Supplementary information

An interpretable learning framework for exploring superelastic degradation of NiTi shape memory alloys using multimodal data

Yuejie Hu[1], Haiyang Wang[1], Ming Chen[4,5], Gang Chen[1,2], Chuanjie Wang[1,2,*], Bin Guo[2], M.W. Fu[3,*]

[1] School of Materials Science and Engineering, Harbin Institute of Technology at Weihai, Weihai 264209, China

[2] Key Laboratory of Micro-Systems and Micro-Structures Manufacturing (Harbin Institute of Technology), Ministry of Education, Harbin 150080, China

[3] Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

[4] State Key Laboratory of Metal Matrix Composites, Shanghai Jiao Tong University, Shanghai 200240, China

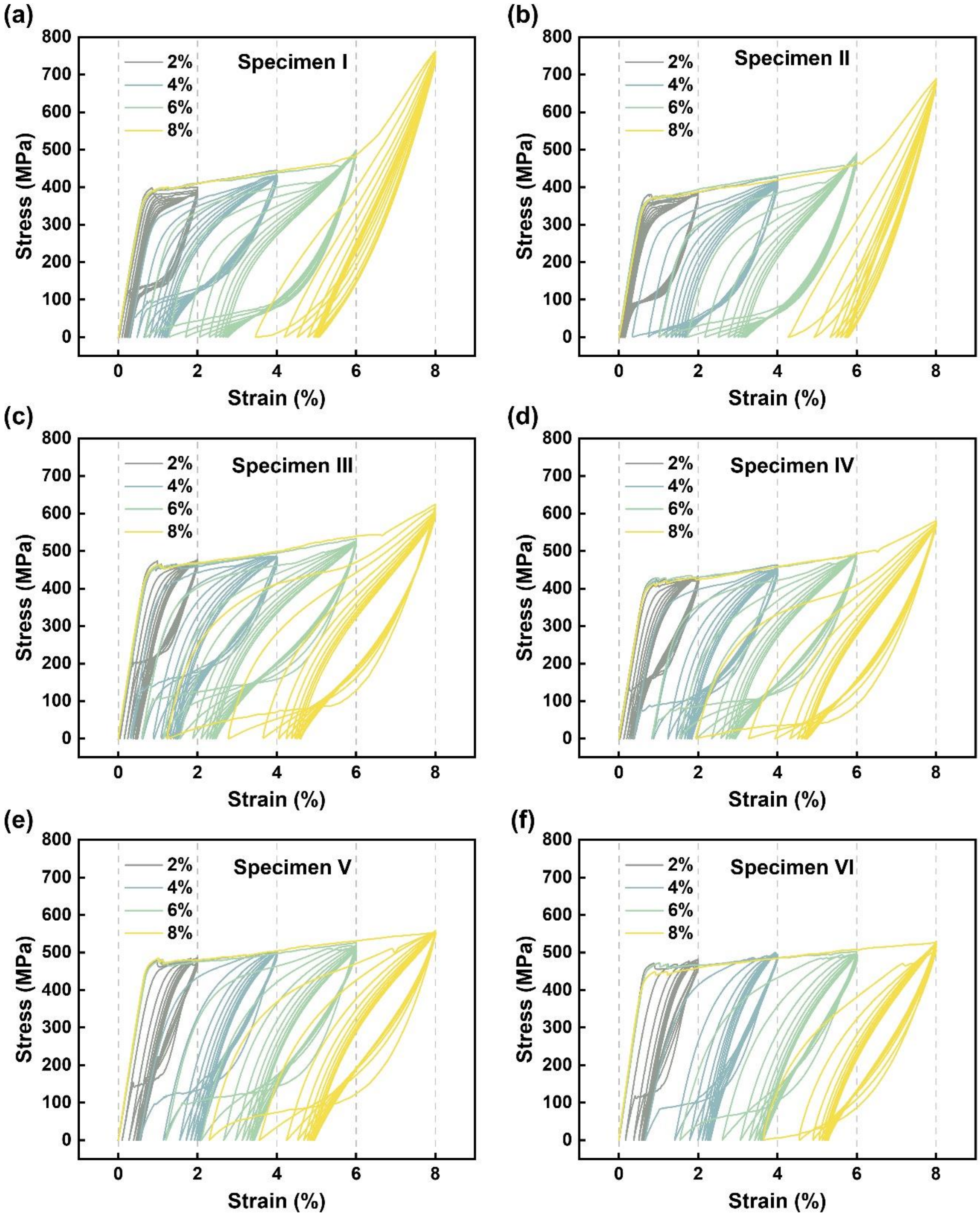
[5] School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

[6] These authors contributed equally: Chuanjie Wang, M.W. Fu

[*]Correspondence authors. E-mail: cjwang@hitwh.edu.cn; mmmwfu@polyu.edu.hk

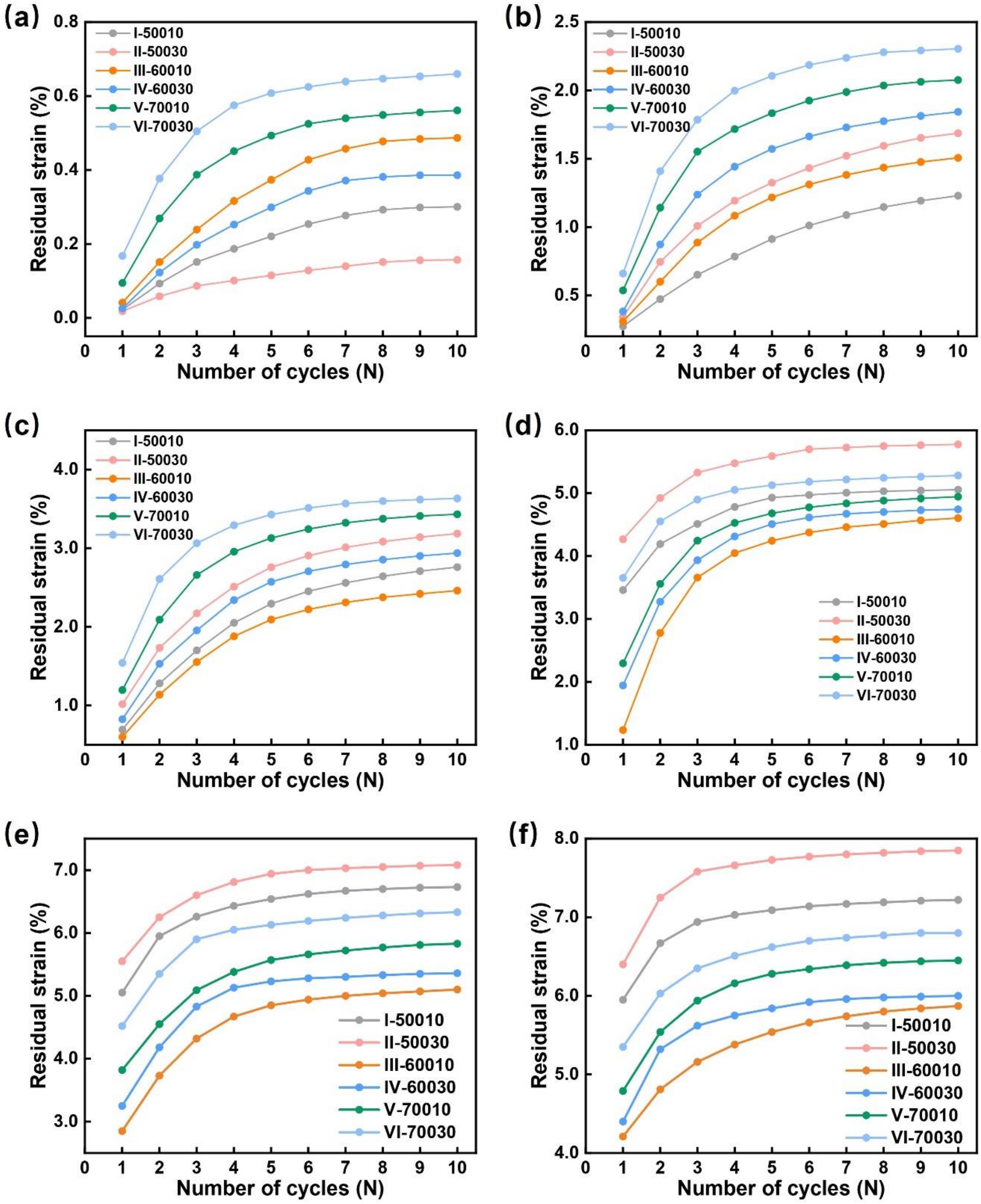
Experimental Section

Cyclic tensile–unloading engineering stress–strain curves



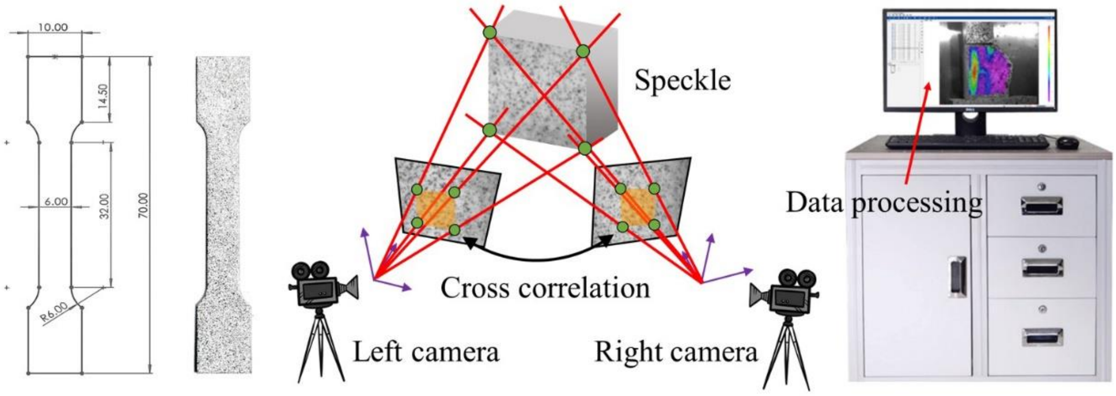
Supplementary Fig. 1 | Cyclic tensile–unloading engineering stress–strain curves of NiTi alloys under different annealing conditions. (**a**–**f**) correspond to specimens I–VI, respectively, showing the cyclic test results under strain amplitudes of 2%, 4%, 6%, and 8%.

Residual strain statistics

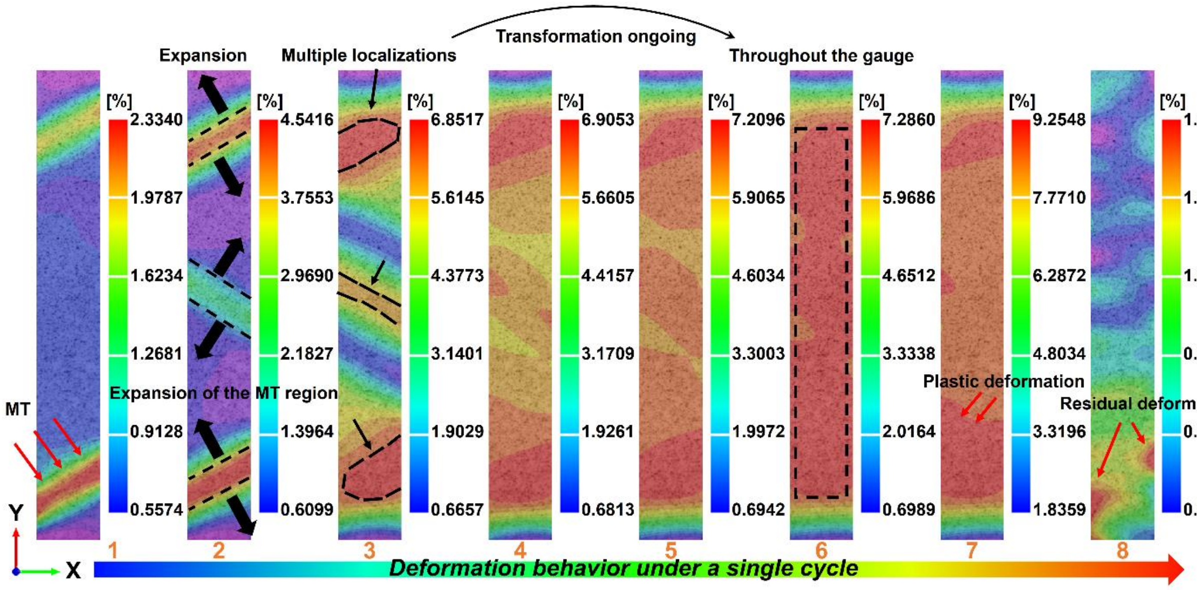


Supplementary Fig. 2 | Residual strain accumulation of NiTi alloys under various annealing conditions. Evolution of residual strain with increasing cycle number at strain amplitudes of (**a**) 2%, (**b**) 4%, (**c**) 6%, (**d**) 8%, (**e**) 10% and (**f**) 12%. The first four amplitudes were employed for model training and validation. The latter two were used to evaluate extensibility, where the material had fully entered the martensitic plastic deformation period.

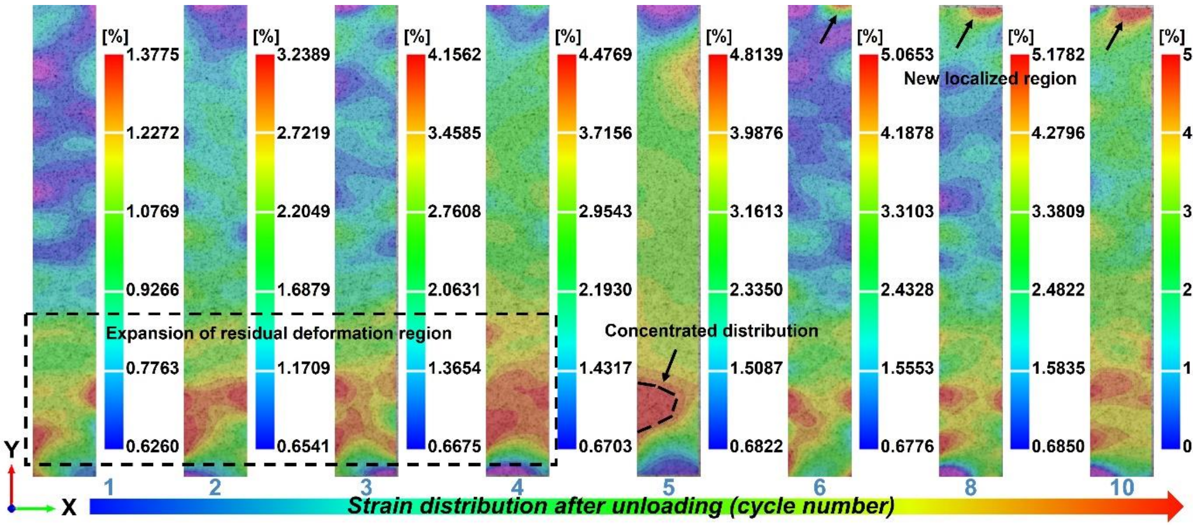
Digital image correlation experiment



Supplementary Fig. 3 | Schematic of the digital image correlation (DIC) setup used to measure full-field strain during cyclic tensile loading. Dog bone specimens with a gauge section of 25 mm × 6 mm were sprayed with a random speckle pattern and recorded by a stereo DIC system comprising two 5-megapixel CCD industrial cameras equipped with 35-mm lenses. Image pairs were continuously acquired and processed to obtain the three-dimensional displacement and strain fields. The target strain amplitudes for the subsequent cyclic tensile unloading tests were first calibrated from monotonic uniaxial tensile experiments using the same DIC system.

