ARTICLE

OPEN

Check for updates

Accelerated discovery of high-strength aluminum alloys by machine learning

Jiaheng Li [1], Yingbo Zhang [1✉], Xinyu Cao[1], Qi Zeng[1], Ye Zhuang[1], Xiaoying Qian[1] & Hui Chen[1]

Aluminum alloys are attractive for a number of applications due to their high specific strength, and developing new compositions is a major goal in the structural materials community. Here, we investigate the Al-Zn-Mg-Cu alloy system (7xxx series) by machine learning-based composition and process optimization. The discovered optimized alloy is compositionally lean with a high ultimate tensile strength of 952 MPa and 6.3% elongation following a cost-effective processing route. We find that the Al_8Cu_4Y phase in wrought 7xxx-T6 alloys exists in the form of a nanoscale network structure along sub-grain boundaries besides the common irregular-shaped particles. Our study demonstrates the feasibility of using machine learning to search for 7xxx alloys with good mechanical performance.

[1] Key Laboratory of Advanced Technologies of Materials, Ministry of Education, School of Materials Science and Engineering, Southwest Jiaotong University, 610031 Chengdu, China. ✉email: yingbozhang@163.com

Al–Zn–Mg–Cu (7xxx series) alloys have been vastly used in the aerospace industry and have shown increasingly appreciable potentials in rail transportation due to their superior physiomechanical properties and manufacturability[1–3]. In recent decades, some emerging engineering materials (e.g., magnesium alloys, titanium alloys, and composites) have achieved rapid developments and challenged 7xxx alloys in many fields[4]. Therefore, further performance advancements are deeply required for 7xxx alloys to persistently remain competitive in their dominant fields and to get more opportunities in new application areas.

Mechanical strength is a basic consideration for structural materials. The ultimate tensile strength (UTS) of commercial 7xxx alloys is typically below 700 MPa[4]. Interests in improving the UTS of 7xxx alloys have never ceased. Some sophisticated techniques [e.g., severe plastic deformation (SPD), rapid solidification and powder metallurgy (RS/PM), spray forming, and multistage heat treatments] have pushed the UTS of 7xxx alloys to extremely high levels, in excess of 750 MPa[4–10]. For instance, an ultrafine-grained 7475 alloy and a nano-grained 7075 alloy processed by high-pressure torsion (HPT, an SPD processing), manifested UTSs of more than 900 MPa[5,11]. However, these techniques are now confined by their inabilities to fabricate large-size products, complex operations, high costs, or high requirements for facilities, which limit their extensive industrial uses. To develop high-strength 7xxx alloys aimed at industrializing, optimizing the alloy composition could be a practical strategy.

7xxx alloys usually contain main alloying elements of Zn, Mg, and Cu, as well as trace elements of Cr, Mn, Zr, Ti, Sc, etc. Many efforts have been made to tailor the contents of the main alloying elements for excellent mechanical properties[12–14]. It now appears that high Zn and Mg contents (Zn > 8 wt.%; Mg > 2.0 wt.%) are necessary for ensuring ultra-high strength, but they synchronously increase localized corrosion susceptibility, and the hot tearing susceptibility (a catastrophic problem in industrial production) and macro-segregation during casting[12,15–17]. As an effective grain refiner and an anti-recrystallization agent, Zr is a nearly indispensable trace element in 7xxx alloys. High-strength 7xxx alloys containing 0.1–0.2 wt.% Zr are often present[18]. Moreover, combined addition of Zr and Ti has a better strengthening effect than single addition of Zr thanks to the formation of $L1_2$-$Al_3(Zr_x,Ti_{1-x})$, which is more stable than $L1_2$-Al_3Zr dispersoids[19]. Micro-alloying of rare-earth elements (e.g., Sc, Yb, Y, Ce, and Gd) is a potent approach to modifying the microstructure and properties of 7xxx alloys given its grain refinement and recrystallization inhibition effects[20–23]. Among all

the frequently-used rare-earth elements in 7xxx alloys, Sc element is deemed as the most effective. Nevertheless, it is imperative to research and find other rare-earth elements cheaper than Sc (e.g., Y and Ce) to replace Sc owing to the prohibitive price of Sc. In a word, combined addition of multiple alloying elements is an important development trend of 7xxx alloys. Given the wide composition range[24] and composition-sensitivity of 7xxx alloys, numerous undiscovered alloys may outperform the existing counterparts (regardless of possible strengthening effects of rare-earth elements). The vast unexplored composition space, thus, presents a good opportunity to develop 7xxx alloys with desired properties. However, in combination with their numerous processing steps, 7xxx alloys are extremely complex and thus hard to optimize via intuition and trial-and-error. Recently, machine-learning-assisted design of multicomponent materials has attracted special interests[25–31]. We believe that machine learning could play an important role in composition optimization of multicomponent 7xxx alloys, although related publications are limited[32].

Our objective here is to discover new 7xxx alloys with desired UTS by machine learning. Here we propose a modified, Kriging model-based[33] efficient global optimization (EGO) algorithm[34] and apply it to composition optimization of 7xxx alloys. A 950 MPa grade 7xxx alloy was developed with Zn content of less than 7 wt.%. Meanwhile, we found the unusual formation of an Al_8Cu_4Y nanoscale network structure in wrought 7xxx-T6 alloys, which may be useful for future alloy design of high-strength aluminum alloys. This study demonstrates the feasibility of using machine learning to search for 7xxx alloys with good mechanical performance. Alloys based on the optimized alloy will be candidates for the mass production of a critical part in high-speed trains.

Results

Summary of the research strategy. Our research roadmap is shown in Fig. 1. First, we prepare a training data set for subsequent model evaluations and constructions. It contains some selected Al–Zn–Mg–Cu–(Ti)–(Y)–(Ce) alloys with known UTSs. Then, we evaluate or validate the machine-learning model through "leave-one-out cross-validation"[34]—where one observation in the original data set is treated as a test point and predicted back based on the remaining observations—to determine its feasibility. An iterative process (the so-called "adaptive design loop") will be introduced if the machine-learning model is valid. It includes steps as follows: (π) construct the machine-learning model based on the training data set to establish the composition–UTS relationship; (θ) use the "expected improvement" function[34] as a global selector to recommend the next experiments; (ρ) decide whether or not to stop iteration according to the maximum value of "expected improvement" function and the current-highest UTS; and (σ) refit the machine-learning model by incorporating newly generated data points if the stopping criterion is not satisfied. We obtain the optimized alloy by adding 0.2 wt.% Zr (a designed content) into the current-best Al–Zn–Mg–Cu–Ti–(Y)–(Ce) alloy after we stop iteration. Note that we exclude the key alloying element Zr (which is beneficial to improving the strength if suitably added) at the stage of composition optimization, mainly because of its hard-to-control burning loss during casting. Finally, we prepare the extrusion rods of the optimized alloy through a range of traditional processing techniques and characterize the extrusions systematically.
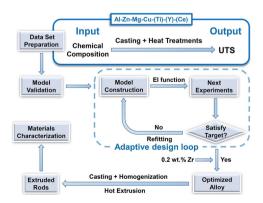

Fig. 1 Research roadmap of this study. Machine-learning-assisted composition optimization and a range of traditional processing techniques are integrated for the rapid discovery of Al–Zn–Mg–Cu–Ti–(Y)–(Ce) alloys with desired ultimate tensile strength (UTS) and characterizations of the optimized Al–Zn–Mg–Cu–Zr–Ti–(Y)–(Ce) alloy, respectively.

