement observed in this metric.

(3) Utilizing the RBOALM, several novel high-performance and low-cost Mg-Mn-based series wrought Mg alloys were designed and prepared. After mechanical property testing of the designed alloys, the Mg-2.1Zn-2.0Mn-0.5Sn-0.1Ca alloy displayed an ultimate tensile strength of 406 MPa, a yield tensile strength of 287 MPa, and a fracture elongation of 23%. Furthermore, the Mg-2.7Mn-0.5Al-0.1Ca alloy exhibited an ultimate tensile strength of 211 MPa, coupled with a remarkable fracture elongation of 41%.

Declaration of competing interest

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

CRediT authorship contribution statement

Xiaoxi Mi: Conceptualization, Data curation, Methodology, Software, Writing – original draft. **Lili Dai:** Data curation, Validation. **Xuerui Jing:** Investigation, Validation. **Jia She:** Data curation, Resources. **Bjørn Holmedal:** Methodology, Supervision. **Aitao Tang:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **Fusheng Pan:** Funding acquisition, Supervision.

Acknowledgments

This work was supported by the National Natural the Science Foundation of China (51971042, 51901028), the Chongqing Academician Special Fund (cstc2020yszx-jcyjX0001). Xiaoxi Mi is also grateful to the China Scholarship Council (CSC) and Norwegian University of Science and Technology (NTNU) for their financial and technical support.

References

[1] Y. Yang, X. Xiong, J. Chen, et al., J. Magnesium. Alloys 9 (3) (2021) 705–747.
[2] J.F. Song, J. She, D. Chen, et al., J. Magnesium. Alloys 8 (1) (2020) 1–41.
[3] Z. Zeng, N. Stanford, C.H.J. Davies, et al., Int. Mater. Rev. 64 (1) (2018) 27–62.
[4] S. Li, X. Yang, J. Hou, et al., J. Magnesium. Alloys 8 (1) (2020) 78–90.
[5] H. Pan, Y. Ren, F. He, et al., J. Alloys Compounds 663 (2016) 321–331.
[6] L. Shao, C. Zhang, C. Li, et al., Mater. Charact. 183 (2022) 111651.
[7] P. Peng, X. He, J. She, et al., Mater. Sci. Eng. A Struct. Mater. 766 (2019) 138332.
[8] F. Hu, S. Zhao, G. Gu, et al., Mater. Sci. Eng. A Struct. Mater. 795 (2020) 139926.
[9] P. Peng, A.T. Tang, B. Wang, et al., J. Mater. Res. Technol. 15 (2021) 1252–1265.
[10] J. She, P. Peng, L. Xiao, et al., Mater. Sci. Eng. A 765 (2019) 138203.
[11] X.X. Mi, X.R. Jing, H.L. Wang, et al., J. Mater. Res. Technol. 23 (2023) 4576–4590.
[12] J.Y. Zhang, P. Peng, A.A. Luo, et al., Mater. Sci. Eng. A Struct. Mater. 829 (2022) 142143.
[13] K. Choudhary, B. DeCost, C. Chen, et al., NPJ Comput. Mater. 8 (1) (2022) 59.
[14] K.T. Butler, D.W. Davies, C. Hugh, et al., Nature 559 (7715) (2018) 547–555.
[15] P. Raccuglia, K.C. Elbert, P.D.F. Adler, et al., Nature 533 (7601) (2016) 73–76.
[16] R. Ramprasad, R. Batra, G. Pilania, et al., NPJ Comput. Mater. 3 (1) (2017) 54.
[17] M. Hu, Q. Tan, R. Knibbe, et al., Mater. Sci. Eng.: R: Reports 155 (2023) 100746.
[18] X.X. Mi, L.J. Tian, A.T. Tang, et al., Comput. Mater. Sci. 201 (2022) 110881.
[19] C. Wang, H. Fu, L. Jiang, et al., NPJ Comput. Mater. 5 (1) (2019) 2057–3960.
[20] K.L.M. Elder, J. Berry, B. Bocklund, et al., NPJ Comput. Mater. 9 (1) (2023) 84.
[21] D. Xue, P.V. Balachandran, J. Hogden, et al., Nat. Commun. 7 (2016) 11241.
[22] Q. Liang, A.E. Gongora, Z. Ren, et al., NPJ Comput. Mater. 7 (1) (2021) 188.
[23] B. Lei, T.Q. Kirk, A. Bhattacharya, et al., NPJ Comput. Mater. 7 (1) (2021) 194.
[24] Y. Liu, L. Wang, H. Zhang, et al., Metall. Mater. Trans. A 52 (2021) 943–954.
[25] W.C. Chen, J.N. Schmidt, D. Yan, et al., NPJ Comput. Mater. 7 (1) (2021) 02038.
[26] J. Ojih, M. Al-Fahdi, A.D. Rodriguez, et al., NPJ Comput. Mater. 8 (1) (2022) 143.
[27] Z.R. Zeng, Y.M. Zhu, R.L. Liu, et al., Acta Mater. 160 (2018) 97–108.
[28] L. Guo, D.F Zhang, Z.J. Hong, et al., Trans. Nonferrous Metals Soc. China 23 (3) (2013) 586–592.
[29] D.F. Zhang, G.L. Shi, X.B. Zhao, et al., Trans. Nonferrous Metals Soc. China 21 (1) (2011) 15–25.
[30] M.G. Jiang, C. Xu, T. Nakata, et al., J Alloys Compd 668 (2016) 13–21.
[31] T. Nakata, C. Xu, Y.C. Matsumoto, et al., Mater. Sci. Eng.: A 673 (2016) 443–449.
[32] J.D. Robson, D.T. Henry, B. Davis, Acta Mater. 57 (9) (2009) 2739–2747.
[33] P. Peng, A.T. Tang, B. Wang, et al., J. Mater. Res. Technol. 15 (2021) 1252–1265.
[34] X. Wu, X.R. Jing, H. Xiao, et al., J. Mater. Res. Technol.J. Mater. Res. Technol. 21 (2022) 1395–1407.
[35] A.A. Luo, A.K. Sachdev, Metall. Mater. Trans. A 38 (2007) 1184–1192.
[36] H. Yu, S.H. Park, B.S. You, Mater. Sci. Eng. A 610 (2014) 445–449.