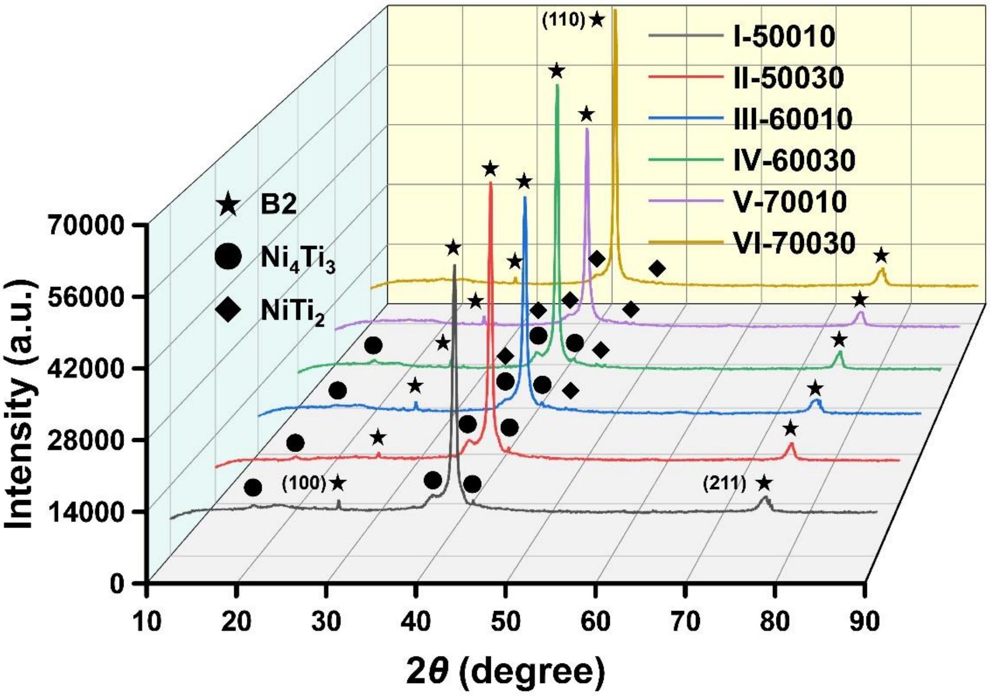


Supplementary Fig. 4 | First cycle deformation analysis. DIC maps 1–8 mark key deformation stages and correspond to the symbols on the stress strain curve in Fig. 5b. At 2% strain, the specimen has entered the stress plateau and the initial martensitic transformation is activated. The transformation evolves from local nucleation to multiple localization and eventually to a fully transformed gauge section. Notably, the end of the plateau at 6.66% strain does not coincide with the completion of transformation: by 6.82% the gauge section is fully transformed, and at 8% strain martensitic plastic deformation emerges.



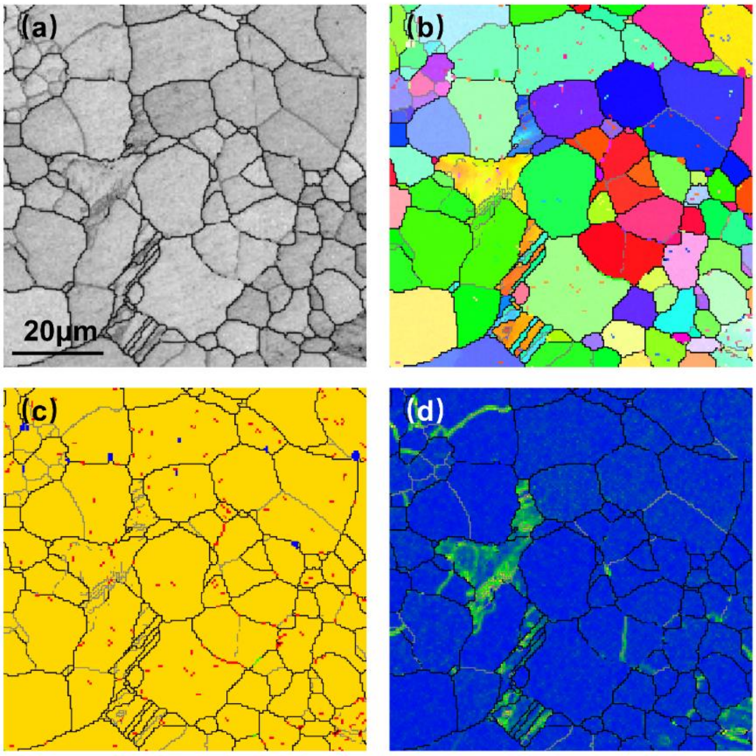
Supplementary Fig. 5 | Cyclic deformation analysis. DIC maps for successive cycles show the local deformation of the specimen at the end of each loading unloading loop, corresponding to the blue markers on the stress strain curve in Fig. 5b and visualizing the progressive accumulation of residual strain. As shown in the figure, the residual deformation region continuously expands during the first five loading cycles, corresponding to the observed orientation-dependent variant transfer phenomenon in the main text.

X-ray diffraction pattern



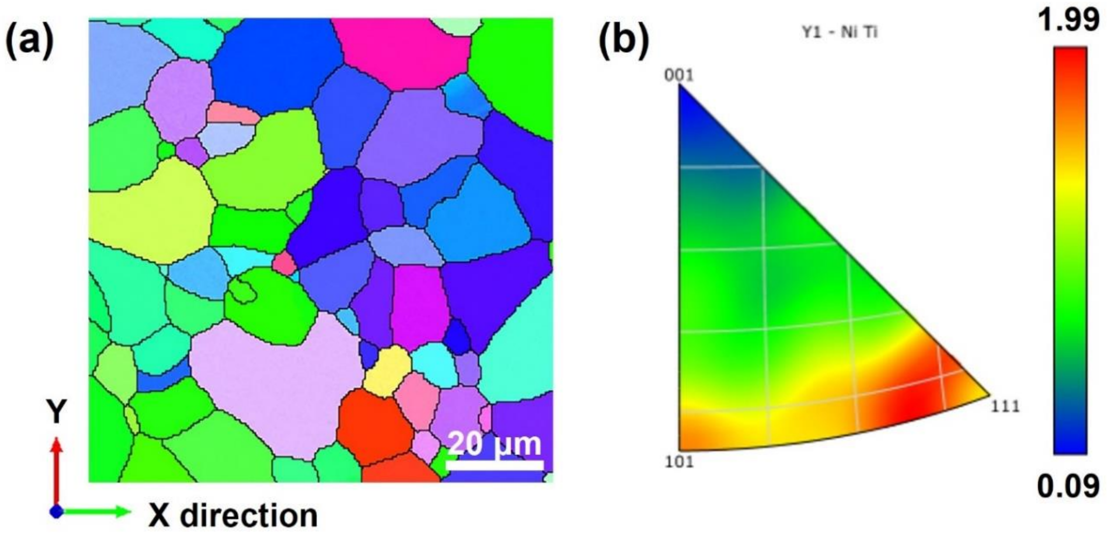
Supplementary Fig. 6 | XRD patterns of NiTi alloys annealed under various short-term annealing conditions. Specimens I–VI correspond to distinct annealing regimes: I–50010 (500 °C, 10 min), II–50030 (500 °C, 30 min), III–60010 (600 °C, 10 min), IV–60030 (600 °C, 30 min), V–70010 (700 °C, 10 min), and VI–70030 (700 °C, 30 min). All specimens exhibit a B2 austenitic matrix at room temperature, accompanied by different precipitate species. With increasing annealing temperature, the fine Ni_4Ti_3 precipitates dominant at 500 °C evolve into a multiphase microstructure containing both Ni_4Ti_3 and $NiTi_2$ at 600 °C, and $NiTi_2$ becomes the prevailing phase at 700 °C. Prolonging the holding time mainly promotes the coarsening and redistribution of precipitates without altering their types.

Electron backscatter diffraction: The EBSD results from Specimen IV are presented as representative. The microstructure exhibits a multiphase configuration comprising the B2 matrix (yellow) together with uniformly dispersed Ni_4Ti_3 (red) and $NiTi_2$ (blue) precipitates. Due to cold rolling deformation prior to annealing, residual internal stresses were not completely relieved, resulting in a pronounced {101} texture orientation. Nevertheless, the overall structure remains predominantly uniform equiaxed grains with an average size of approximately 10.3 μm. In the phase image, $NiTi_2$ precipitates are generally identifiable, while Ni_4Ti_3 can only be effectively resolved by EBSD when sufficiently large (>200 nm). EBSD struggles to fully analyze smaller Ni_4Ti_3 precipitates. This is why we incorporated TEM results into the numerical feature to achieve a more comprehensive microstructural characterization. Furthermore, KMA analysis reveals the presence of certain orientation gradients within the grains, which is also related to the retention of residual stresses from cold working.



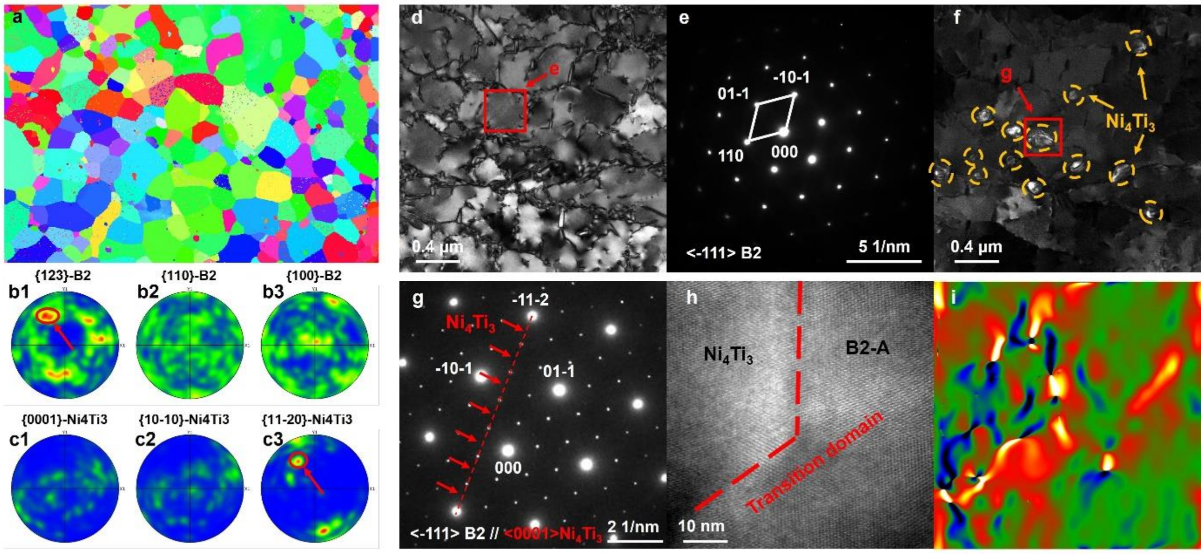
Supplementary Fig. 7 | EBSD characterization of specimen IV annealed at 600 °C for 30 min. a Band contrast image. **b** Inverse pole figure (IPF) map showing an equiaxed grain structure. **c** Phase map indicating that the matrix is primarily B2 austenite. **d** Kernel average misorientation (KAM) map revealing the local strain distribution and orientation gradients.

Pole figure of the selected region



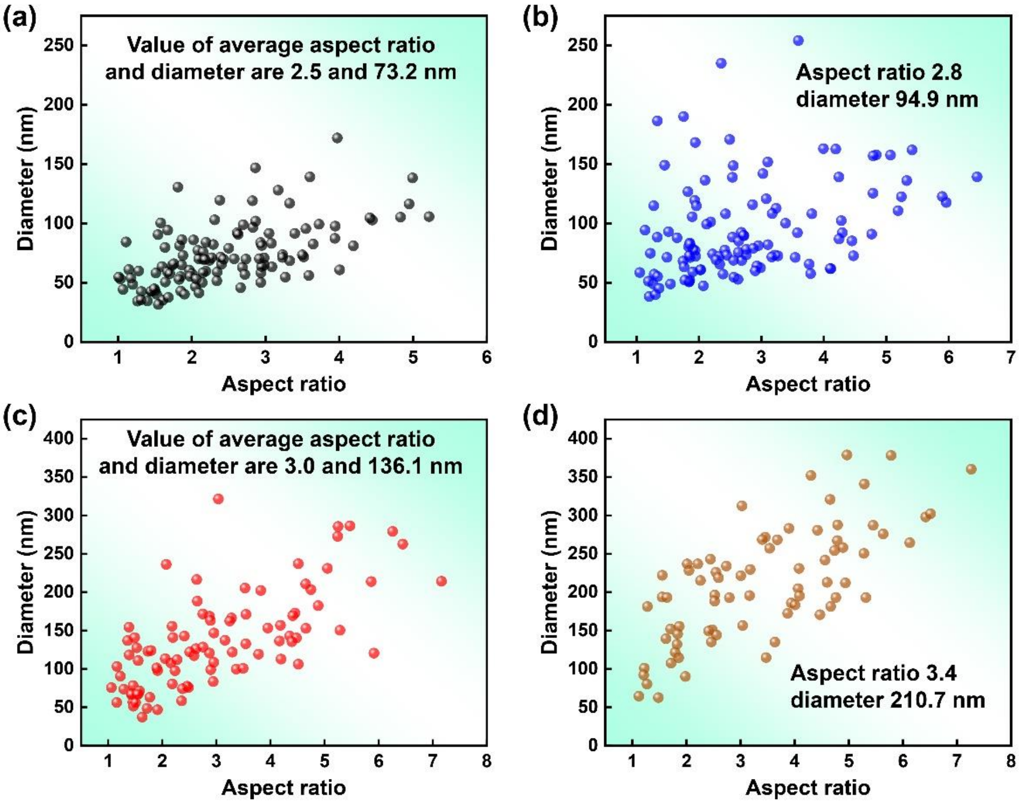
Supplementary Fig. 8 | EBSD results of the region selected for grain orientation analysis. **a** Inverse pole figure (IPF) map of the NiTi alloy subjected to heat treatment at 600 °C for 10 min. The Y direction of the EBSD map is parallel to the tensile axis. The microstructure consists of equiaxed B2 austenite grains. **b** The corresponding IPF along the Y direction shows a relatively strong <111>$_{B2}$ fiber texture in the B2 grains, with a multiple of uniform distribution (MUD) ≈ 1.99, a texture commonly observed in NiTi sheets[1].