Search space and data set. We considered Al–xZn–yMg–zCu–uTi–vY–wCe with x, y, z, u, v, and w compositions, where the alloying elements were constrained within $5.1 \leq x \leq 8.4$, $1.2 \leq y \leq$
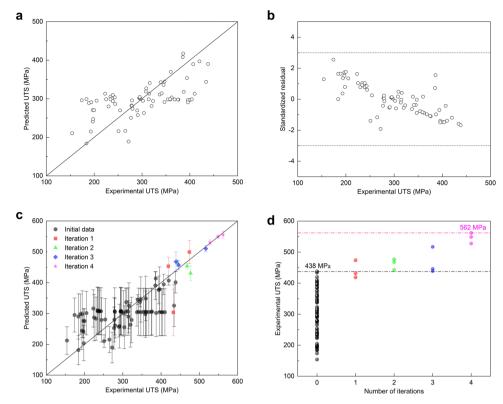
Fig. 2 Results of the model evaluations and adaptive design loop. a and **b** Diagnostic test plots for the machine-learning model before iteration, based on the initial training data set. **c** Diagnostic test plot for the updated machine-learning model after four iterations, based on the overall data set. The error bars correspond to standard error given by the machine-learning model. **d** Results of the adaptive design loop, showing the escalating trend of the ultimate tensile strength (UTS).

Table 1 Our newly synthesized alloys during the iterative process.

Alloy no.	Nominal elemental-content (wt.%)							UTS (MPa)	
	Zn	Mg	Cu	Y	Ce	Ti	EI value	Predicted	Experimental
1–1	7.67	1.65	2.38	0.40	0	0.04	19.037	454 ± 68	419 ± 6
1–2	7.85	1.68	2.38	0.38	0.06	0.05	15.767	446 ± 86	432 ± 8
1–3	7.57	1.68	2.38	0.38	0.01	0.07	10.008	411 ± 156	474 ± 12
2–1	6.89	1.70	2.38	0.35	0	0.08	11.261	477 ± 73	443 ± 2
2–2	7.36	1.69	2.38	0.36	0	0.08	9.854	482 ± 39	468 ± 2
2–3	6.20	1.71	2.37	0.35	0.01	0.08	8.741	456 ± 120	477 ± 8
3–1	6.08	1.72	2.35	0.38	0.01	0.08	28.260	502 ± 93	439 ± 8
3–2	7.13	1.67	2.35	0.41	0	0.08	24.060	492 ± 116	446 ± 2
3–3	7.43	1.78	2.36	0.41	0	0.09	20.584	486 ± 118	517 ± 7
4–1	8.20	1.82	2.37	0.40	0	0.10	26.822	543 ± 66	528 ± 1
4–2	7.84	1.86	2.37	0.40	0	0.10	22.741	538 ± 65	549 ± 13
4–3	7.47	1.93	2.37	0.39	0	0.09	10.529	509 ± 107	562 ± 18

Notes: "1–2" represents the second alloy in the first iteration. The predictive confidence-interval is the prediction plus or minus three standard errors given by the machine-learning model.
EI expected improvement function.

2.9, $0.8 \leq z \leq 2.6$, $0.02 \leq u \leq 0.2$, $0 \leq v \leq 0.5$, and $0 \leq w \leq 0.5$ wt.%. The aforementioned ranges of Zn, Mg, Cu, and Ti were determined by referring to five typical commercial 7xxx alloys (Supplementary Table 1) to narrow the search space moderately. The search space of Al–xZn–yMg–zCu–uTi–vY–wCe, thus, corresponds to a huge space and our objective is to uncover alloys with desired strength over this space as rapidly as possible.

A data set containing proper features and associated properties is required for machine-learning models to capture the relationship between features and properties. Proper features should be able to define a material quantitatively but not redundantly and

are usually the object to be optimized or discovered. Here, we take the alloy composition and UTS as the input features and output property, respectively. Before the tensile test, each cast alloy in the data set underwent identical heat treatments (see "Methods") to estimate its capacity of ageing strengthening, which is the main strengthening source of 7xxx alloys. To minimize the disturbance of processing on the property, all the alloys were fabricated by our laboratory under identical conditions. The data set, as a whole, is scattered and representative, containing 38 quaternary alloys, 25 six-component alloys, and 10 seven-component alloys (Supplementary Table 2).
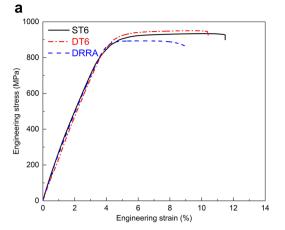
Model evaluation and iteration results. We performed diagnostic tests for the machine-learning model before and after iteration to evaluate their performance through cross-validation (Fig. 2a–c). The cross-validation result of the machine-learning model before iteration is shown in Fig. 2a, where most points distribute on both sides of a 45° line and few lie on the 45° line. The points should lean close to the 45° line if the machine-learning model were good enough. However, it was tough to make an impersonal and black-and-white judgment only according to Fig. 2a. Thus, we introduced the "standardized cross-validated residual"[34] in Fig. 2b, where all the standardized residuals of the points are within the interval of [−3, +3], demonstrating that the machine-learning model is valid and the data points in the training data set coexist harmoniously.

We re-evaluated the machine-learning model after four iterations to examine its improvement (Fig. 2c). The machine-learning model was locally upgraded, resulting in more accurate predictions at points with high experimental UTSs but similarly rough predictions at most of the initial data points compared with the early model evaluation result. We can find that some points, with experimental UTSs ranging from 174 to 410 MPa, possess exceptionally similar predicted UTSs of ~300 MPa (Fig. 2a, c). This suggests the inability of the machine-learning model to capture the relative variations of these points due to their isolated positions in the search space. Consequently, the machine-learning model helplessly predicted them with values close to the average value of all the known experimental UTSs at that time, accompanied by large uncertainties.

