From Image to Properties: Deep Learning for Microstructure Property Prediction

Shih-Chieh Chou[1], I-Chieh Cheng[2], Bo-Shiuan Li[1]

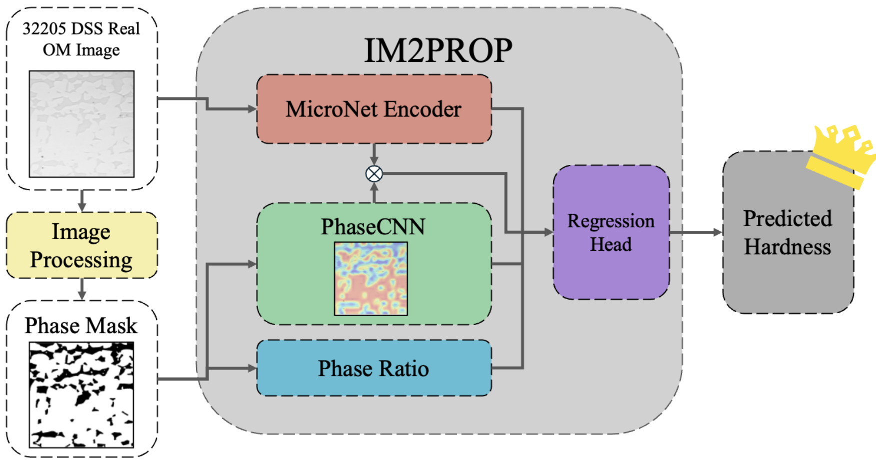


Fig. 1. Overview of the IM2PROP framework. Real optical microscope images of S32205 duplex stainless steel are processed through three complementary branches — a MicroNet-pretrained encoder, a PhaseCNN with spatial attention, and macroscopic phase ratios — whose fused representations are regressed to predict Micro-Vickers hardness.

***Abstract*—Predicting local mechanical properties from microstructural images is essential for reducing the experimental burden in materials development, yet existing deep learning approaches predominantly rely on synthetic training data, while domain-specific pretraining remains underexplored for image-based property regression. This paper presents IM2PROP, a tri-branch regression framework that predicts Micro-Vickers hardness of S32205 duplex stainless steel directly from real optical microscopy images. The architecture fuses a MicroNet-pretrained ResNet50 encoder with a phase-driven spatial attention mechanism derived from binary phase masks and macroscopic phase ratio inputs, embedding physical domain knowledge into the feature extraction process. A physically grounded dataset of 94 samples is constructed using the expanding cavity model to define regions of interest that match each indentation's plastic deformation zone. A full-factorial ablation study with repeated five-fold cross-validation (360 training runs) demonstrates that the best configuration, which couples phase-driven attention with phase-mask features, achieves a cross-validated mean absolute error of 0.175 GPa — a 12.7% improvement over the single-branch baseline — and that adding macroscopic phase ratios lowers the average error at every training duration.**

***Index Terms*—Materials informatics, deep learning, microstructure–property relationship, duplex stainless steel, Vickers hardness**

[1]*Shih-Chieh Chou and Bo-Shiuan Li are with Department of Mechanical and Electro-Mechanical Engineering, National Sun Yat-sen University.*
[2]*I-Chieh Cheng is with Department of Materials and Optoelectronic Science, National Sun Yat-sen University.*

I. Introduction

Micromechanical properties serve as a critical foundation for predicting the overall performance of materials. Microstructural features, including distinct crystal orientations, grain sizes, chemical compositions, and phase distributions, directly determine local mechanical behavior. Traditionally, evaluating these local micromechanical properties relies heavily on destructive testing using physical instruments, such as the Micro-Vickers hardness test. Although such tests can effectively quantify local mechanical responses, they remain constrained by complex and time-consuming specimen preparation and testing procedures. Particularly during the stages of new material development and process optimization, relying solely on physical measurements for iterative validation testing inevitably causes unnecessary delays in research and development cycles and keeps testing costs prohibitively high.

In recent years, deep learning technologies have introduced unprecedented opportunities in materials science. However, most current deep learning applications for microstructural images are still limited to grain classification or semantic segmentation tasks. Research on regression tasks that directly predict continuous mechanical properties, such as hardness, remains relatively scarce; the few existing frameworks that explore the prediction of material mechanical properties often rely heavily on computer-generated synthetic microstructural images for training, which struggle to fully capture real-world conditions. Furthermore, the features extracted by traditional models pre-trained on ImageNet [1] often fail to correspond with complex material microstructures. To address this domain adaptation issue, recent pre-trained models like MicroNet [2], which are trained on massive datasets of real microscopy images, have been proven to provide more physically meaningful feature representations. Nevertheless, how to effectively utilize such domain-specific encoders for robust, end-to-end prediction of local mechanical properties directly from real material images remains a critical research gap that urgently needs to be filled.