Orientation analysis of Ni_4Ti_3 precipitates



Supplementary Fig. 9 | Orientation relationship between Ni_4Ti_3 precipitates and the B2 austenite matrix. **a** Orientation distribution map. **b_1–b_3** {123}, {110}, and {100} pole figures of the B2 austenite matrix. **c_1–c_3** {0001}, {10-10}, and {11-20} pole figures of the Ni_4Ti_3 precipitates. **d** Bright-field TEM image of the overall region, and (**e**) the selected area electron diffraction (SAED) pattern taken from the red boxed area in (**d**). **f** Corresponding dark field image, with (**g**) and (**h**) showing the SAED pattern and high-resolution TEM (HRTEM) image from the red boxed region in (**f**), respectively. **i** Geometric phase analysis[2] (GPA) strain map corresponding to (h). No common-index crystal plane was identified between Ni_4Ti_3 and the B2 matrix in the pole figures. Nevertheless, the pole figure and SAED analyses reveal a specific orientation relationship of $\{123\}\langle 111\rangle_{B2}$ // $\{11\text{-}20\}\langle 0001\rangle_{Ni_4Ti_3}$. Dark field image demonstrates that Ni_4Ti_3 precipitates are uniformly distributed as nanoscale particles. HRTEM and GPA strain mapping indicate that the formation of Ni_4Ti_3 precipitates induces strain inhomogeneity in the NiTi matrix, with significant local tensile strain in continuous distribution regions.

Statistical information of Ni_4Ti_3 precipitates



Supplementary Fig. 10 | Quantitative statistics of Ni_4Ti_3 precipitates under different annealing conditions. a–d Aspect ratio–diameter distributions of Ni_4Ti_3 precipitates in Specimens I–IV, corresponding to 500 °C × 10 min, 500 °C × 30 min, 600 °C × 10 min and 600 °C × 30 min, respectively. Each symbol denotes one precipitate, and the insets give the mean aspect ratio and equivalent diameter. From Specimen I to IV, the data clouds shift toward larger diameters and higher aspect ratios, and the distributions gradually broaden, indicating progressive coarsening and slight elongation of Ni_4Ti_3. The increase in average diameter and aspect ratio is much more pronounced when the annealing temperature rises from 500 °C to 600 °C than when the holding time is extended from 10 min to 30 min, showing that temperature exhibits a stronger influence on the precipitation behavior than time.

Calculation and correction of the Schmitt factor: For multiphase NiTi alloys, traditional Schmid theory struggles to accurately predict critical transformation stresses. The primary limitation arises from the localized stress field introduced by precipitates, which modify the effective resolved shear stress distribution within grains. As a result, the actual slip and transformation behaviors depart from idealized crystallographic expectations. Previous studies indicate that the coherency mismatch stress of Ni_4Ti_3 precipitates can elevate the critical stress for dislocation slip while simultaneously lowering the onset stress required to martensitic transformation[3-5]. With Ni_4Ti_3 precipitates distributed nearly uniformly and the analysis centered on preferentially oriented grains that control superelastic degradation, the use of this correction is well-justified and broadly applicable. In principle, the Ni_4Ti_3 precipitates larger than ~300 nm, once they lose coherency with the matrix, should contribute only weakly to the internal stress field[6], similar to $NiTi_2$ particles. However, TEM observations (**Supplementary Fig. 10**) show that such oversized Ni_4Ti_3 particles are extremely scarce, so the overall adjustment is still formulated by treating the Ni_4Ti_3 population as effectively coherent. The specific derivation is as follows. We first assume that dislocation slip or SIMT is triggered when the total resolved shear stress attains a critical threshold τ_{c}, such that:

$$\tau_{\text{rss-total}} = m \cdot \sigma_{\text{app}} + \tau_{\text{rss-i}} \geq \tau_{\text{c}} \tag{1}$$

Where $\tau_{\text{rss-total}}$ denotes the total critical resolved shear stress. *m* represents the traditional Schmid factor, σ_{app} corresponds to the nominal stress under applied load, and $\tau_{\text{rss-i}}$ signifies the local resolved shear stress introduced by the coherent stress field of the Ni_4Ti_3 precipitates. To analyze the activation behavior in different regions, the above relationship can be equivalently expressed as a modified Schmid factor:

$$m_{\text{eff}} \cdot \sigma_{\text{app}} = m \cdot \sigma_{\text{app}} + \tau_{\text{rss-i}} \tag{2}$$

$$m_{\text{eff}} = m + \frac{\tau_{\text{rss-i}}}{\sigma_{\text{app}}} \tag{3}$$

Here, m_{eff} denotes the corrected Schmid factor, which integrates the combined effects of the external load and the local coherency stress field. For different deformation mechanisms, the corrected form can be expressed as follows:

$$\begin{cases} m_{\text{eff,slip}} = m_{\text{slip}} - \dfrac{\tau_{\text{rss-i,slip}}}{\sigma_{\text{app,slip}}} \\ m_{\text{eff,trans}} = m_{\text{trans}} - \dfrac{\tau_{\text{rss-i,trans}}}{\sigma_{\text{app,trans}}} \end{cases} \tag{4}$$

Among these, m_{slip} and m_{trans} represent the conventional Schmid factors for slip and transformation, respectively. $\tau_{\text{rss-i,slip}}$ and $\tau_{\text{rss-i,trans}}$ denote the local shear stress corrections for the precipitates, with the former corresponding to the slip inhibition effect and the latter to the transformation promotion effect. $\sigma_{\text{app,slip}}$ and $\sigma_{\text{app,trans}}$ represent the applied stress, respectively. For transformation, $\sigma_{\text{app,trans}}$ employs the stress corresponding to the first tensile transformation plateau. To quantify the local stress fields generated by precipitates, the Eshelby approach is employed：

$$\tau_{\text{rss-i,trans}} = \alpha'_{ij} C_{ijkl} D_{klmn}(x) \varepsilon^{\text{p}}_{mn} \tag{5}$$

Here, α'_{ij} is the orientation tensor transformed into the principal coordinate system of the precipitate from α_{ij} :

$$\alpha_{ij} = \frac{1}{2}(m_i n_j + n_i m_j) \tag{6}$$

m and *n* denote the shear direction and the habit plane normal associated with each martensitic twin variant. Each martensite twin variant pair comprises two martensite single-crystal variants V*i* and V*j* connected by a twin relationship, maintaining a specific volume fraction ratio *x*:(1-*x*). Although the interface between the austenite and each individual martensite variant (V*i* or V*j*) is not an invariant plane, the interface between the austenite and the martensite variant pair (V*i*–V*j*) is an invariant habit plane. In this study, we consider the only two types of martensitic twins with invariant habit planes: the <011> type II twin and the {11-1} type I twin, yielding a total of 48 variant pairs[7,8]. For the first type, the crystallographic parameters are *m* = [0.41145, −0.76331, 0.49807] and *n* = [−0.88888, −0.21524, 0.40443]; for the second type, *m* = [−0.82044, 0.02068, 0.57136] and *n* = [0.27789, −0.85894, 0.43012]. The parameters for the other 46 variants can be obtained through cyclic permutation of the indices. **Supplementary Table 1 and Table 2** summarize the theoretical solutions for all habit planes of these two twin types. Additionally, C_{ijkl} represents the elastic constant tensor for the precipitate and the surrounding matrix (assumed to be consistent). $D_{klmn}(x)$ denotes the Eshelby tensor describing the transmission of lattice misfit strain from the precipitate within the matrix[9], characterizing the elastic field response at any position *x*. $\varepsilon^{\text{p}}_{mn}$ represents the intrinsic misfit strain of the Ni_4Ti_3 precipitate, reflecting its spontaneous lattice distortion in the absence of external loading[6].

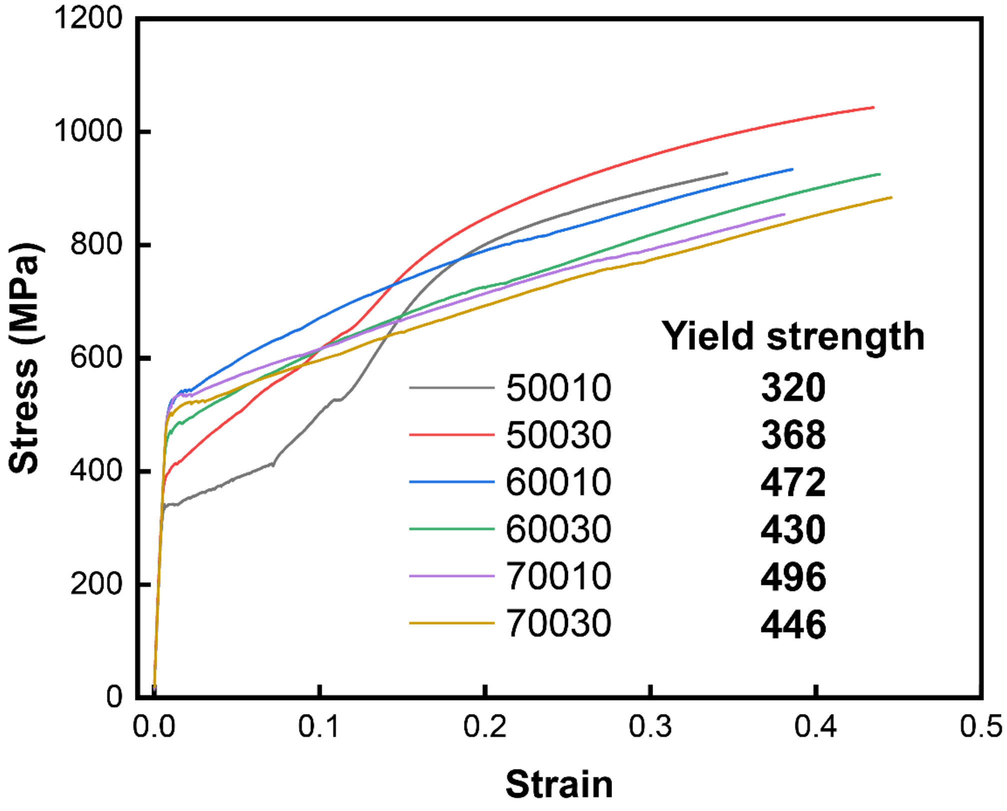
For slip-related correction, the Eshelby's calculation method remains consistent, employing only the crystallographic parameters of the slip system in the orientation projection section. Specifically, in Equations (6), the α_{ij} no longer represents the orientation tensor of the transformation variant, but instead denotes the Schmid orientation tensor $\alpha_{ij} = s_i n_j$, defined by the slip direction s_i and slip plane normal n_j. The most representative slip systems in B2 NiTi, {011}<001> and {011}<111>, are selected. Moreover, the $\sigma_{\mathrm{app,slip}}$ is the conventional yield stress[10], extracted from tensile tests conducted at 200 °C (**Supplementary Fig. 11**), a temperature at which no stress-induced transformation occurs and aging effects are still negligible.

Supplementary Table 1 | Theoretical solutions of the habit plane system for <011> type II twin. The volume fraction ratio of the twinning x:(1-x)=0.72898:0.27102

Variant pairs	Invariant habit plane	Invariant shear direction
1-2	(−0.88888 −0.21524 0.40443)	[0.41145 −0.76331 0.49807]
1-2′	(−0.88888 −0.40443 0.21524)	[0.41145 −0.49807 0.76331]
1'-2	(0.88888 −0.40443 0.21524)	[−0.41145 −0.49807 0.76331]
1'-2′	(0.88888 −0.21524 0.40443)	[−0.41145 −0.76331 0.49807]
2-1	(−0.88888 −0.21524 −0.40443)	[0.41145 −0.76331 −0.49807]
2-1′	(−0.88888 −0.40443 −0.21524)	[0.41145 −0.49807 −0.76331]
2′-1	(0.88888 −0.40443 −0.21524)	[−0.41145 −0.49807 −0.76331]
2′-1′	(0.88888 −0.21524 −0.40443)	[−0.41145 −0.76331 −0.49807]
3-4	(0.40443 −0.88888 −0.21524)	[0.49807 0.41145 −0.76331]
3-4′	(0.21524 −0.88888 −0.40443)	[0.76331 0.41145 −0.49807]
3′-4	(0.21524 0.88888 −0.40443)	[0.76331 −0.41145 −0.49807]
3′-4′	(0.40443 0.88888 −0.21524)	[0.49807 −0.41145 −0.76331]
4-3	(−0.40443 −0.88888 −0.21524)	[−0.49807 0.41145 −0.76331]
4-3′	(−0.21524 −0.88888 −0.40443)	[−0.76331 0.41145 −0.49807]
4′-3	(−0.21524 0.88888 −0.40443)	[−0.76331 −0.41145 −0.49807]
4′-3′	(−0.40443 0.88888 −0.21524)	[−0.49807 −0.41145 −0.76331]
5-6	(−0.21524 0.40443 −0.88888)	[−0.76331 0.49807 0.41145]
5-6′	(−0.40443 0.21524 −0.88888)	[−0.49807 0.76331 0.41145]
5′-6	(−0.40443 0.21524 0.88888)	[−0.49807 0.76331 −0.41145]
5′-6′	(−0.21524 0.40443 0.88888)	[−0.76331 0.49807 −0.41145]
6-5	(−0.21524 −0.40443 −0.88888)	[−0.76331 −0.49807 0.41145]
6-5′	(−0.40443 −0.21524 −0.88888)	[−0.49807 −0.76331 0.41145]
6′-5	(−0.40443 −0.21524 0.88888)	[−0.49807 −0.76331 −0.41145]
6′-5′	(−0.21524 −0.40443 0.88888)	[−0.76331 −0.49807 −0.41145]

Supplementary Table 2 | Theoretical solutions of the habit plane system for {11-1} type I twin. The volume fraction ratio of the twinning x:(1-x)=0.67994:0.32006