The results of our adaptive design loop are shown in Fig. 2d and Table 1. The UTSs of 7xxx alloys with different compositions can vary on a large scale (here from 154 to 438 MPa, Fig. 2d), demonstrating the composition-sensitivity of 7xxx alloys. On the whole, the UTS increases with the iteration. All but two alloys (alloys 1–1 and 1–2) of the newly synthesized alloys have UTSs exceeding 438 MPa (the best value in the initial data set). The UTS of alloy 3–3 exceeds 500 MPa, and all the alloys synthesized in iteration 4 have UTSs over 520 MPa. In particular, alloy 4–3 shows a UTS as high as 562 MPa, with an increment of 124 MPa compared with the best alloy in the initial data set. Indeed, we were reluctant to part with such a successful iterative process as the "expected improvement" function values indicated the remaining optimization space—the "1% stopping criterion"[34] was not satisfied. However, our targeted UTS of 550 MPa (see "Methods") was reached. Next, we will focus on the microstructure and tensile properties of the optimized alloy (with the actual composition of Al–6.49Zn–2.52Mg–1.92Cu–0.25Zr–0.07Ti–0.29Y in wt.%; alloy 4–3 + Zr). It should be pointed out that due to the disparity between the actual and our predetermined burning loss rates of the alloying elements, the actual chemical compositions of the alloys involved at the stage of composition optimization (including alloy 4–3) are expected to deviate from their nominal ones given in Table 1 and Supplementary Table 2. However, we emphasize that the use of the predetermined burning loss rates had little influence on the composition optimization as we used consistent values in all the alloy preparations. The following is for the extrusion rods if not otherwise specified.

Tensile properties and microstructure of the optimized alloy. After a double-stage solution treatment and a subsequent retrogression and re-ageing (RRA) treatment (see "Methods") used at the stage of composition optimization, the optimized alloy shows a UTS of 896 MPa and 4.7% elongation (Fig. 3a, b). To the best of our knowledge, the UTS near 900 MPa is a record one for traditionally processed 7xxx alloys based on ingot metallurgy. A widely used T6 treatment (120 °C/24 h) was applied to maximize
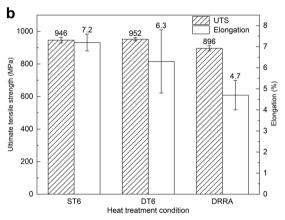
Fig. 3 Room-temperature tensile properties of the optimized alloy under various heat treatment conditions. a Typical engineering stress-strain curves. **b** Results of the ultimate tensile strength (UTS) and elongation. The error bars indicate standard deviation. ST6: single-stage solution treatment of 470 °C/2 h + room-temperature water quench (RTWQ) + peak ageing of 120 °C/24 h. DT6: double-stage solution treatment of (450 °C/1 h + 470 °C/2 h) + RTWQ + peak ageing of 120 °C/24 h. DRRA: double-stage solution treatment of (450 °C/1 h + 470 °C/2 h) + RTWQ + retrogression and re-ageing (RRA) of (120 °C/24 h + 190 °C/30 min + 120 °C/24 h).

the strength of the extrusions subjected to single- and double-stage solution treatments, culminating in the UTSs of 946 and 952 MPa, respectively. It is also found that the ST6 and DT6 treatments outperform the DRRA treatment in terms of both UTS and elongation. The DT6 treatment was therefore selected for the extrusions in the subsequent research.

To better understand the excellent mechanical properties obtained and the role of Y in the optimized alloy, systematical microstructural analysis was performed. The XRD pattern in the as-cast condition suggests that, apart from the common $MgZn_2$ phase, Al_8Cu_4Y and $Al_{20}Ti_2Y$ phases are formed (Fig. 4a). Figure 4b shows the quite heterogeneous microstructure of the α-Al matrix, consisting of coarse (>30 μm) grains and fine (<10 μm) recrystallized grains or sub-grains. The bimodal grain size distribution is usually beneficial to mechanical properties, especially the ductility[35,36]. Combined with the SEM analysis results (Fig. 4d, e), the second phases observed in Fig. 4b, c can be classified into three types: (π) tiny (<5 μm) Al_8Cu_4Y particles on grain boundaries; (θ) coarse (5–30 μm) $Al_{20}Ti_2Y$ particles with polygonal outlines, such as parallelogram and hexagon; and (ρ) coarse (5–50 μm) primary Al_3Zr and $Al_3(Zr,Ti)$ particles. Note that both Al_8Cu_4Y and $Al_{20}Ti_2Y$ particles contain strengthening solute atoms of Zn and Mg (Fig. 4e).
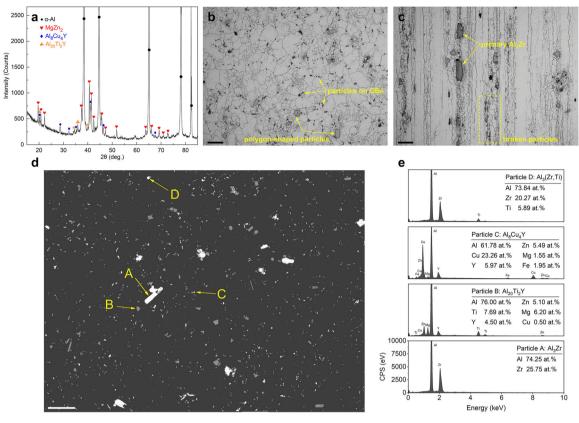
Fig. 4 X-ray, OM, and SEM characterizations of the optimized alloy. a XRD pattern in the as-cast condition. **b** and **c** OM images of the DT6-treated alloy on cross and longitudinal sections, respectively. Scale bars, 50 μm. **d** SEM-BSE image of the DT6-treated alloy on cross section showing four typical remaining constituent particles (indicated by yellow arrows A-D). Scale bar, 100 μm. **e** EDS analysis results of particles in **d**.