To bridge this gap, this study proposes IM2PROP (Fig. 1), a novel tri-branch deep learning regression framework for predicting the local Micro-Vickers hardness of duplex stainless steel (DSS) using real optical microscope (OM) images. The core innovation of this framework lies in embedding microstructural priors into the feature extraction process. To achieve this, IM2PROP goes beyond standard microscopy-encoded RGB features by simultaneously extracting topological representations from binary phase masks generated by an image processing pipeline, and directly integrating macroscopic phase ratios. A phase-derived spatial attention mechanism modulates the encoder response based on phase distribution, allowing the network to incorporate phase-level information beyond what RGB intensity alone provides. This physically-informed fusion strategy not only overcomes the black-box nature of traditional convolutional neural networks (CNN) but also establishes a highly interpretable and robust mapping for accelerating material evaluation.

II. Related Work

A. Microstructure–Property Linkage via Deep Learning

Establishing quantitative linkages between microstructure and mechanical properties has long been a central pursuit in materials science. Early efforts combined hand-crafted statistical features with traditional machine learning—for example, Pathan et al. [3] used PCA of two-point correlation functions with gradient-boosted tree regression to predict composite stiffness and yield strength. With the advent of CNNs, end-to-end image-to-property regression became feasible: Sun et al. [4] predicted stress fields in fiber-reinforced polymers using a modified StressNet, Abueidda et al. [5] achieved 99.2% computational speedup over FE modeling with a customized CNN, and Larmuseau et al. [6] predicted hardness of martensitic steels directly from scanning electron microscope (SEM) images. Kondo et al. [7] further demonstrated CNN-based regression for predicting ionic conductivity from microstructure image quality maps.

However, most existing regression frameworks rely on computer-generated synthetic microstructure images for training [3]–[5]. A significant domain gap exists between such idealized images and real micrographs, which contain noise, uneven illumination, and etching variations that synthetic data fails to capture [8]. In the context of dual-phase steels, Cheloee Darabi et al. [9] trained a hybrid ResNet50–VGG16 framework entirely on phase-field-simulated microstructures and explicitly noted that conventional data augmentation (rotation, flipping) is inapplicable to microstructure–property regression, as such transformations disrupt the physical correspondence between phase morphology and mechanical response.

An alternative line of research trains directly on real experimental images. Wang et al. [10] integrated real microscopy images with compositional data to predict Vickers hardness of Mg–Gd alloys, and Yekta et al. [11] demonstrated that pre-trained vision transformers (CLIP, DINOv2, SAM) can extract task-agnostic features from real SEM images for superalloy microhardness prediction. Nevertheless, end-to-end regression frameworks that predict local micromechanical properties—such as Micro-Vickers hardness—from real material images remain scarce, representing a critical gap that the present study aims to address.

B. Domain-Specific Pre-training in Materials Science

ImageNet pre-trained encoders are the default feature extractors in most deep learning workflows; however, their learned representations—tuned to natural images of animals, vehicles, and everyday objects—differ fundamentally from the microstructural features (grain boundaries, phase distributions, precipitates) found in microscopy images [2]. To address this domain mismatch, Stuckner et al. [2] created MicroNet, a dataset of over 110,000 labeled microscopy images spanning 54 material classes, and showed that MicroNet pre-trained encoders generalize better to out-of-distribution micrographs and are more accurate in few-shot scenarios—reducing relative IoU error by 72.2% when training with only a single image. Subsequent studies confirmed these advantages: Alrfou et al. [12] showed that in-domain pre-training benefits one-shot and out-of-distribution learning via CS-UNet, and Marcjan et al. [13] validated MicroNet-initialized DeepLabv3 for segmenting irregular microstructural elements.

Notably, MicroNet and its derivatives have been used almost exclusively for segmentation tasks, while their potential as feature encoders for regression—particularly for predicting mechanical properties—remains largely unexplored. This study leverages a MicroNet pre-trained encoder as the primary feature extractor in IM2PROP, extending domain-specific pre-training to microstructure-to-property regression for the first time.

C. Domain Knowledge via Multi-Branch Fusion

Multi-branch fusion architectures provide an effective vehicle for integrating domain knowledge into deep learning models. Xue et al. [14] fused visual microstructure data with macroscopic properties via a hybrid CNN–MLP model, Wang et al. [10] combined image features with elemental composition for hardness prediction, and CrabNet [15] and Takeshita et al. [16] employed attention mechanisms to reveal element-level and diffraction-peak-level contributions to property predictions, respectively. This architectural integration aligns with the broader paradigm of physics-informed deep learning, which embeds domain knowledge directly into model design [17].