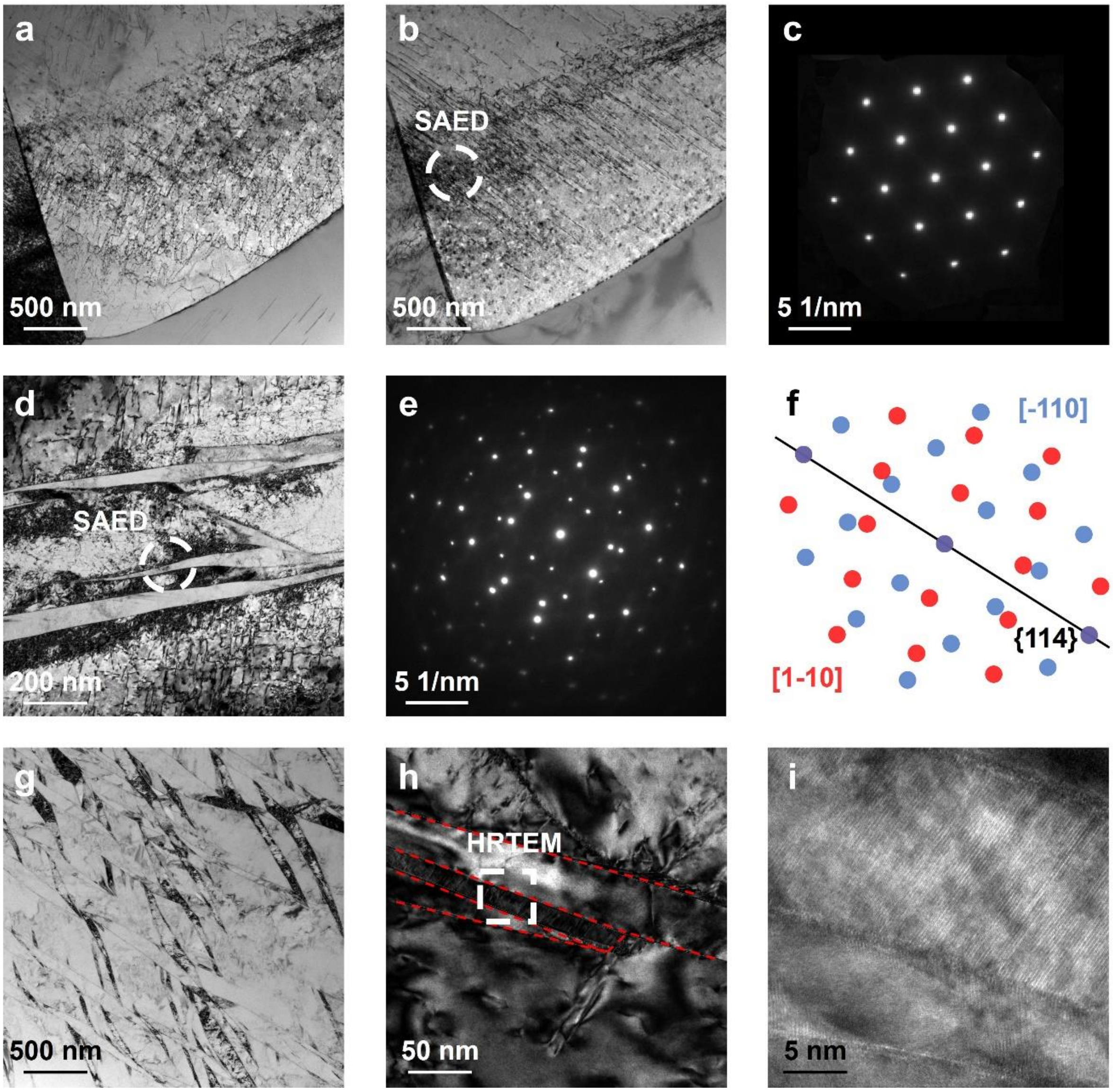
Variant pairs	Invariant habit plane	Invariant shear direction
1–4	(0.27789 −0.85894 0.43012)	[−0.82044 0.02068 0.57136]
1–5′	(0.27789 −0.43012 0.85894)	[−0.82044 −0.57136 −0.02068]
1′–3	(−0.27789 −0.85894 0.43012)	[0.82044 0.02068 0.57136]
1′–6	(−0.27789 −0.43012 0.85894)	[0.82044 −0.57136 −0.02068]
2–3′	(0.27789 −0.85894 −0.43012)	[−0.82044 0.02068 −0.57136]
2–5	(0.27789 −0.43012 −0.85894)	[−0.82044 −0.57136 0.02068]
2′–4′	(−0.27789 −0.85894 −0.43012)	[0.82044 0.02068 −0.57136]
2′–6′	(−0.27789 −0.43012 −0.85894)	[0.82044 −0.57136 0.02068]
3–1′	(0.85894 0.27789 −0.43012)	[−0.02068 −0.82044 −0.57136]
3–6	(0.43012 0.27789 −0.85894)	[0.57136 −0.82044 0.02068]
3′–2	(0.85894 −0.27789 −0.43012)	[−0.02068 0.82044 −0.57136]
3′–5	(0.43012 −0.27789 −0.85894)	[0.57136 0.82044 0.02068]
4–1	(−0.85894 0.27789 −0.43012)	[0.02068 −0.82044 −0.57136]
4–5′	(−0.43012 0.27789 −0.85894)	[−0.57136 −0.82044 0.02068]
4′–2′	(−0.85894 −0.27789 −0.43012)	[0.02068 0.82044 −0.57136]
4′–6′	(−0.43012 −0.27789 −0.85894)	[−0.57136 0.82044 0.02068]
5–2	(−0.85894 0.43012 0.27789)	[0.02068 0.57136 −0.82044]
5–3′	(−0.43012 0.85894 0.27789)	[−0.57136 −0.02068 −0.82044]
5′–1	(−0.85894 0.43012 −0.27789)	[0.02068 0.57136 0.82044]
5′–4	(−0.43012 0.85894 −0.27789)	[−0.57136 −0.02068 0.82044]
6–1′	(−0.85894 −0.43012 0.27789)	[0.02068 −0.57136 −0.82044]
6–3	(−0.43012 −0.85894 0.27789)	[−0.57136 0.02068 −0.82044]
6′–2′	(−0.85894 −0.43012 −0.27789)	[0.02068 −0.57136 0.82044]
6′–4′	(−0.43012 −0.85894 −0.27789)	[−0.57136 0.02068 0.82044]



Supplementary Fig. 11 | Tensile testing at 200 °C is used to determine the critical stress $\sigma_{\text{app,slip}}$ for slip (conventional yield strength). At this temperature, no stress-induced transformation occurs and aging effects are still negligible.

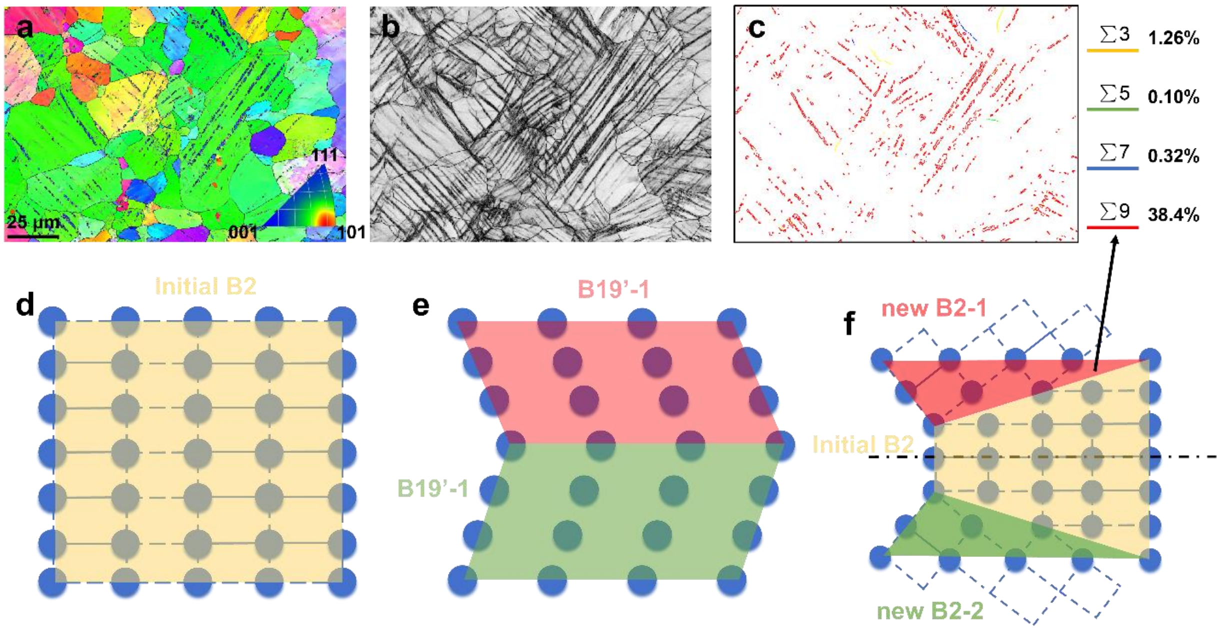
5

Transmission electron microscope characterization after cyclic loading



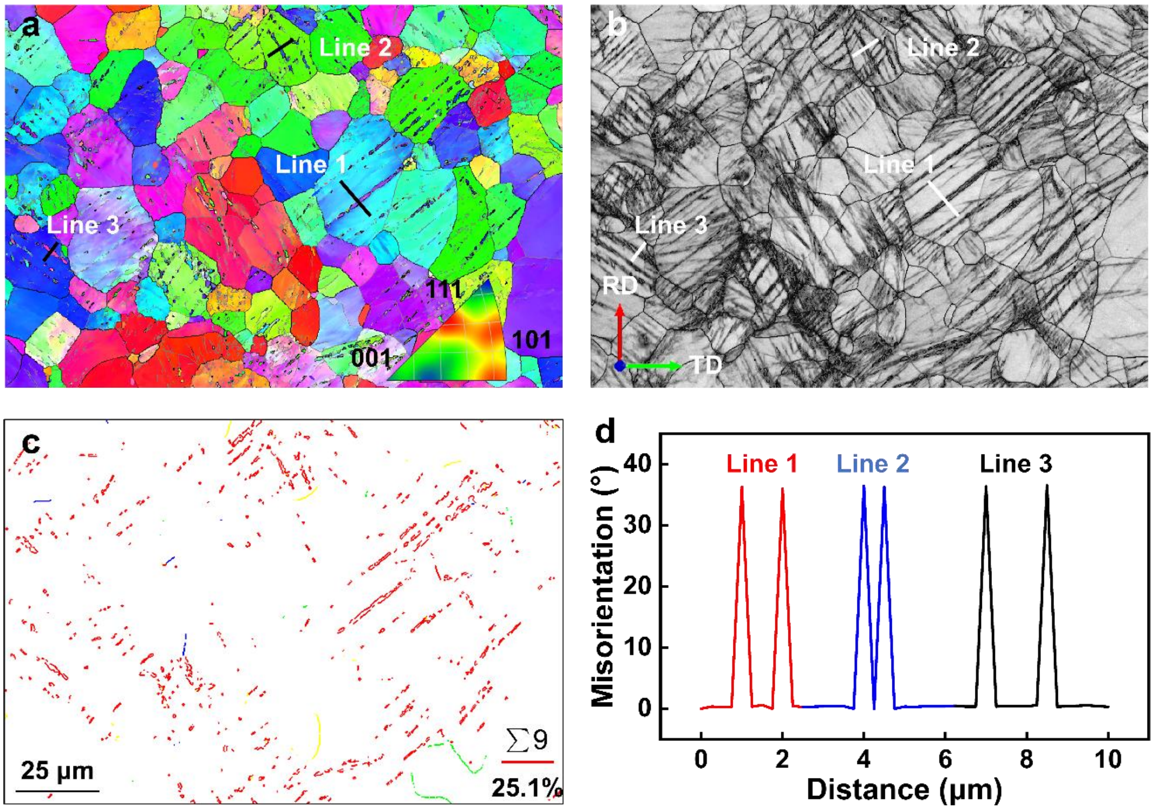
Supplementary Fig. 12 | TEM characterization of NiTi after cyclic loading at different strain amplitudes. a, **b** The bright field images of the specimen cycled at 4% strain amplitude under different diffraction conditions, revealing a high density of dislocations but no discernible austenite twins or retained martensite plates. **c** SAED pattern from the region marked in (**b**). **d**–**f** TEM results for the specimen cycled at 6% strain amplitude, showing the appearance of {114} B2 austenite twins. **g** BF image of the specimen cycled at 8% strain, displaying extensive austenite twins. **h** Local observation of a representative twin band, and (**i**) corresponding HRTEM image from the boxed region in (**h**), where the twin boundaries are clearly resolved. Overall, superelastic degradation of the NiTi alloy under cyclic loading is primarily governed by dislocation accumulation and austenite twinning, with no retained martensite detected over the entire strain amplitude range.

EBSD characterization after cyclic loading (n=5)



Supplementary Fig. 13 | EBSD characterization of the NiTi alloy annealed at 700 °C for 30 min after 5 cycles at 8% strain amplitude. a IPF map showing the development of a pronounced {101} texture after annealing. **b** Band contrast image, where dense deformation bands aligned along specific orientations traverse the grains after high strain cycling. **c** Coincidence site lattice (CSL) boundary map revealing the extensive formation of Σ9 boundaries, reaching 38.4%, indicative of substantial grain boundary reconfiguration during deformation. We provide a schematic illustration of the formation mechanism of Σ9 special boundaries in B2 austenite during cyclic phase transformation. **d** The initial B2 phase (schematized using the {011} plane) exhibits a regular lattice. **e** Cyclic loading induces the formation of a composite B19′ martensite twin structure. **f** Subsequent reverse transformation produces two newly formed B2 regions with a relative lattice rotation that satisfies the Σ9 relation, thereby generating high densities of Σ9 boundaries at the microscale.

EBSD characterization after cyclic loading (n=10)



Supplementary Fig. 14 | EBSD characterization of the NiTi alloy annealed at 700 °C for 30 min after 10 cycles at 8% strain amplitude. a IPF orientation map, with Line 1–3 indicating the paths used for subsequent misorientation profiles. **b** Corresponding BC image on the RD–TD section, showing numerous deformation twin bands traversing the grains along specific orientations after high strain cyclic loading. **c** CSL boundary map, revealing a high density of Σ9 special deformation boundaries with an area fraction of 25.1%. **d** Intragrain misorientation profiles extracted along Line 1–3 in (a, b) show the mismatch angle of approximately 36° when passing through the grain boundary, confirming the existence of deformation twins.

Machine learning framework construction

Dataset Description: During dataset construction, EBSD maps obtained under various annealing conditions were segmented and augmented offline, then paired with the corresponding Ni_4Ti_3 descriptor, cycle number, and strain amplitude, yielding 31,536 labeled sequences. The dataset was divided into training, validation, and test sets in a ratio of 8:1:1. The test set contained only original data to ensure independent evaluation. Images were partitioned by field of view to avoid splitting the same region across subsets.