Figure 5a shows the fine (<3 nm), dense matrix precipitates, the discontinuous grain boundary precipitates of η-$MgZn_2$, and the high-density dislocations in the DT6-treated alloy. Figure 5b reveals a typical deformed area filled with ultrafine (100–800 nm) sub-grains. The SADPs confirm the presence of Al_8Cu_4Y (body-centered tetragonal structure) and $Al_{20}Ti_2Y$ (diamond cubic structure) phases (Fig. 5c, d). The ultrafine $Al_{20}Ti_2Y$ particle in Fig. 5d lies on sub-grain boundaries, which is different from the coarse polygon-shaped $Al_{20}Ti_2Y$ particles observed in Fig. 4d. In addition, the dislocation entanglements can be observed in Fig. 5d. More interestingly, we find that Al_8Cu_4Y phase has two distinct morphologies: the irregular-shaped particles on or near grain boundaries (Fig. 5e) and a quasi-continuous, nanoscale network structure along the sub-grain boundaries near Al_8Cu_4Y particles (Fig. 5f). For 7xxx alloys, second phases presenting as network structures are common in as-cast microstructures. To the best of our knowledge, such a network structure, however, has not been reported in as-deformed microstructures as yet—its particularity lies in its size and distribution.

Discussion

For comparison, we enumerated five previously reported 7xxx alloys with ultra-high (>750 MPa) UTSs in Table 2. A 7xxx alloy with UTS of >900 MPa came out as early as 1995 by an RS/PM approach (alloy #1), which was capable of fabricating highly-alloyed 7xxx alloys. However, the intrinsic drawbacks of the traditional RS/PM approaches, (e.g., tedious technology process, limitation in sample dimension, and bad plasticity) restrict their industrial applications to some extent. HPT can greatly improve the UTS and maintain considerable elongation for even moderately-alloyed 7xxx alloys (alloy #2). Unfortunately, there is still a long way for SPD methods from being industrially used,

despite their superior grain refinement effect. Combining the advantages of both ingot metallurgy and RS/PM, spray forming is a promising forming technique for large-size, highly-alloyed 7xxx alloys. It can produce 7xxx alloys with UTSs over 800 MPa (alloys #3 and 4). As mentioned above, it is reasonable to deduce that the ultra-high strength of 7xxx alloys is guaranteed by high contents of alloying elements (especially Zn and Mg) and/or advanced forming techniques. In sharp contrast, the optimized alloy in this work exhibits excellent mechanical properties by only a range of traditional processing techniques. Its comprehensive mechanical property is more superior to alloys #1, 3, 4, and 5, while its Zn content is much lower. Thus, 7xxx alloys that are modified on the basis of the optimized alloy are expected to have a bright prospect in industrial applications considering their advantages in mechanical properties and alloy composition.

It is well known that the strength of 7xxx alloys relies mainly on the formation of matrix precipitates (GP zones and/or η'-$MgZn_2$ phase) during ageing treatment. Moreover, the formation and strengthening mechanisms of the matrix precipitates have been well understood now[37]. As previously stated, our optimization framework works primarily by identifying alloys with an excellent ageing strengthening effect. The resulting cast alloy 4–3 exhibits the UTS of 562 MPa (Table 1), which is comparable with some commercial 7xxx alloys subjected to a complete processing route [casting → homogenization treatment → plastic deformation (e.g., extrusion, rolling, and forging) → solution treatment → ageing treatment]. That is, alloy 4–3 owns a significantly powerful ageing strengthening capacity, which set the tone for the high strength of the optimized alloy (alloy 4–3 + Zr). The carefully-designed extrusion parameters and heat treatment schedules also contributed to the high strength of the optimized alloy, beyond the machine-learning-assisted composition optimization.
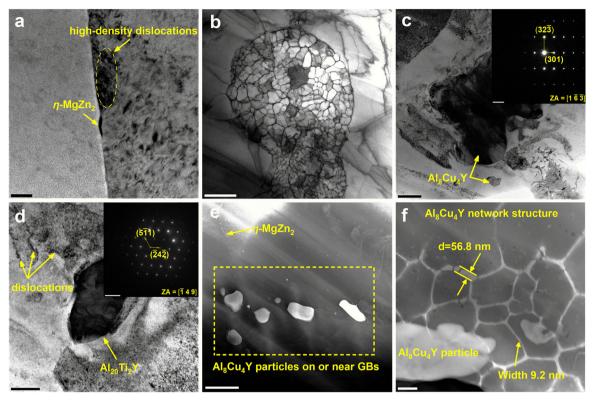
Fig. 5 TEM images of the optimized alloy in the DT6 condition. a Bright-field image showing intragranular and intergranular precipitates, as well as high-density dislocations. Scale bar, 100 nm. **b** Bright-field image showing typical deformed microstructure. Scale bar, 2 μm. **c** and **d** Bright-field images and corresponding selected area diffraction patterns (SADPs) of Al_8Cu_4Y and $Al_{20}Ti_2Y$ particles, respectively. Scale bars, 1 μm in **c**, 500 nm in **d**, and 5 1/nm in the inserted SADPs. **e** and **f** HAADF images showing two different morphologies of Al_8Cu_4Y phase. Scale bars, 2 μm in **e**, and 200 nm in **f**.

Table 2 Tensile properties of ultra-high-strength 7xxx alloys in this work and other researches.

Alloy composition (wt.%) and code	Processing	UTS (MPa)	EL (%)	Ref.
Al–9Zn–3Mg–1.5Cu–4Mn–0.5Zr–0.04Ag (#1)	RS/PM + Extr. + T6	916	0.6	[6]
Al–5.93Zn–2.5Mg–1.38Cu–0.18Cr–0.02Ti (#2)	IM + ST + HPT	>1000	9.0	[5]
Al–11.3Zn–2.65Mg–1.06Cu–0.14Zr (#3)	SF + Extr. + T6	823	5.0	[9]
Al–10.89Zn–2.43Mg–1.67Cu–0.17Zr–0.2Sc (#4)	SF + Extr. + T6	831	10.2	[10]
Al–9.3Zn–2.4 Mg–1.5Cu–0.16Zr (#5)	IM + Extr. + T6	753	8.4	[13]
Al–6.49Zn–2.52Mg–1.92Cu–0.25Zr–0.07Ti–0.29Y	IM + Extr. + DT6	952	6.3	This work

Notes: *UTS* ultimate tensile strength, *EL* elongation to failure, *RS/PM* rapid solidification and powder metallurgy, *Extr.* extrusion, *T6* a combinative processing of solution treatment, room-temperature water quench (RTWQ), and peak ageing (typically 120 °C/24 h) to maximize the tensile strength, *IM* ingot metallurgy, *ST* solution treatment including RTWQ here. *HPT* high-pressure torsion, *SF* spray forming. *DT6* double-stage ST of (450 °C/1 h + 470 °C/2 h + RTWQ) + peak ageing of 120 °C/24 h.