[37] F. Pan, J. Mao, G. Zhang, et al., Prog. Natural Sci.: Mater. Int. 26 (6) (2016) 630–635.
[38] Z.T. Li, X.D. Zhang, M.Y. Zheng, et al., Mater. Sci. Eng. A 682 (2017) 423–432.
[39] N. Stanford, Mater. Sci. Eng. A 528 (1) (2010) 314–322.
[40] Z.W. Yu, A.T. Tang, J. He, et al., Mater. Charact. 136 (2018) 310–317.
[41] J. She, F.S. Pan, W. Guo, et al., Mater. Des. 90 (2016) 7–12.
[42] J. She, F.S. Pan, H.H. Hu, et al., J Mater Eng Perform 24 (2015) 2937–2943.
[43] C. Liu, X. Chen, J. Chen, et al., J. Magnesium. Alloys 9 (3) (2021) 1084–1097.
[44] H.X. Liao, J. Kim, T. Liu, et al., Mater. Sci. Eng. A 754 (2019) 778–785.
[45] Z.W. Yu, A.T. Tang, W. Qin, et al., Mater. Sci. Eng. A 648 (11) (2015) 202–207.
[46] T. Nakata, T. Mezaki, R. Ajima, et al., Scr. Mater. 101 (2015) 28–31.
[47] T. Nakata, C. Xu, R. Ajima, et al., Mater. Sci. Eng. A 712 (2018) 12–19.
[48] J. She, F.S. Pan, P. Peng, et al., Mater. Sci. Technol. 31 (3) (2015) 344–348.
[49] D.F. Zhang, G. Shi, Q. Dai, et al., Trans. Nonferrous Metals Soc. China 18 (2008) 59–63.
[50] P. Peng, J. She, A.T. Tang, et al., Mater. Sci. Eng. A Struct. Mater. 859 (2022) 144229.
[51] A. Zhang, R. Kang, L. Wu, et al., Mater. Sci. Eng. A Struct. Mater. 754 (2019) 269–274.
[52] P. Peng, J. She, A.T. Tang, et al., J. Alloys Compd. 890 (2022) 161789.
[53] S.J. Meng, H. Yu, S.D. Fan, et al., Acta Metallurgica Sinica (English Letters) 32 (2019) 145–168.
[54] F. Hu, S. Zhao, G. Gu, et al., Mater. Sci. Eng. A 795 (2020) 139926.
[55] J. Cai, K. Xu, Y. Zhu, et al., Appl. Energy 262 (2020) 114566.
[56] D.A. Otchere, T.O.A. Ganat, J.O. Ojero, et al., J. Pet. Sci. Eng. 208 (2022) 109244.
[57] M. Mohammed, M.B. Khan, E.B.M. Bashier, Machine Learning: Algorithms and Applications[M], Crc Press, 2016.
[58] G. Bonaccorso, Machine Learning Algorithms[M], Packt Publishing Ltd, 2017.
[59] L. Breiman, Mach. Learn. 45 (2001) 5–32.
[60] P.M. Tagade, S.P. Adiga, S. Pandian, et al., NPJ Comput. Mater. 5 (1) (2019) 127.
[61] E. Frank, L. Trigg, G. Holmes, et al., Mach. Learn. 41 (2000) 5–25.
[62] E. Schulz, M. Speekenbrink, A. Krause, J. Math. Psychol. 85 (2018) 1–16.
[63] V.L. Deringer, A.P. Bartók, N. Bernstein, et al., Chem. Rev. 121 (16) (2021) 10073–10141.
[64] X. Li, Y. Wang, S. Basu, et al., Adv. Neural. Inf. Process Syst. 32 (2019).
[65] J. Kazemitabar, A. Amini, A. Bloniarz, et al., Adv. Neural. Inf. Process. Syst. 30 (2017).
[66] A. Sutera, G. Louppe, V.A. Huynh-Thu, et al., Adv. Neural. Inf. Process. Syst. 34 (2021) 3533–3543.
[67] C. Fare, P. Fenner, M. Benatan, et al., NPJ Comput. Mater. 8 (1) (2022) 257.
[68] W. Scott, P. Frazier, W. Powell, SIAM J. Optim. 21 (3) (2011) 996–1026.