Framework Description: In this study, pronounced modality gaps and semantic misalignment exist between microstructural features and loading conditions. Handling cross-modal entanglement and semantic inconsistency remains a central challenge for effective multimodal feature fusion. For clarity, we take KAM maps and Euler angle matrices (EAM) as a representative pair of modalities. Most existing approaches[11] simply concatenate these features and then perform feature extraction, which can be expressed as:

$$P = h_{\mathrm{pred}}(R \left| F_{KAM}(B_{KAM}), F_{EAM}(B_{EAM}) \right.) \tag{7}$$

Here, $F_{KAM}(B_{KAM})$ and $F_{EAM}(B_{EAM})$ represent KAM and EAM features, respectively. [] denotes the concatenation operation. *R* denotes the extracted feature representation, which is subsequently fed into the prediction head h_{pred} to obtain the prediction result *P*. Although concatenating multimodal features followed by unified feature extraction simplifies the modeling process, this strategy overlooks the latent inconsistencies between modalities, leaving the intrinsic conflicts in complex tasks unresolved. On the other hand, Wu et al.[12] explicitly accounts for such cross-modal inconsistencies by decomposing the feature encoding process into multiple coordinated subtasks, which can be formulated as:

$$P = h_{\mathrm{pred}}([R_{KAM} \left| F_{KAM}(B_{KAM}), R_{EAM} \right| F_{EAM}(B_{EAM})]) \tag{8}$$

Among them, R_{KAM} and R_{EAM} represent the KAM and EAM feature representations, respectively. Apparently, this method only segregates feature exraction through two independent networks, while failing to explore and utilize their mutual cooperation. Distinct from above methods, the construction of our model accommodates the disentanglement and modulation of signal, which is enlightened by cognitive neuroscience[13]. From the perspective of feature learning, it can be expressed as:

$$\begin{cases} f_{KAM} = \varepsilon_{KAM}(R_{KAM} \left| F_{KAM}(B_{KAM}), R_{EAM} \right| F_{EAM}(B_{EAM})) \\ f_{EAM} = \varepsilon_{EAM}(R_{EAM} \left| F_{EAM}(B_{EAM}), R_{KAM} \right| F_{KAM}(B_{KAM})) \\ P = h_{\mathrm{pred}}(\phi(f_{KAM}, f_{EAM})) \end{cases} \tag{9}$$

Here, ε and ϕ denote the feature alignment and mergence operations, respectively. Specifically, we first employ two backbone networks to extract KAM and EAM features independently, as illustrated in **Fig. 1** of the main text. Since features learned by different backbones may exhibit semantic misalignment, we develop a feature synergy module to align and merge these disentangled representations. Note that the design inspiration of this module derives from visual center that focuses on integrating multipartite information, but the specific mechanism is still unknown. Therefore, the method proposed in this paper emphasizes on drawing lessons from its cogitation, while the technical details are incompletely corresponding.

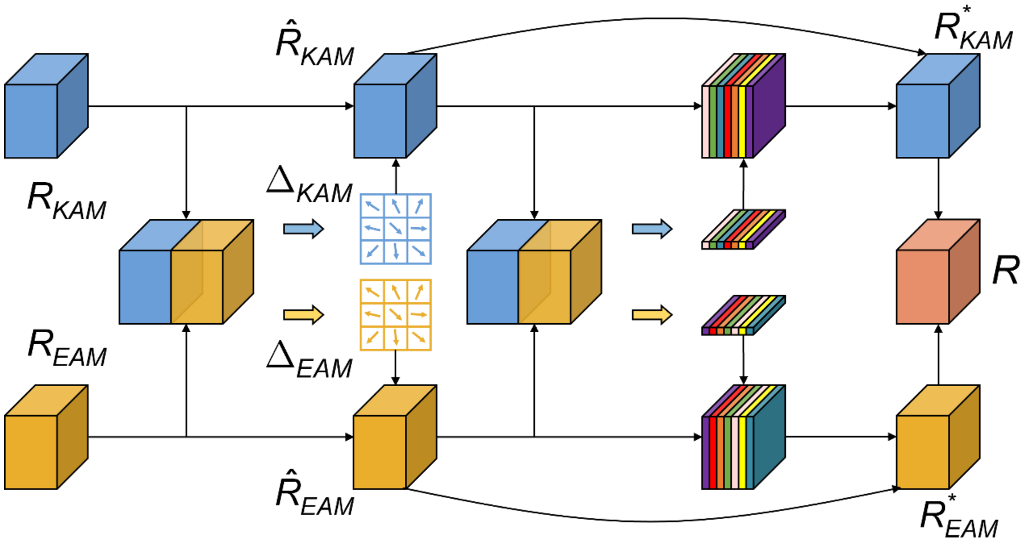
Feature synergy module: We designed a feature synergy module to mitigate the contextual misalignment between modalities. This module comprises feature alignment and mergence, which are detailed in this section. As illustrated in **Supplementary Fig. 15** and Eq. (10), the alignment step primarily involves estimating an offset and applying it to the corresponding feature representations:

$$\begin{cases} \Delta_i = f_0^i([R_{KAM} | F_{KAM}(B_{KAM}), R_{EAM} | F_{EAM}(B_{EAM})]),\ i \in \{KAM, EAM\} \\ \hat{F}_i = f_1^i(R_i | F_i(B_i), \Delta_i)\ i \in \{KAM, EAM\} \end{cases} \tag{10}$$

where f_0 and f_1 represent functions that learn coordinate offsets of the features and align the features with offsets, respectively. In this work, they are implemented using 3×3 deformable convolution, activation and standard convolutions with the same kernel size. In the feature mergence stage, we emphasize the significant feature maps which contains a large amount of spatial information, while suppressing redundant information to achieve precise allocation. The process is shown in Equation (11):

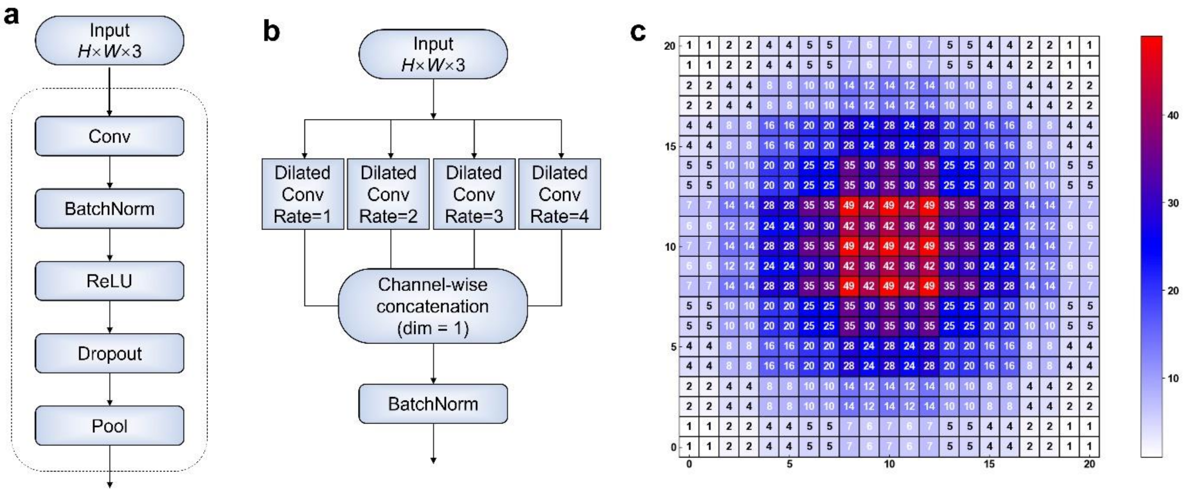
$$\begin{cases} u = \mathrm{GAP}([\hat{R}_{KAM}, \hat{R}_{EAM}]) \\ R_i^* = \hat{R}_i + \sigma(\mathrm{W}_i \otimes u) \odot \hat{R}_i, i \in \{KAM, EAM\} \\ R = R_{KAM}^* \oplus R_{EAM}^* \end{cases} \tag{11}$$

where *u* is obtained by global average pooling after the concat of $\hat{R}_{KAM}$ and $\hat{R}_{EAM}$. σ represents the Sigmoid function, and W_i refers to the 1×1 convolution kernel. Besides, we add residual connections to prevent any response of particular channel from being overly amplified or suppressed. Ultimately, the obtained feature *R* is used for prediction.



Supplementary Fig. 15 | Illustration of the feature synergy module consisting of both feature alignment (*i.e.*, ε_{KAM} and ε_{EAM}) and feature mergence operations (*i.e.*, ϕ). Refer to the text description for more details.

Deep learning model structure: Each branch uses the same feature extraction backbone. The input tensor of size $H \times W \times 3$ is first passed sequentially through five stacked ConvBlocks. Each ConvBlock consists of a 2D convolution, batch normalization, ReLU, dropout, and a pooling layer; downsampling pooling is enabled only in selected blocks to moderately compress spatial resolution and expand the receptive field while controlling network depth, and batch normalization together with dropout helps suppress overfitting. A DilConvBlock is then introduced to reinforce multi-scale context modeling: this module applies four parallel 3×3 dilated convolutions with dilation rates of 1, 2, 3, and 4 to the same feature map, concatenates their outputs along the channel dimension (dim = 1). The output of the DilConvBlock is compressed into a 1D vector by global adaptive average pooling and mapped to three scalar features through several fully connected layers. Applying this pipeline to the Euler, KAM, and phase branches yields three sets of matrix features, which are combined with numerical features and fed into the feature synergy module of the main network.



20	1	1	2	2	4	4	5	5	7	6	7	6	7	5	5	4	4	2	2	1	1
	1	1	2	2	4	4	5	5	7	6	7	6	7	5	5	4	4	2	2	1	1
	2	2	4	4	8	8	10	10	14	12	14	12	14	10	10	8	8	4	4	2	2
	2	2	4	4	8	8	10	10	14	12	14	12	14	10	10	8	8	4	4	2	2
	4	4	8	8	16	16	20	20	28	24	28	24	28	20	20	16	16	8	8	4	4
15	4	4	8	8	16	16	20	20	28	24	28	24	28	20	20	16	16	8	8	4	4
	5	5	10	10	20	20	25	25	35	30	35	30	35	25	25	20	20	10	10	5	5
	5	5	10	10	20	20	25	25	35	30	35	30	35	25	25	20	20	10	10	5	5
	7	7	14	14	28	28	35	35	49	42	49	42	49	35	35	28	28	14	14	7	7
	6	6	12	12	24	24	30	30	42	36	42	36	42	30	30	24	24	12	12	6	6
10	7	7	14	14	28	28	35	35	49	42	49	42	49	35	35	28	28	14	14	7	7
	6	6	12	12	24	24	30	30	42	36	42	36	42	30	30	24	24	12	12	6	6
	7	7	14	14	28	28	35	35	49	42	49	42	49	35	35	28	28	14	14	7	7
	5	5	10	10	20	20	25	25	35	30	35	30	35	25	25	20	20	10	10	5	5
	5	5	10	10	20	20	25	25	35	30	35	30	35	25	25	20	20	10	10	5	5
5	4	4	8	8	16	16	20	20	28	24	28	24	28	20	20	16	16	8	8	4	4
	4	4	8	8	16	16	20	20	28	24	28	24	28	20	20	16	16	8	8	4	4
	2	2	4	4	8	8	10	10	14	12	14	12	14	10	10	8	8	4	4	2	2
	2	2	4	4	8	8	10	10	14	12	14	12	14	10	10	8	8	4	4	2	2
	1	1	2	2	4	4	5	5	7	6	7	6	7	5	5	4	4	2	2	1	1
0	1	1	2	2	4	4	5	5	7	6	7	6	7	5	5	4	4	2	2	1	1
	0					5					10					15					20

Supplementary Fig. 16 | Schematic of the single branch network architecture. a Structure of the ConvBlock, consisting of convolution, batch normalizationn, ReLU, dropout, and pooling layers in sequence. **b** Structure of the DilConvBlock, comprising four parallel 3×3 dilated convolutions with dilation rates of 1, 2, 3, and 4; their outputs are concatenated along the channel dimension (dim = 1) and followed by batch normalization. **c** Illustration of the sampling pattern and effective receptive field of the dilated convolutions, showing the number of times each spatial location is used.

Hyper-parameter setting

Supplementary Table 3 | Hyper-parameters and their descriptions

Hyper-parameter	Description
Data configuration	
num_images=3	Number of input images, the matrix is treated as similar
num_parameters=3	Number of input numerical parameters
num_outputs=1	Number of predicted outputs
image_size=336	Input image size
Model configuration	
backbone	Custom CNN backbone; optional pretrained ResNet-18/34/50
num_blocks=5	Number of convolution blocks
feature_dim=512	Feature dimension after convolution
num_feature=3	Number of features extracted from a single input image
dropout=0.5	Dropout rate for regularization
ReLU	Activation function
Training configuration	
batch size = 32	Number of samples used in each iteration
epochs = 150	Number of model iterations
Loss = MSE	Loss function
Adam	Adaptive moment estimation optimizer
LR = 0.001	Initial learning rate
WD=0.0001	Weight decay
Cosine	Learning rate scheduler
early_stopping=10	Early stopping patience

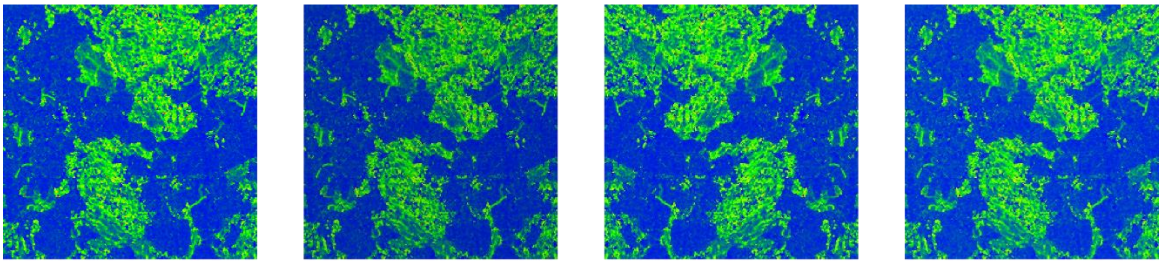
Parameter configuration

Supplementary Table 4 | Initialization parameters of different ML models. For each comparative network (MCC-Net, ECMMF-Net, Base-w/oSM, MBNI-Net), we conducted a small grid search over the Adam hyperparameters on the validation set, while keeping all other optimizer settings at their default values. The final configuration for each model is summarized

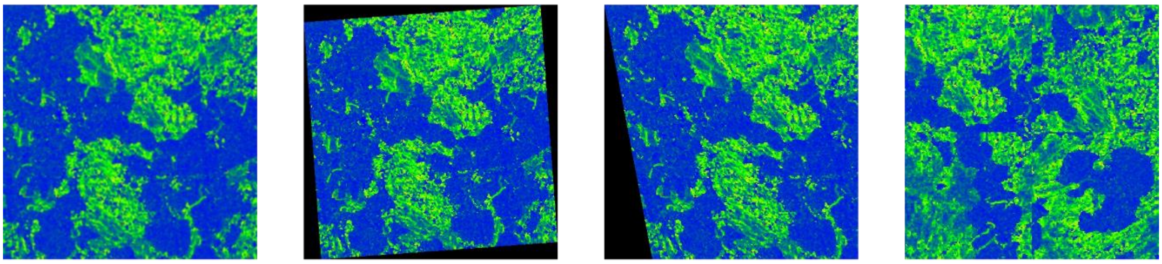
Model	Batch size	Optimizer	Loss function	Initial LR	Weight decay
MCC-Net	32	Adam	MSE	1.0×10^{-3}	1.0×10^{-4}
ECMMF-Net	32	Adam	MSE	5.0×10^{-4}	2.0×10^{-4}
Base-w/oSM	32	Adam	MSE	1.5×10^{-3}	5.0×10^{-5}
MBNI-Net	32	Adam	MSE	5.0×10^{-4}	2.0×10^{-4}
Ours	32	Adam	MSE	1.0×10^{-3}	1.0×10^{-4}

Data augmentation: Given the limited size of the dataset in this study, we adopted offline data augmentation to expand the effective sample size. Images were preprocessed using the Imgaug Python library, and the augmented samples were then fed into the network for training. Seven augmentation strategies were used, including geometric transforms (horizontal flipping, small-angle random rotation of approximately ±10°–15°, and shear affine transformation) and photometric perturbations (Gaussian blur with Gaussian noise, random brightness adjustment, and contrast/gamma perturbation), as well as Mosaic augmentation, in which four images were randomly cropped and stitched into composite scenes.