Table 3 Compositional comparison between the optimized alloy and commercial alloys[24] chemically similar to it.

Alloy designation	Chemical composition (wt.%)							
	Zn	**Mg**	**Cu**	**Zr**	**Ti**	**Fe**	**Si**	**Y**
7010	5.7–6.7	2.1–2.6	1.5–2.0	0.10–0.16	0.06	0.15	0.12	—
7050A	5.7–6.9	1.7–2.6	1.7–2.4	0.05–0.12	0.06	0.15	0.12	—
7150	5.9–6.9	2.0–2.7	1.9–2.5	0.08–0.15	0.06	0.15	0.12	—
7278A	6.4–7.4	2.3–3.2	1.3–2.1	0.05–0.25	0.05	0.15	0.12	—
The optimized alloy	6.49	2.52	1.92	0.25	0.07	0.10	0.03	0.29

Notes: The contents of Ti, Fe, and Si in the commercial alloys correspond to the maximum acceptable values.

We can find in Table 3 that the optimized alloy falls precisely within the composition windows of commercial 7010, 7050A, 7150, and 7278A alloys apart from Zr, Ti, and Y elements. However, the optimized alloy shows a much higher UTS than the commonly reported UTS of these commercial

alloys. In particular, a commercial 7050-T6 alloy (Al–6.2Zn–2.1Mg–2.2Cu–0.1Zr–0.03Ti–0.06Fe–0.02Si, wt.%) was tested in our laboratory with a UTS of only 651 MPa under the same test conditions. Moreover, the optimized alloy has been found even stronger than some high-Zn-Mg-content alloys prepared by RS/PM approaches
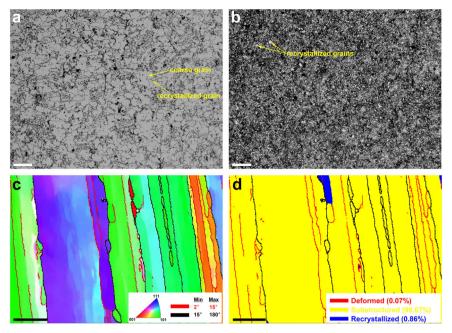
Fig. 6 OM images (cross section) and EBSD maps (longitudinal section) of the optimized alloy. a As-quenched microstructure after a double-stage solution treatment of (450 °C/1 h + 470 °C/2 h). **b** As-aged microstructure after the double-stage solution treatment of (450 °C/1 h + 470 °C/2 h) and an ageing treatment of 150 °C/2 h. The large white patches represent recrystallized grains. **c** EBSD orientation map in the DT6 condition. The insert is the IPF color code. **d** EBSD recrystallization distribution map in the DT6 condition. Scale bars, 100 μm in **a** and **b**, 20 μm in **c** and **d**.

(alloys #1, 3, and 4 in Table 2). What could the high strength of the optimized alloy be ascribed to except for the ageing strengthening? Let us focus on the effects of the combined addition of Zr, Ti, and Y.

As observed in Fig. 4, the combined addition of Zr, Ti, Y forms primary Al_3Zr, $Al_3(Zr,Ti)$, $Al_{20}Ti_2Y$, and Al_8Cu_4Y phases. Our original intention of adding 0.25 wt.% Zr into the optimized was (π) to refine the as-cast microstructure by dispersively forming small Al_3Zr or $Al_3(Zr,Ti)$ nucleants and (θ) to inhibit the recrystallization by introducing the $L1_2$-Al_3Zr or $L1_2$-$Al_3(Zr,Ti)$ dispersoids. Unfortunately, some coarse Zr-containing primary phases were formed due to the insufficient stirring during casting, as well as the high Zr content (Fig. 4d, e). In this particular case, we believe that the grain refinement and recrystallization inhibition effects of Zr element are reduced greatly. Hence, the 0.25 wt.% Zr addition in the optimized alloy contributed little to the high strength. The coarse Al_3Zr, $Al_3(Zr,Ti)$, and $Al_{20}Ti_2Y$ (polygon-shaped) particles, as potential crack initiators, can also serve as particle-stimulated nucleation[38] (PSN) sites for recrystallization of the α-Al matrix, resulting in soft fine-grained regions around them (Fig. 4b, c). In addition, the coarse $Al_{20}Ti_2Y$ particles could trap Zn and Mg atoms (Fig. 4e), which may further make the fine-grained regions soft due to the lack of matrix precipitates. The stress relaxation brought by the fine-grained regions can partly weaken the adverse effects of these coarse particles. As verified by ref. [39], we emphasize that the unwanted coarse particles in the optimized alloy have a very limited detrimental effect on UTS. The tiny Al_8Cu_4Y and $Al_{20}Ti_2Y$ particles (Fig. 5d, e) on or near (sub-)grain boundaries can stabilize substructures and inhibit the growth of recrystallized grains. Consequently, the optimized alloy shows a very low degree of recrystallization (independent of the ageing treatment) (Fig. 6a–d). In the as-quenched sample (where the sub-gains are difficult to reveal due to the absence of η-$MgZn_2$ precipitates on their boundaries), the α-Al matrix is bimodally grained with fine (<10 μm) recrystallized grains at the periphery of the coarse (>30 μm) deformed grains. A high-temperature ageing treatment (150 °C/2 h) was applied to fully reveal sub-gains and the result

are shown in Fig. 6b. The optimized alloy contains a great number of fine (<5 μm) sub-grains and a few recrystallized grains with slightly larger sizes. The observed high-density dislocations (Fig. 5a, d) and sub-grains (Figs. 5b and 6b) indicate the considerable substructure strengthening and dislocation strengthening effects in the optimized alloy. Meanwhile, the tiny particles themselves can also improve the UTS by dispersion strengthening. Last but not least, we strongly believe that the Al_8Cu_4Y nanoscale network structure also contributed to the high strength of the optimized alloy.

The satisfactory iteration results we gained here have demonstrated the high efficiency of machine learning in searching for 7xxx alloys with desired strength. The maximum "expected improvement" value in iteration 4 ($EI^4_{max} = 26.822$, see Table 1) is much >1% of the current-highest value of the UTS—a recommended convergence value of "expected improvement"[34], indicating that the remaining optimization space is still plenty. On the other hand, the choice of alloying elements and their ranges can be adjusted flexibly according to the predetermined preparation process. For instance, if we aim to optimize 7xxx alloys that are planned to be fabricated using advanced techniques such as spray forming and electromagnetic casting, the limits of the main alloying elements can be extended significantly. Thus, we believe that the UTS of 7xxx alloys can be further improved using machine-learning methods, provided the following strategies are adopted.