Such architectural integration contrasts with post-hoc explainable AI methods such as Grad-CAM [18], which have been widely applied in materials science [5], [7], [19], [20] but operate retrospectively on already-trained models. While retrospective explanation provides valuable diagnostic insight — though Adebayo et al. [21] noted limitations such as insensitivity to model parameter randomization — it does not influence what the model learns to attend to during training.

IM2PROP follows the architectural integration paradigm: its phase-derived spatial attention mechanism injects phase-level information directly into the encoder during training, and its tri-branch design fuses MicroNet-encoded RGB features, phase mask topology, and macroscopic phase ratios into a unified representation.

D. Micromechanical Behavior of Duplex Stainless Steels

DSSs are Fe–Cr–Ni alloys with a ferritic–austenitic dual-phase microstructure, typically containing 22–25% Cr and 4–8% Ni [22], [23]. The ideal 50:50 phase ratio provides a combination of the high strength of BCC ferrite and the superior ductility of FCC austenite. Nanoindentation studies have established that ferrite generally possesses higher elastic modulus and hardness than austenite [24], [25], with Xie et al. [26] further showing that ferrite develops a shallower deformation zone under cyclic loading. Argandoña et al. [27], [28] demonstrated that heat treatment temperature governs sigma-phase precipitation and, consequently, the overall hardness distribution.

The selection of S32205 cold-rolled DSS as the target material aligns directly with the IM2PROP architecture in three ways: (i) after polishing, the two phases exhibit clear bright–dark contrast under OM, enabling reliable binary phase mask generation without SEM or Electron Backscatter Diffraction (EBSD) ; (ii) heat treatment within the dual-phase region allows systematic control over grain size and phase ratio [29], providing diverse training data from a single material system; and (iii) the distinct local mechanical properties of the two phases ensure that Micro-Vickers hardness is inherently governed by the underlying phase distribution — precisely the relationship that the phase-driven spatial attention mechanism is designed to capture.

III. Methodology

A. Material Preparation and Microstructure Characterization

This study employs S32205 DSS cold-rolled plates as the base material. In the as-received cold-rolled condition, the extremely fine grains of S32205 DSS render the BCC ferrite and FCC austenite phases difficult to distinguish under OM. To address this, a homogenization heat treatment was applied following the approach [29], specimens were held at 1100°C for 24 hours in a high-temperature furnace, then furnace-cooled to room temperature within the tube. Both specimens were subsequently ground and polished to achieve a mirror-like surface finish. Notably, no chemical etching was performed, as the two phases of the homogenized DSS already exhibit sufficient bright–dark contrast under OM in the as-polished condition.

To verify the microstructural evolution induced by the heat treatment, both the as-received cold-rolled and the homogenized specimens were characterized using OM, SEM, and EBSD (Fig. 2). EBSD analysis confirmed that the homogenization treatment effectively promoted grain growth and rendered the phase morphology clearly distinguishable, thereby providing high-quality OM images suitable for model training. The corresponding engineering stress–strain curves (Fig. 3) further confirm that the heat treatment altered the mechanical response of the material, consistent with the observed microstructural changes.

B. Dataset Construction: Linking Indentation to Microstructure

This section describes how a physically grounded correspondence between Micro-Vickers hardness measurements and local microstructural images is established, forming the data foundation of the IM2PROP framework.

1) Hardness Testing: Hardness measurements were performed using a Micro-Vickers indenter under a fixed load of $F = 1$ kgf. For each indentation, the two diagonals d_1 and d_2 (in μm) were measured, and the Vickers hardness was computed in GPa as

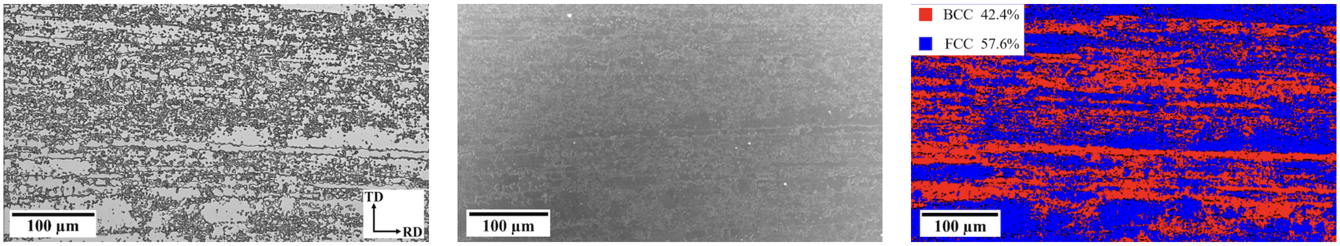
$$H_{\mathrm{GPa}} = \frac{18.1848523 \cdot F_{\mathrm{kgf}}}{1000 \cdot d^2}, \tag{1}$$

where $d = (d_1 + d_2)/2$ is the mean diagonal length in mm.