Horizontal flipping, small-angle rotation, and shear primarily enhance the robustness to variations in orientation and pose, whereas blur+noise, brightness, and contrast/Gamma perturbations enrich the distribution of imaging conditions and appearance. Mosaic augmentation further increases the diversity of spatial compositions and structural patterns[14]. Together, these seven strategies provide a balanced and semantically consistent expansion of the small-scale dataset, as illustrated in **Supplementary Fig. 17**.



(a) Original (b) Brightness (c) Horizontal flip (d) Contrast



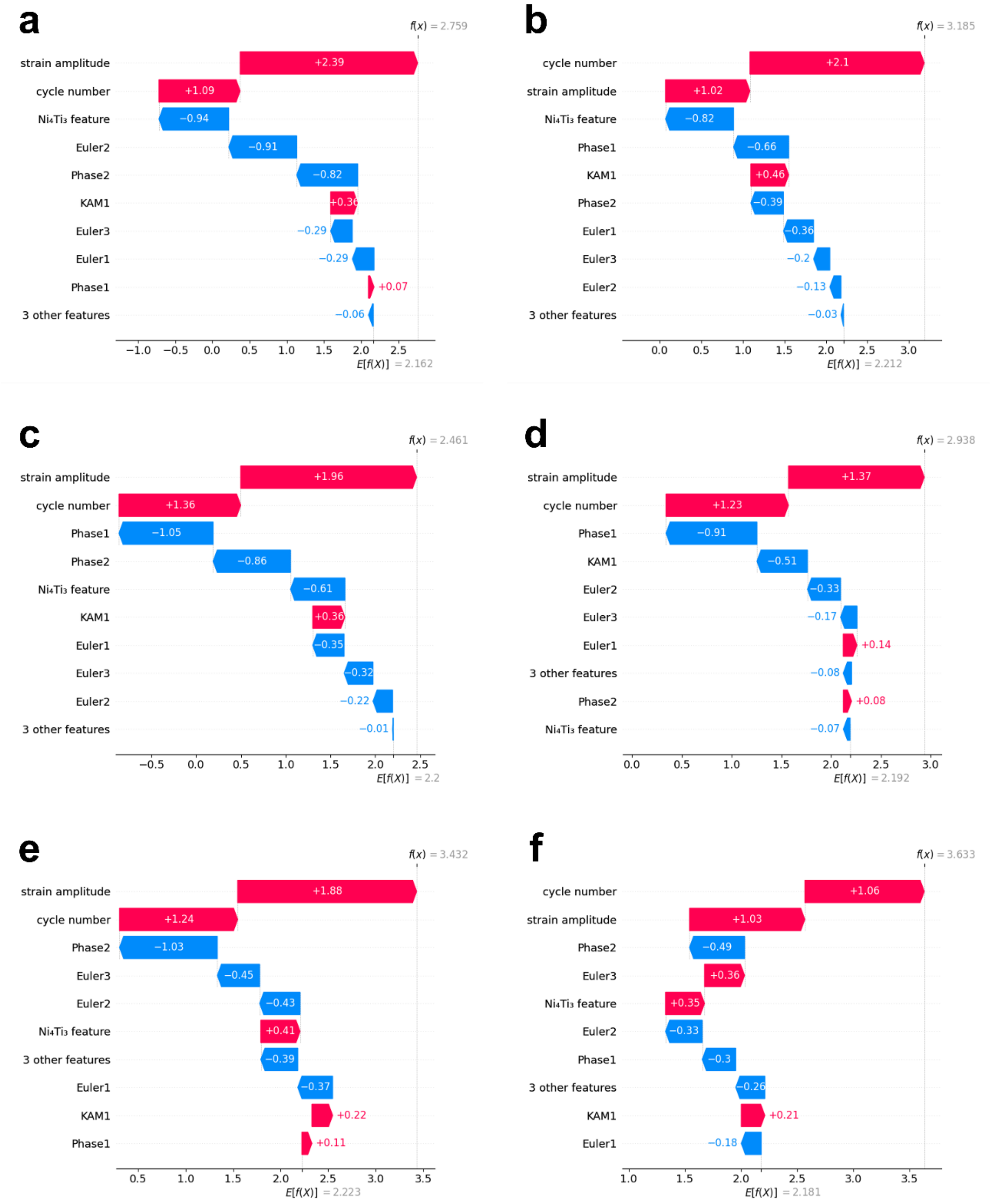
(e) Gaussian blur (f) Rotation (g) Shear transform (h) Mosaic

Supplementary Fig. 17 | Schematic illustration of the data augmentation effect on the KAM map.

Supplementary Table 5 | Statistical comparison of model performance in terms of MSE (×10^{-2}) and MAE (×10^{-1}) for six specimens tested at strain amplitudes of 2 %, 4 %, 6 %, and 8 %. The best result is represented in bold, followed by underline. Our model achieves the best overall accuracy with an average rank of 1.521. After removing the feature synergy module (BMN-w/oSM), the errors increase to 7.39 × 10^{-2} and 2.43 × 10^{-1}, corresponding to 14.9 % and 21.5 % degradations, respectively, which highlights the essential role of this module in enhancing overall predictive reliability

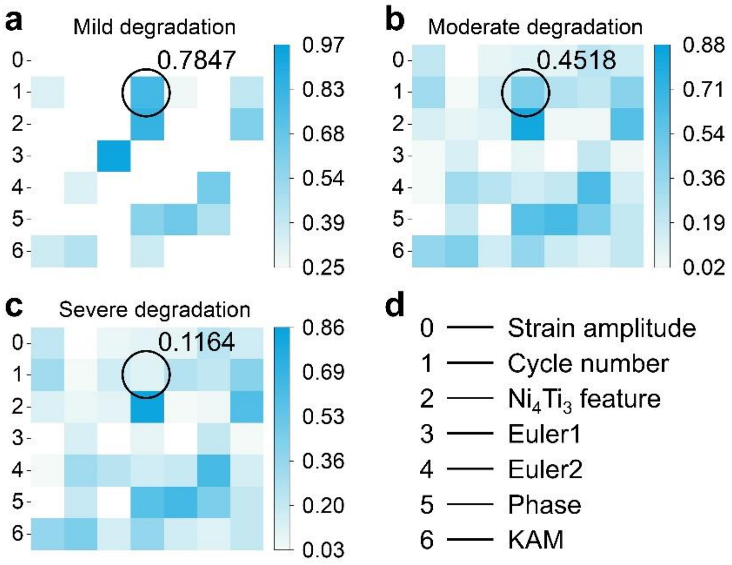
Models		MBFFA		Base-w/oSM		MBNI-Net		MCC-Net		ECMMF-Net	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Specimen I	2	**0.124**	**0.256**	0.158	0.345	0.149	0.298	0.183	0.412	0.201	0.455
	4	**2.620**	**1.315**	3.351	1.897	2.764	1.462	3.213	1.843	4.512	2.771
	6	7.851	2.278	9.112	3.045	**7.743**	**2.113**	10.94	3.681	12.37	4.268
	8	**10.27**	**2.646**	11.30	3.154	13.94	3.721	14.97	4.742	14.83	5.325
Specimen II	2	**0.128**	**0.264**	0.165	0.361	0.152	0.329	0.168	0.358	0.245	0.538
	4	2.703	1.357	**2.654**	**1.329**	3.387	1.742	3.223	1.812	5.032	2.391
	6	8.599	2.350	8.524	2.983	**8.342**	**2.203**	9.964	3.252	13.86	4.422
	8	**10.60**	**2.726**	10.81	3.025	13.74	3.505	14.28	3.792	14.24	4.783
Specimen III	2	0.133	0.275	**0.126**	**0.249**	0.175	0.333	0.131	0.412	0.233	0.564
	4	2.813	1.412	2.851	1.603	3.411	1.773	**2.745**	**1.389**	5.376	2.642
	6	**8.429**	**2.446**	9.524	3.246	10.38	3.201	8.68	2.589	13.15	3.893
	8	11.03	2.833	**10.92**	**2.698**	14.32	4.186	15.42	4.284	15.13	4.732
Specimen IV	2	0.137	0.283	0.144	0.322	**0.128**	**0.279**	0.157	0.353	0.282	0.598
	4	2.896	1.453	3.016	1.734	3.509	1.811	**2.852**	**1.452**	7.421	2.672
	6	**8.677**	2.518	8.764	3.063	9.684	**2.423**	8.989	2.893	11.34	4.462
	8	11.36	2.914	**10.77**	**2.739**	13.54	4.352	13.28	4.882	14.90	5.431
Specimen V	2	0.143	**0.297**	**0.142**	0.333	0.145	0.317	0.205	0.545	0.264	0.621
	4	3.034	**1.523**	**2.904**	1.653	3.107	1.644	3.364	1.918	7.506	2.745
	6	**9.090**	**2.637**	9.765	3.003	9.344	2.825	10.81	3.664	11.67	3.892
	8	11.90	**2.948**	15.03	4.566	**11.68**	3.125	14.32	3.808	14.51	5.324
Specimen VI	2	0.117	0.243	0.119	0.266	0.181	0.381	**0.112**	**0.243**	0.238	0.594
	4	**2.482**	**1.246**	3.501	1.945	3.352	1.812	4.283	2.474	4.762	2.876
	6	**7.438**	**2.158**	9.352	2.742	9.452	2.902	10.45	3.341	11.98	3.538
	8	**9.733**	**2.512**	14.91	3.874	11.36	2.973	14.46	4.163	12.88	4.511
Avg Rank		**1.521**		2.521		2.583		3.417		4.938	

Local interpretability



Supplementary Fig. 18 | Local interpretability of residual strain predictions for six annealing conditions at 6% strain after 10 cycles. a–f SHAP waterfall plots for specimens I–VI, respectively. Overall, strain amplitude and cycle number act as the dominant positive contributors across all specimens. The Ni_4Ti_3 feature generally provides a strong negative contribution, indicating that fine, dispersed Ni_4Ti_3 precipitates mitigate degradation, consistent with previous results. In contrast, KAM1 contributes more strongly in lightly annealed specimens, where larger residual stresses and lattice distortion more readily trigger superelastic degradation, but its influence diminishes in more heavily annealed states, where recovery lowers stored energy and moderates local damage.

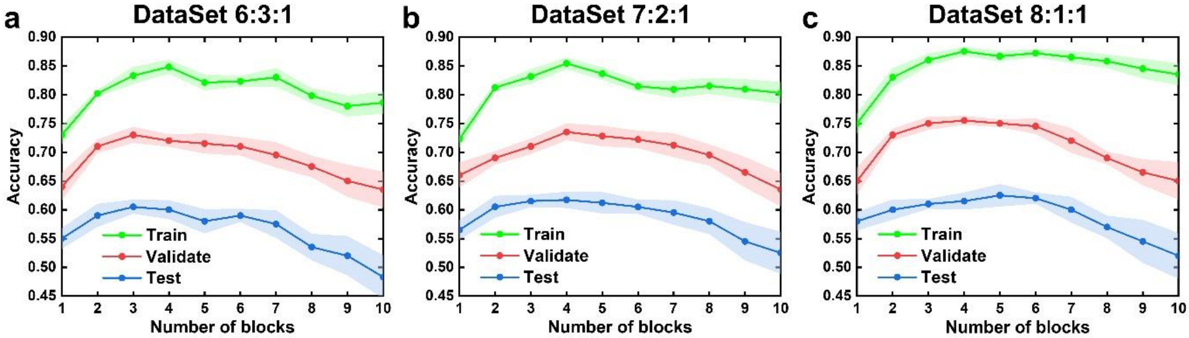
Feature correlation analysis



Supplementary Fig. 19 | Feature interaction matrices across different levels of superelastic degradation. a–c Feature interaction matrices corresponding to mild, moderate, and severe degradation. Colors denote the normalized interaction intensity between feature pairs. Black circles highlight the term between cycle number and orientation features, whose magnitude decreases progressively from 0.7847 (mild) to 0.4518 (moderate) and 0.1164 (severe). This systematic decline indicates that cyclically accumulated local damage and irreversible deformation increasingly dominate the material response, diminishing the influence of initial orientation and thereby weakening the coupling between cycle number and Euler orientation features as degradation advances. **d** Mapping between matrix indices and specific feature definitions.

Ablation experiments

Assess the impact of number of blocks and dataset allocation



Supplementary Fig. 20 | Influence of data-splitting ratios and network block numbers on model accuracy. Training, validation, and test accuracies are shown under three dataset partitions: (**a**) 6:3:1, (**b**) 7:2:1, and (**c**) 8:1:1. Accuracy was computed following the method reported in our previous study[15]. The test set was kept fixed, while the remaining samples were allocated to training and validation according to each ratio. Increasing the proportion of training data led to a progressive improvement in performance, with the 8:1:1 split achieving the highest accuracy. The effect of network depth exhibited a characteristic non-monotonic trend: performance peaked when the network comprised approximately 3–5 blocks, whereas further increasing the number of blocks did not yield additional gains. Instead, excessive depth caused feature representations to become increasingly redundant and saturated, resulting in reduced model stability and a decline in predictive accuracy.