First, it is worthy to prepare a high-quality data set. The data set used in this study, although scattered enough, is less than ideal because the input alloy compositions were not sampled in ways the machine-learning model prefers. Therefore, the input compositions should be sampled in the most ideal way, such as Latin hypercube sampling[40,41]. The scale of the data set is also essential. It should not be too large, as collecting experimental data is time-consuming and expensive. Also, it should not be too small, to avoid inducing too many iterations (which is frustratingly slow). We suggest using a medium-size data set containing $n = 2d + 1$ (d is the dimension of composition space) initial points for design

problems of 7xxx alloys on the condition of a nicely-sized search space.

Second, we should evaluate the effects of alloy composition on the UTS fully. For simplicity and efficiency, at the stage of composition optimization the cast billets were only solution-treated and aged to screen out 7xxx alloys with a good ageing strengthening effect. That is, we ignored homogenization treatment and plastic deformation (e.g., extrusion, rolling, and forging), which are necessary ingredients in the industrial production of 7xxx alloys. Consequently, the effects of alloy composition were not evaluated thoroughly. Especially, the discrepancy in recrystallization behavior of different 7xxx alloys was ignored, which can significantly affect the UTS. Hence, a complete processing pathway is strongly recommended.

Last, we need to choose the most suitable machine-learning model. Although previous research suggested that the adaptive design was very forgiving to the quality of the machine-learning model or regressor[26], there could be other machine-learning models superior to the Kriging model against composition optimization of 7xxx alloys, given the famous "no-free-lunch theorems"[42]. Therefore, concurrent consideration of multiple models (e.g., Kriging, support vector regression, backpropagation neural network) is also necessary for further improvements.

Conclusions

Through the development of nearly a century, 7xxx alloys have been demanded higher in mechanical performance. These alloys have traditionally been developed by screening various compositions and heat treatments based on trial-and-error. They are today being designed with higher contents and more species of alloying elements to improve mechanical properties. Often, advanced forming techniques and new-type heat treatment schedules are introduced to process highly-alloyed alloys. In this paper, however, we find that very high tensile strength can be achieved in a dilute 7xxx alloy based on traditional processing techniques. Additionally, we demonstrate the feasibility of using machine learning to accelerate the discovery of 7xxx alloys and the potential strengthening effects of Y element. Especially, we show the unusual Al_8Cu_4Y nanoscale network structure, which may be of great value to alloy design of high-strength aluminum alloys.

For the optimized alloy, further composition or process adjustment against Zr and Ti is required to eliminate the coarse constituent particles. As inspired by the findings presented here, it can be expected that multi-objective composition optimization promotes comprehensive properties of 7xxx alloys. Moreover, the effects of Y on microstructure and mechanical properties of 7xxx alloys need further investigating to give more insights, and comparing with Sc. More in-depth microstructural studies are necessary to know more about the Al_8Cu_4Y network structure. Since the formation of the Al_8Cu_4Y network structure is intimately related to Al_8Cu_4Y particles, it may be meaningful to control the proportion of the two kinds of Al_8Cu_4Y phase for tailoring different properties. Results regarding these topics will be the subjects of our future publications.

Methods

Materials and processing. The experimental alloys were prepared by melting raw materials including 99.9% pure Al, Zn, and Mg ingots, and master alloys (wt. %) of Al–50Cu, Al–10Zr, Al–10Ti, Al–20Y, and Al–20Ce in a clay-bonded graphite crucible using a 7.5 KW well resistance furnace. The melts were poured into a zinc oxide (ZnO)-coated iron mold, preheated at ~200 °C, after a three-minute manual stirring, removal of slag, and a ten-minute heat preservation at ~720 °C. Two molds with inner dimensions of 120 × 40 × 20 mm and Φ95 mm were used to fabricate 1 kg slabs for composition optimization and a 6 kg cylindrical ingot for extrusion, respectively. The slabs were successively solution-

treated at a relatively low temperature of 465 °C for 4 h to avoid overburning, water quenched at room-temperature, and immediately RRA-treated (120 °C/24 h + 190 °C/30 min + 120 °C/24 h) after quenching. The cylindrical ingot first underwent a two-step homogenization treatment (400 °C/12 h + 465 °C/24 h), followed by furnace cooling. The homogenized ingot was then machined into Φ90 mm for scales removal and extruded at ~420 °C with an extrusion ratio of ~32 and a ram speed of 0.4 mm/s. The solution and artificial ageing treatments of the extrusion rods were performed in a vacuum furnace and a silicone oil-bath furnace, respectively.

Microstructural characterization. The actual chemical composition of the optimized alloy was calibrated through an inductively coupled plasma-atomic emission spectroscopy (ICP-AES; 7300DV). Phase constitution was determined using an X-ray diffractometer (XRD; Empyrean) with Cu Kα radiation in the 2-theta span of 15–85°. Metallographic specimens were polished to a mirror finish, then immersed in Graff Sargent etchant[43] (1 mL HF, 16 mL HNO₃, 3 g CrO₃, and 83 mL H₂O) and observed by an optical microscope (OM; Model Zeiss Lab A1). Specimens for scanning electron microscope (SEM; JSM 7800F) analysis were unetched. Thin foils for transmission electron microscope (TEM; FEI Tecnai G2 F30, coupled with a high-angle annular dark-field detector) examination were mechanically polished to ~60 μm and further jet-polished at the temperature between −20 °C and −25 °C in an electrolyte solution of 25 vol.% HNO₃ and 75 vol.% CH₃OH. The electrolyte solution was also used to electro-polish the electron backscatter diffraction (EBSD) specimen (5 × 5 × 7mm) at ~10 V and at the temperature of less than −30 °C. The step size and acceleration voltage used for EBSD measurements were 0.15 μm and 15 kV, respectively. Kikuchi patterns was identified with the final success rate of 90.2%. EBSD data were analyzed by HKL Channel 5 software. Image-Pro plus software was used to measure the size of second phases and α-Al grains.

Mechanical property testing. The tensile tests were conducted on a computerized universal material testing machine (CMT5105) at room-temperature with a deformation rate of 1 mm/min. Tensile specimens, 15 mm in gauge length and 3 mm in gauge diameter, were selected according to the ISO 6892-1:2016 standard. The specimens (with axial directions along the length of the slaps) were machined from the same regions of the slaps at the stage of composition optimization. For the specimens machined from the extrusion rods, the axial direction aligned with the extrusion direction. In each condition, at least three specimens were tested in parallel and their average value was adopted.