Why Euler orientation matrix instead of IPF map? We encode the three Euler angles into three channels to construct the Euler matrix that represents the actual crystal orientation, rather than directly using the more readily available IPF map. The IPF map essentially compresses the orientation of a selected reference direction in the crystal frame into RGB colors, where a single color may correspond to multiple crystallographically orientations (**Supplementary Fig. 21a**). In contrast, the Euler matrix fully describes all three degrees of freedom. Interestingly, this approach unexpectedly introduces a new modality, yet it can still be processed in the same manner as an image matrix.

re-IPF: In this ablation variant, we replace the Euler matrix with the IPF map as the orientation input.

Why is PCA used to reduce the dimensionality of Ni_4Ti_3 features? We extract multiple descriptors of Ni_4Ti_3 from TEM, including diameter, aspect ratio, and area, and then compress them into a single principal component via PCA. Similar strategies have been applied in other material systems. For instance, researchers have used PCA to reduce the dimensionality of γ'-phase morphological descriptors, improving the robustness of creep strain prediction in Ni-based superalloys[16]. For the present work, if all precipitate features are directly concatenated into the numerical input rows, the performance degrades significantly (**Supplementary Table 6**). This likely occurs because the inflated dimensionality of the numerical branch interferes with weight allocation among other dominant factors.

w/o-PCA: We omit the PCA step and directly concatenate all individual Ni_4Ti_3 precipitate features into the numerical input branch.

Importance of KAM and phase maps. For NiTi alloys, high KAM regions are prone to damage evolution and functional degradation. Here KAM is used as a scalar field to quantify local damage, while detailed mechanistic discussion is beyond the scope of this study. $NiTi_2$ appears as incoherent, brittle second-phase particles along grain boundaries or within grains, acting as stress concentrators, promoting dislocation slip and microcrack initiation, and thereby degrading superelastic stability[17,18]; our observations align with this view (**Supplementary Fig. 21b**). Ablation experiments show that removing either the KAM or phase map input leads to a marked drop in accuracy.

w/o-KAM: we remove the KAM channel and retain only the Euler matrices and phase maps as matrix inputs.

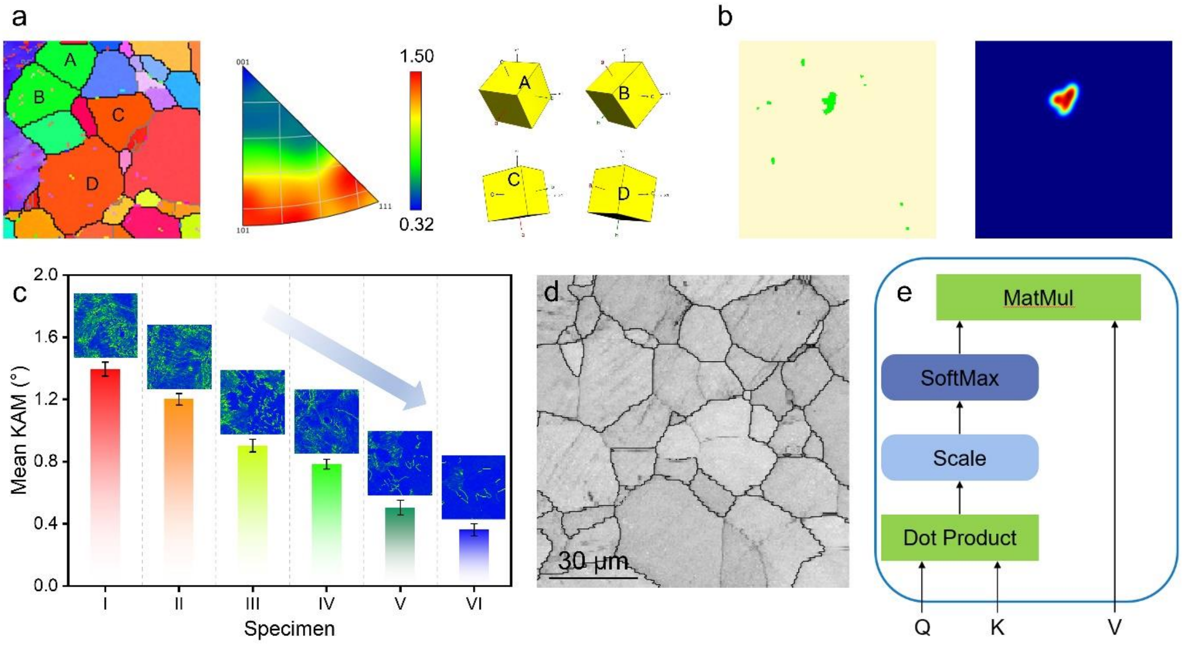
w/o-Phase: we remove the phase map channel and retain only the Euler matrices and KAM maps as matrix inputs.

Other available microstructural information. BC maps mainly capture the contrast and sharpness of backscattered electron patterns (**Supplementary Fig. 21d**), allowing visualization of grain morphology, grain boundaries and surface condition. We tested BC as a fourth image channel, but the gain in prediction accuracy was negligible, so it was not retained in the main framework. A plausible explanation is that degradation-related information is already encoded in the existing inputs, whereas BC contributes little independent signal and is highly sensitive to sample preparation and imaging conditions, thus introducing task-irrelevant noise.

ad/BC: we additionally feed the BC map as a fourth image channel, with the remaining network architecture and training configuration unchanged.

Feature synergy module. The extracted matrix features and numerical features are projected into a shared embedding space to mitigate cross-modal context misalignment, thereby improving the effectiveness and stability of multimodal fusion.

re-FSM: in this ablation variant, the feature synergy module is replaced with a multi-head self-attention module[19].

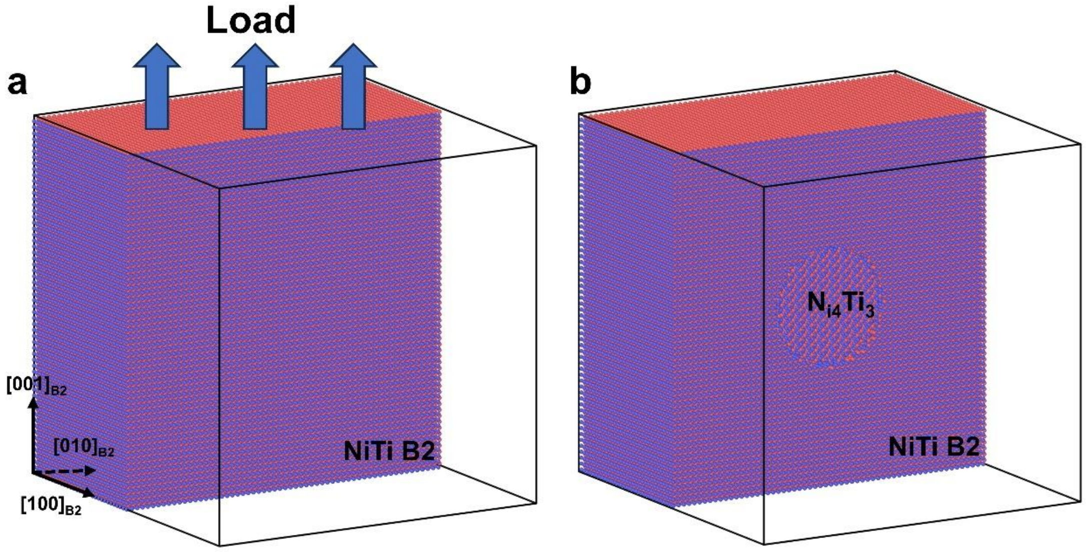


Supplementary Fig. 21 | Supplementary illustration of different input modalities. **a** IPF map and crystallographic orientation schematics of four selected grains A–D, illustrating how grains A/B or C/D differ markedly in orientation yet still exhibit similar colours, reflecting the information simplification in the IPF representation. **b** Phase map, where green denotes $NiTi_2$ precipitates and the light background the matrix, together with the corresponding response heatmap derived from this map. **c** Mean KAM values for specimens I–VI subjected to various heat treatments, with representative KAM maps for each specimen shown above. **d** BC image highlighting the grain morphology and grain boundaries. **e** Schematic of the self-attention mechanism. The Q matrix of each element is dot-producted with the K matrix of the other elements, and then the result of the dot-product is scaled and fed into the Softmax function for normalization. Finally, the V matrix is multiplied by the output of the Softmax function to determine the relative importance of each element.

Supplementary Table 6 | Ablation study of model components in MBFFA

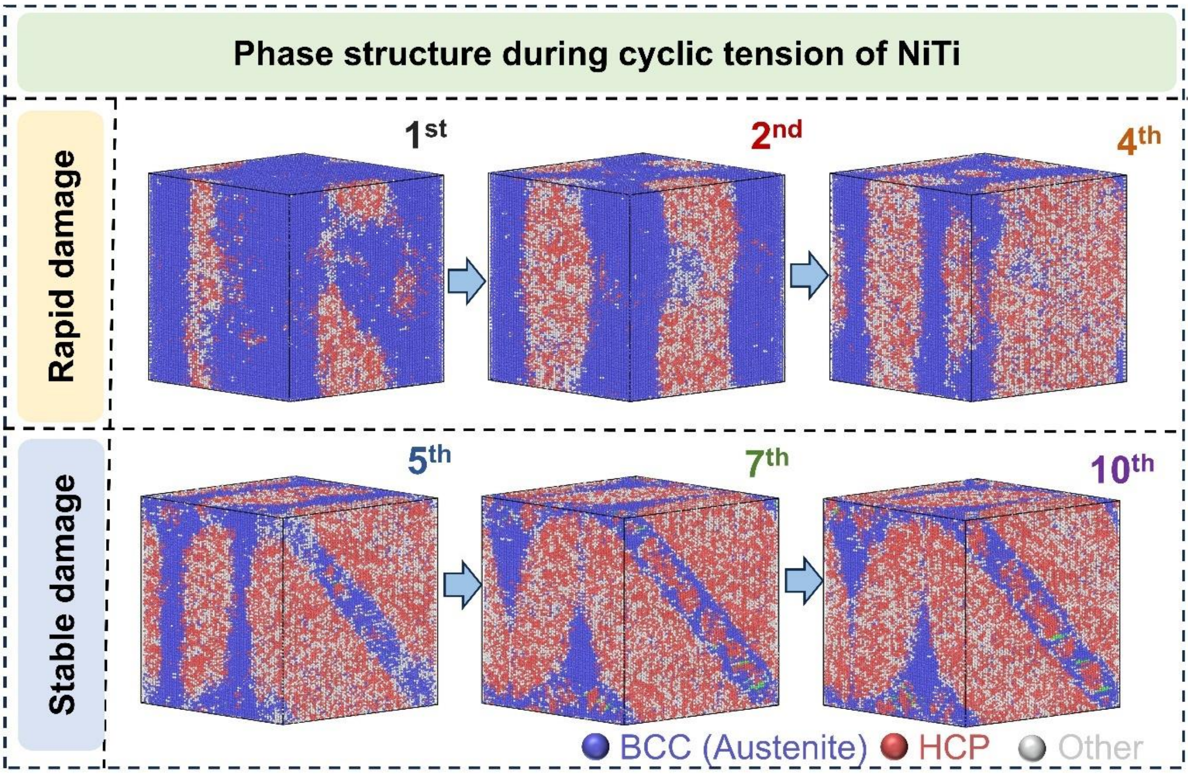
Method	MSE (10^{-1})	MAE	R^2 (10^{-2})
w/o-PCA	0.647	0.318	84.91
w/o-KAM	0.634	0.307	86.26
w/o-Phase	0.587	0.207	89.79
re-FSM	0.579	0.201	93.48
re-IPF	0.567	0.190	94.35
ad/BC	0.556	0.178	97.11
Base	0.552	0.172	98.02

Molecular dynamics

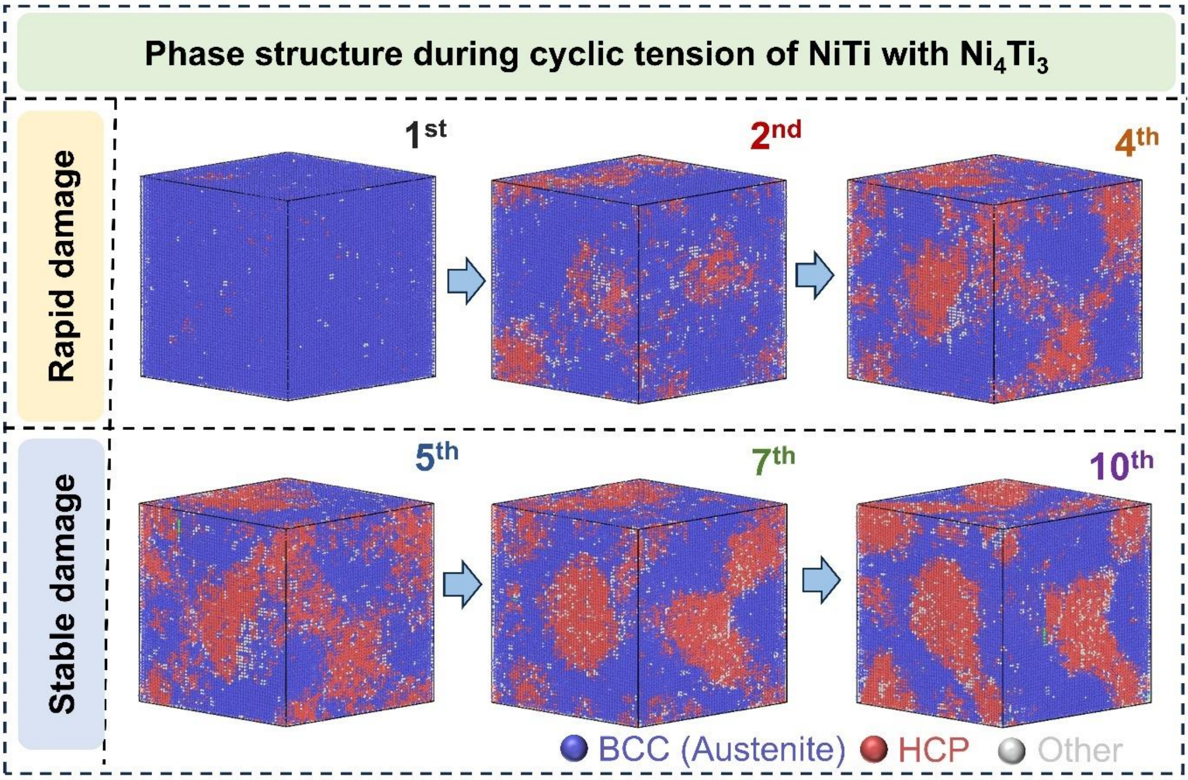


Supplementary Fig. 22 | Atomistic models used in this study (sliced at 1/2 along the x-axis). a NiTi without Ni_4Ti_3 precipitates. **b** NiTi with Ni_4Ti_3 precipitates.

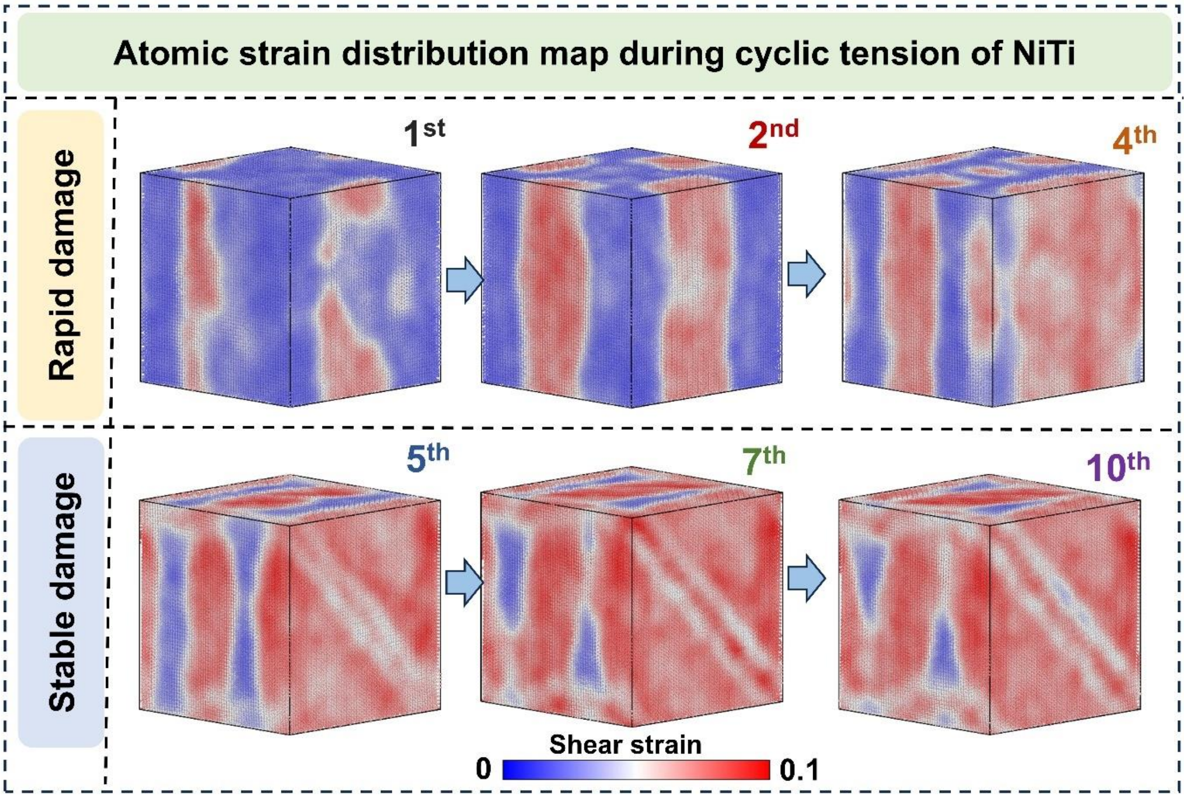
5



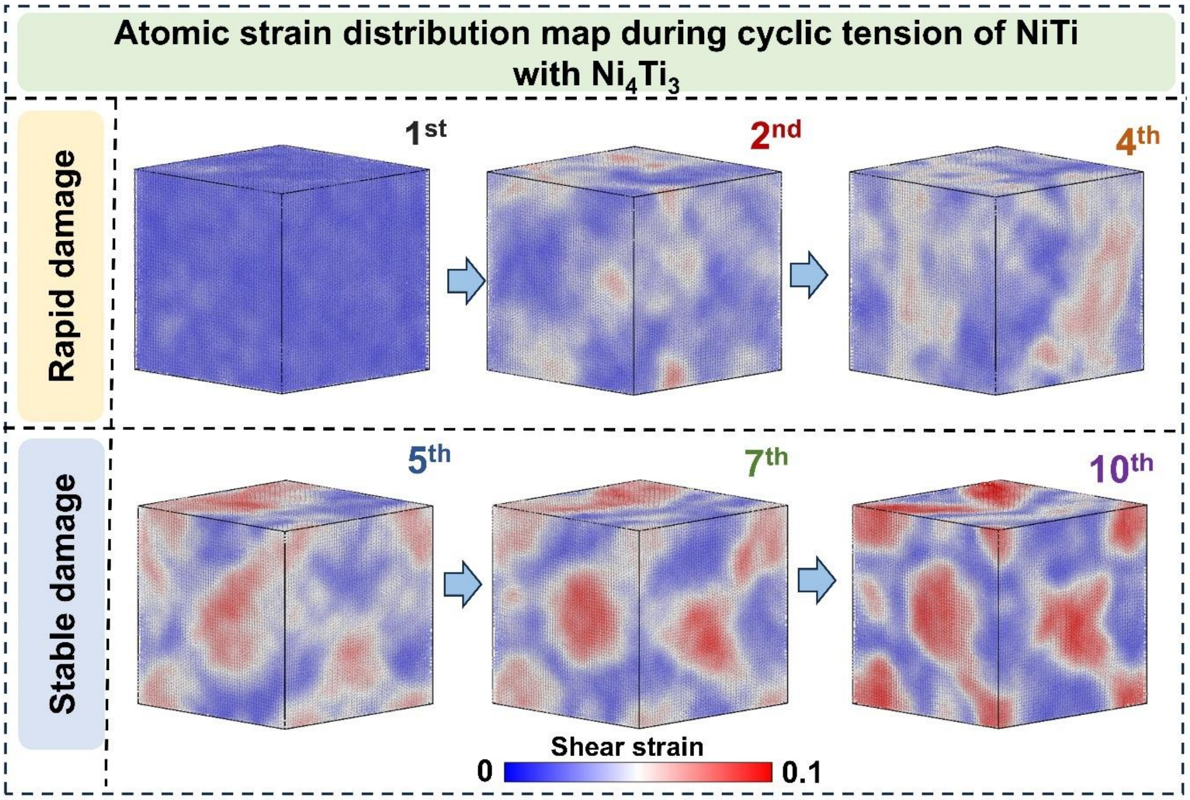
Supplementary Fig. 23 | Atomistic configurations of NiTi after each cycle in the cyclic simulation.



Supplementary Fig. 24 | Atomistic configurations of NiTi with Ni_4Ti_3 precipitates after each cycle in the cyclic simulation.



Supplementary Fig. 25 | Atomic strain maps of NiTi after each cycle in the cyclic simulation.

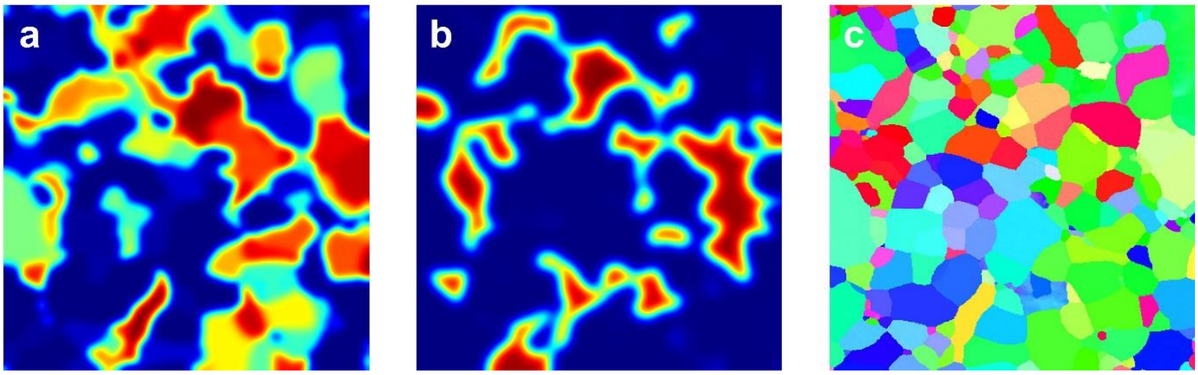


Supplementary Fig. 26 | Atomic strain maps of NiTi with Ni_4Ti_3 precipitates after each cycle in the cyclic simulation.

5

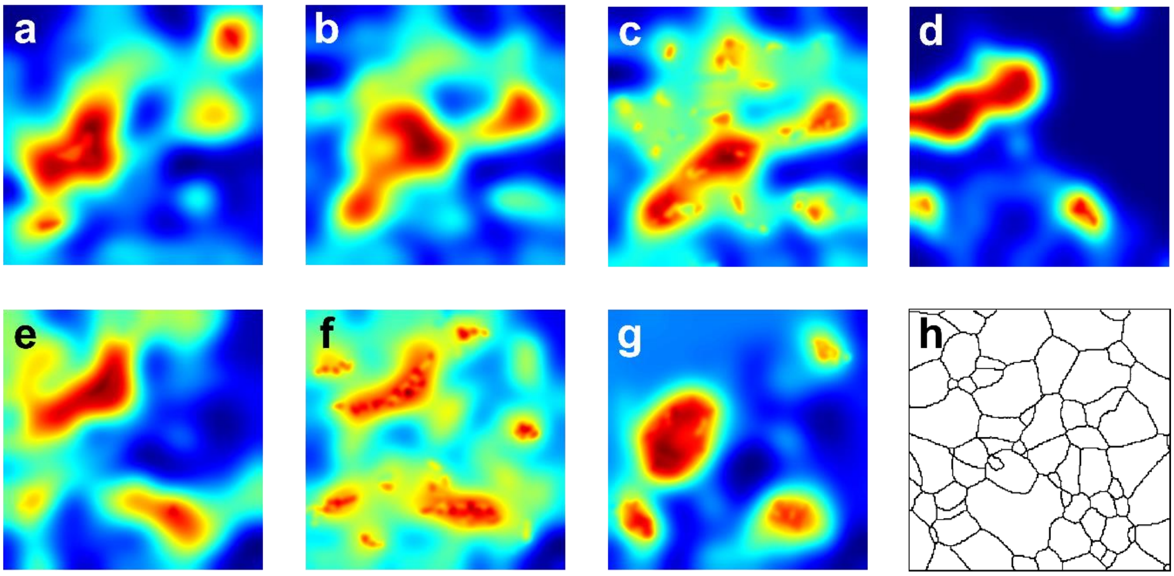
Visualization results

Raw heatmap used for visual illustration



Supplementary Fig. 27 | Schematic of the heatmap visualization workflow. **a** Response heatmap obtained by applying the channel separation strategy proposed in this work to visualize features from the Euler orientation matrix. **b** Reference heatmap generated using the conventional channel stacking strategy. **c** IPF orientation map of the corresponding region. In the main text (**Fig. 2**), panels (**a**) and (**b**) are respectively overlaid with (**c**) to illustrate how high response regions in the heatmaps align with grain morphology and crystallographic orientation.

Raw heatmap used to elucidate degradation mechanism



Supplementary Fig. 28 | Euler matrix response heatmaps. **a–g** Response heatmaps obtained by feature visualization of the Euler orientation matrices in the same region, corresponding to different strain amplitudes and cycle numbers: (**a**) 2%–1 cycle; (**b**) 2%–3 cycles; (**c**) 2%–10 cycles; (**d**) 8%–1 cycle; (**e**) 8%–3 cycles; (**f**) 8%–10 cycles; (**g**) 6.38%–1 cycle. **h** Grain boundary map of the same region. In **Fig. 4** of the main text, the heatmaps in (**a–g**) are spatially superimposed with the grain boundary contours in (**h**), and the high response regions are further compared with the Schmid factor distribution to identify the grain orientations and associated deformation mechanisms that dominate superelastic degradation.

References

1. Inoue, H., Miwa, N. & Inakazu, N. Texture and shape memory strain in TiNi alloy sheets. *Acta Mater.* **44**, 4825–4834 (1996).
2. Hytch, M., Snoeck, E. & Kilaas, R. Quantitative measurement of displacement and strain fields from HREM micrographs. *Ultramicroscopy* **74**, 131–146 (1998).
3. Wang, X. et al. Improved functional stability of a coarse-grained Ti-50.8 at.% Ni shape memory alloy achieved by precipitation on dislocation networks. *Scr. Mater.* **163**, 57–61 (2019).
4. Xu, B. et al. Effect of Ni4Ti3 precipitates on the functional properties of NiTi shape memory alloys: A phase field study. *Int. J. Plast.* **177**, 103993 (2024).
5. Zhou, N. et al. Effect of Ni4Ti3 precipitation on martensitic transformation in Ti–Ni. *Acta Mater.* **58**, 6685–6694 (2010).
6. Gall, K., Sehitoglu, H., Chumlyakov, Y. & Kireeva, I. Tension-compression asymmetry of the stress-strain response in aged single crystal and polycrystalline NiTi. *Acta Mater.* **47**, 1203–1217 (1999).
7. Jin, Y. & Weng, G. A direct method for the crystallography of martensitic transformation and its application to TiNi and AuCd. *Acta Mater.* **50**, 2967–2987 (2002).
8. Liu, Y. The superelastic anisotropy in a NiTi shape memory alloy thin sheet. *Acta Mater.* **95**, 411–427 (2015).
9. Tirry, W. & Schryvers, D. Quantitative determination of strain fields around Ni4Ti3 precipitates in NiTi. *Acta Mater.* **53**, 1041–1049 (2005).
10. Hornbogen, E., Mertinger, V. & Wurzel, D. Microstructure and tensile properties of two binary NiTi-alloys. *Scr. Mater.* **44**, 171–178 (2001).
11. Ren, D., Wang, C., Wei, X., Lai, Q. & Xu, W. Building a quantitative composition-microstructure-property relationship of dual-phase steels via multimodal data mining. *Acta Mater.* **252**, 118954 (2023).
12. Wu, Y. et al. A versatile multimodal learning framework bridging multiscale knowledge for material design. *npj Comput. Mater.* **11**, 276 (2025).
13. Goodale, M., Milner, A., Jakobson, L. & Carey, D. A neurological dissociation between perceiving objects and grasping them. *Nature* **349**, 154–156 (1991).
14. Xu, M., Yoon, S., Fuentes, A. & Park, D. A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognit.* **137**, 12 (2023).
15. Hu, Y. et al. Investigation of full-field strain evolution behavior of Cu/Ni clad foils by interpretable machine learning. *Int. J. Plast.* **184**, 104181 (2025).

16. Song, J. et al. Phase-field-informed machine learning on creep behavior of Ni-based single-crystal superalloys. *J. Mater. Inform.* **5**, 19 (2025).
17. Li, S., Hassanin, H., Attallah, M., Adkins, N. & Essa, K. The development of TiNi-based negative Poisson's ratio structure using selective laser melting. *Acta Mater.* **105**, 75–83 (2016).
18. Otsuka, K. & Ren, X. Physical metallurgy of Ti-Ni-based shape memory alloys. *Prog. Mater. Sci.* **50**, 511–678 (2005).
19. Han, K. et al. A Survey on Vision Transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 87–110 (2023).