Modification and implementation of EGO algorithm. We chose the Kriging surrogate model to make predictions at unsynthesized alloys in the search space, with associated predictive uncertainty provided synchronously. We constructed the Kriging model on a DACE[44] (Design and Analysis of Computer Experiments) toolbox, in which we selected the zero-order polynomial and gauss as the regression and correlation functions, respectively, as is the case in most engineering design optimization[45]. Apart from the data set, the determination of the correlation parameter θ is also essential for constructing the Kriging model[46]. Unfortunately, DACE simply uses the pattern search method (Hooke and Jeeves method[47]) to optimize θ, which renders the optimum result intimately dependent on the starting point and may finally achieve a suboptimal Kriging model[48]. In this study, we used a genetic algorithm to surmount this shortcoming. Every time we constructed or reconstructed the Kriging model, θ was optimized beforehand except for cross-validation, during which we used an identical θ calculated from all the observations we had then (Supplementary Table 3). The key concept of the EGO algorithm is to use "expected improvement" (EI) [equation (1)] as a figure of merit to balance global and local search, and thus to determine new sampling points reasonably[34].

$$\mathrm{EI}(\text{expected improvement}) = (y - y_{\max})\Phi\left(\frac{y - y_{\max}}{\sigma}\right) + \sigma\varphi\left(\frac{y - y_{\max}}{\sigma}\right), \quad (1)$$

where y is the predicted UTS, $y_{\max}$ is the maximum value of the UTSs in the training data set, and σ is the predictive uncertainty denoted by the root mean squared error. Φ and φ are the standard normal cumulant distribution and probability density functions, respectively. Considering the time-consuming and expensive essence of the iterative process (including new materials syntheses and characterizations) and the possibility of genetic algorithm to fall into local optimum, at each iteration we greedily chose three sampling points (EGO usually selects only one data point by maximizing "expected improvement"[34]) with different local optimum values of "expected improvement". Note that the compositional difference between the three points should be big enough to be guaranteed during casting. Similarly, we used genetic algorithm to calculate "expected improvement". Additional details about genetic algorithm calculations are summarized in Supplementary Table 4. Finally, instead of using the "1% stopping criterion"[34] alone, we considered an additional stopping criterion—stop iteration once the targeted UTS of 550 MPa is reached.

Data availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

Code availability

References

1. Williams, J. C. & Starke, E. A. Jr. Progress in structural materials for aerospace systems. *Acta Mater.* **51**, 5775–5799 (2003).
2. Tang, J., Zhang, H., Teng, J., Fu, D. & Jiang, F. Effect of Zn content on the static softening behavior and kinetics of Al–Zn–Mg–Cu alloys during double-stage hot deformation. *J. Alloys Compd.* **806**, 1081–1096 (2019).
3. Sokoluk, M., Cao, C., Pan, S. & Li, X. Nanoparticle-enabled phase control for arc welding of unweldable aluminum alloy 7075. *Nat. Commun.* **10**, 98 (2019).
4. Sharma, M. M. Microstructural and mechanical characterization of various modified 7XXX series spray formed alloys. *Mater. Charact.* **59**, 91–99 (2008).
5. Liddicoat, P. V. et al. Nanostructural hierarchy increases the strength of aluminium alloys. *Nat. Commun.* **1**, 1–7 (2010).
6. Osamura, K. et al. Development of high-strength aluminum alloys by mesoscopic structure control. *Metall. Mater. Trans. A* **26**, 1597–1599 (1995).
7. Chen, K., Liu, H., Zhang, Z., Li, S. & Todd, R. I. The improvement of constituent dissolution and mechanical properties of 7055 aluminum alloy by stepped heat treatments. *J. Mater. Process. Technol.* **142**, 190–196 (2003).
8. Ning, A., Liu, Z., Peng, B. & Zeng, S. Redistribution and re-precipitation of solute atom during retrogression and reaging of Al-Zn-Mg-Cu alloys. *Trans. Nonferrous Met. Soc. China* **17**, 1005–1011 (2007).
9. Ditta, A., Wei, L., Xu, Y. & Wu, S. Effect of hot extrusion and optimal solution treatment on microstructure and properties of spray-formed Al-11.3Zn-2.65Mg-1Cu alloy. *J. Alloys Compd.* **797**, 558–565 (2019).
10. Li, H., Cao, F., Shu, G., Ning, Z. & Sun, J. Microstructures and properties evolution of spray-deposited Al-Zn-Mg-Cu-Zr alloys with scandium addition. *J. Alloys Compd.* **691**, 482–488 (2016).
11. Valiev, R. Z., Enikeev, N. A., Murashkin, M. Y., Kazykhanov, V. U. & Sauvage, X. On the origin of the extremely high strength of ultrafine-grained Al alloys produced by severe plastic deformation. *Scr. Mater.* **63**, 949–952 (2010).
12. Chen, Z., Mo, Y. & Nie, Z. Effect of Zn content on the microstructure and properties of super-high strength Al-Zn-Mg-Cu alloys. *Metall. Mater. Trans. A* **44**, 3910–3920 (2013).
13. Dong, P., Chen, S. & Chen, K. Effects of Cu content on microstructure and properties of super-high-strength Al-9.3Zn-2.4Mg-xCu-Zr alloy. *J. Alloys Compd.* **788**, 329–337 (2019).
14. Shu, W. X. et al. Tailored Mg and Cu contents affecting the microstructures and mechanical properties of high-strength Al-Zn-Mg-Cu alloys. *Mater. Sci. Eng. A* **657**, 269–283 (2016).
15. Li, Y., Zhang, Z. R., Zhao, Z. Y., Li, H. & Zhuang, L. Effect of main elements (Zn, Mg, and Cu) on hot tearing susceptibility during direct-chill casting of 7xxx aluminum alloys. *Metall. Mater. Trans. A* **50**, 3603–3616 (2019).
16. Ying, D. et al. Effects of Sc and Zr microalloying additions and aging time at 120 °C on the corrosion behaviour of an Al-Zn-Mg alloy. *Corros. Sci.* **65**, 288–298 (2012).
17. Xiao, Y. P., Pan, Q. L., Li, W. B., Liu, X. Y. & He, Y. B. Influence of retrogression and re-aging treatment on corrosion behaviour of an Al–Zn–Mg–Cu alloy. *Mater. Des.* **32**, 2149–2156 (2011).
18. Qiu, D., Taylor, J. A. & Zhang, M. X. Understanding the co-poisoning effect of Zr and Ti on the grain refinement of cast aluminum alloys. *Metall. Mater. Trans. A* **41**, 3412–3421 (2010).
19. Gao, T., Zhang, Y. & Liu, X. Influence of trace Ti on the microstructure, age hardening behavior and mechanical properties of an Al-Zn-Mg-Cu-Zr alloy. *Mater. Sci. Eng. A* **598**, 293–298 (2014).
20. Liu, J. et al. Effect of minor Sc and Zr on recrystallization behavior and mechanical properties of novel Al-Zn-Mg-Cu alloys. *J. Alloys Compd.* **657**, 717–725 (2016).
21. Zhang, X. G., Mei, F. Q., Zhang, H. Y., Wang, S. H. & Hao, H. Effects of Gd and Y additions on microstructure and properties of Al–Zn–Mg–Cu–Zr alloys. *Mater. Sci. Eng. A* **552**, 230–235 (2012).
22. Chen, K. H., Fang, H. C., Zhang, Z., Chen, X. & Liu, G. Effect of Yb, Cr and Zr additions on recrystallization and corrosion resistance of Al–Zn–Mg–Cu alloys. *Mater. Sci. Eng. A* **497**, 426–431 (2008).
23. Yu, X. X. et al. Solidification behavior and elimination of undissolved Al2CuMg phase during homogenization in Ce-modified Al-Zn-Mg-Cu alloy. *Rare Met.* **39**, 1279–1287 (2020).
24. Association, A. *International Alloy Designations and Chemical Composition Limits for Wrought Aluminum and Wrought Aluminum Alloys 10-12* (Aluminum Association, 2009).
25. Xue, D. et al. Accelerated search for materials with targeted properties by adaptive design. *Nat. Commun.* **7**, 11241 (2016).
26. Balachandran, P. V., Xue, D., Theiler, J., Hogden, J. & Lookman, T. Adaptive strategies for materials design using uncertainties. *Sci. Rep.* **6**, 19660 (2016).
27. Yuan, R. et al. Accelerated discovery of large electrostrains in BaTiO3-based piezoelectrics using active learning. *Adv. Mater.* **30**, 1702884 (2018).
28. Zhang, Z. et al. Accelerated discoveries of mechanical properties of graphene using machine learning and high-throughput computation. *Carbon* **148** 115–123 (2019).
29. Wen, C. et al. Machine learning assisted design of high entropy alloys with desired property. *Acta Mater.* **170**, 109–117 (2019).
30. Zhang, Y. et al. Unsupervised discovery of solid-state lithium ion conductors. *Nat. Commun.* **10**, 5260 (2019).
31. Wang, C., Fu, H., Jiang, L., Xue, D. & Xie, J. A property-oriented design strategy for high performance copper alloys via machine learning. *NPJ Comput. Mater.* **5**, 87 (2019).
32. Alireza, V. et al. New bayesian-optimization-based design of high-strength 7xxx-series alloys from recycled aluminum. *JOM* **70**, 2704–2709 (2018).
33. Krige, D. G. A statistical approach to some basic mine valuation problems on the witwatersrand. *J. South. Afr. Inst. Min. Metall.* **52**, 119–139 (1951).
34. Jones, D. R., Schonlau, M. & Welch, W. J. Efficient global optimization of expensive black-box functions. *J. Global Optim.* **13**, 455–492 (1998).
35. Wang, Y., Chen, M., Zhou, F. & Ma, E. High tensile ductility in a nanostructured metal. *Nature* **419**, 912–915 (2002).
36. Tian, W., Songmei, L. I., Liu, J., Mei, Y. U. & Yujie, D. U. Preparation of bimodal grain size 7075 aluminum alloys and their corrosion properties. *Chin. J. Aeronaut.* **30**, 1777–1788 (2017).
37. Chinh, N. Q., Lendvai, J., Ping, D. H. & Hono, K. The effect of Cu on mechanical and precipitation properties of Al-Zn-Mg alloys. *J. Alloys Compd.* **378**, 52–60 (2004).
38. Engler, O., Kong, X. W. & Yang, P. Influence of particle stimulated nucleation on the recrystallization textures in cold deformed Al-alloys Part I—Experimental observations. *Scr. Mater.* **37**, 1665–1674 (1997).
39. Hu, G., Chen, S., Jiang, H., Wang, P. & Chen, K. Effect of RE Ce on microstructure and properties of 7A52 aluminum alloy. *Trans. Nonferrous Met. Soc. China* **26**, 1372–1382 (2016).
40. Suzuki, A., Shen, C. & Chennimalai Kumar, N. Application of computational tools in alloy design. *MRS Bull.* **44**, 247–251 (2019).
41. Jin, R., Chen, W. & Sudjianto, A. An efficient algorithm for constructing optimal design of computer experiments. *J. Stat. Plan. Infer.* **134**, 268–287 (2005).
42. Wolpert, D. H. & Macready, W. G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1**, 67–82 (1997).
43. Graff, W. R. & Sargent, D. C. A new grain-boundary etchant for aluminum alloys. *Metallography* **14**, 69–72 (1981).
44. Lophaven, S. N., Nielsen, H. B. & Sondergaard, J. *DACE—A Matlab Kriging Toolbox, Version 2.0. Report No. IMM-REP-2002-12* (Technical University of Denmark, 2002).
45. Simpson, T. W., Maurey, T. M., Korte, J. J. & Mistree, F. Kriging metamodels for global approximation in simulation-based multidisciplinary design optimization. *AIAA J.* **39**, 2233–2241 (2001).
46. Martin, J. D. & Simpson, T. W. Use of kriging models to approximate deterministic computer models. *AIAA J.* **43**, 853–863 (2005).
47. Hooke, R. & Jeeves, T. A. "Direct search" solution of numerical and statistical problems. *J. Assoc. Comput. Mach.* **8**, 212–229 (1961).
48. Xiong, X., Li, S. & Wu, F. Robust parameter design for nonlinear signal–response systems using kriging models. *Eng. Optim.* **4**, 1–18 (2019).

Acknowledgements

Author contributions

J.H.L. and Y.B.Z. conceived the experimental scheme. J.H.L., Y.B.Z., and X.Y.C. carried out the experiments concerning alloy preparations and discussed how to modify and use the EGO algorithm for optimizing 7xxx Al alloys. J.H.L. and Y.Z. performed the tensile tests. J.H.L. prepared and characterized the extrusion rods with the help of Q.Z. and X.Y.Q. J.H.L., Y.B.Z., and H.C. analyzed the results and wrote the first version of the manuscript. All authors contributed to the improvement and finalization of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s43246-020-00074-2.

Correspondence and requests for materials should be addressed to Y.Z.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.