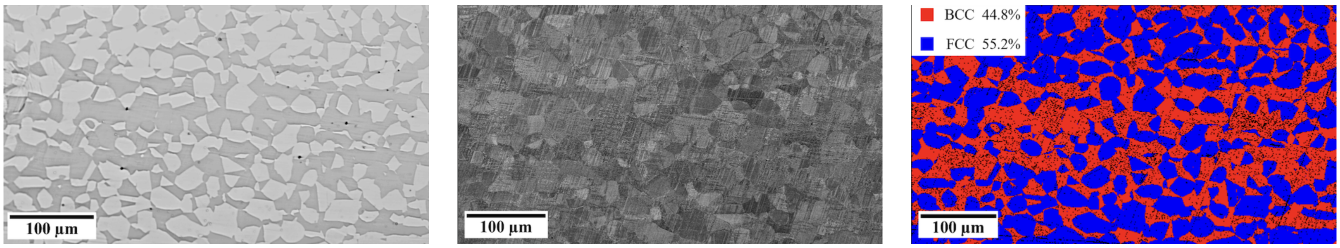
2) Plastic Zone Estimation and ROI Definition: A critical question in building a microstructure–property dataset is: what spatial extent of the microstructure does a single hardness measurement actually represent? To answer this, the expanding cavity model proposed by Mata et al. [30] was adopted to estimate the surface plastic zone radius C_s generated by the Vickers indentation. This model draws an analogy between the indentation process and the inflation of a spherical cavity, yielding closed-form solutions for the plastic zone size in terms of the indenter geometry and material properties (for the Vickers indenter, the projection factor $f = 1.079$). The complete derivation is provided in Appendix VII-A.

Based on this analysis, the region of interest (ROI) for each indentation site is defined as a square area centered on the indentation, with dimensions determined by the mean C_s across all 94 measurements. The resulting ROI size is 720×720 pixels, ensuring that each image encompasses the full extent of the plastic deformation zone and thus captures the microstructural features that physically govern the measured hardness value (Fig 4).

3) Dataset Composition: A total of 94 valid data samples were collected, each consisting of four elements: (1) a 720×720 pixel RGB OM image cropped around the indentation center; (2) a binary phase mask generated by the image processing pipeline described in Section III-C; (3) the light/dark phase ratios computed from the mask (in percent); and (4) the corresponding Vickers hardness label (in GPa). The descriptive statistics of the dataset are summarized in Table I.



(a) Cold-Rolled / OM (b) Cold-Rolled / SEM (c) Cold-Rolled / EBSD



(d) Homogenized / OM (e) Homogenized / SEM (f) Homogenized / EBSD

Fig. 2. Microstructural comparison of S32205 DSS before (top) and after (bottom) homogenization at 1100°C for 24 h. In the as-received cold-rolled condition, the ferrite and austenite phases are indistinguishable under OM (a) due to the extremely fine elongated grain structure, whereas SEM (b) and EBSD (c) confirm the dual-phase nature. After homogenization, significant grain growth produces clear bright–dark phase contrast visible even under OM (d), which is verified by the corresponding SEM (e) and EBSD (f) images.

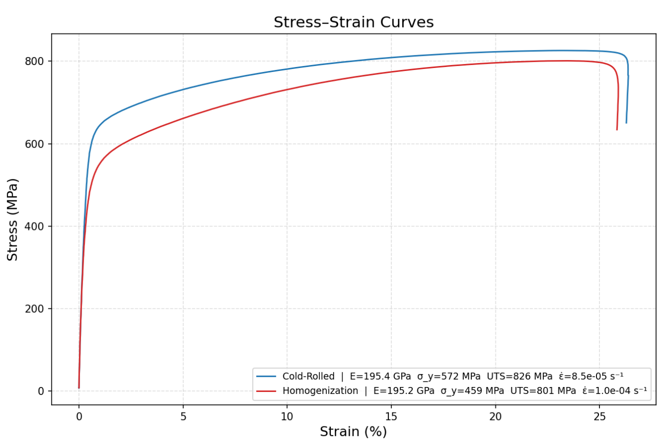


Fig. 3. Engineering stress–strain curves of S32205 DSS in the as-received cold-rolled and homogenized conditions. Homogenization reduces the yield strength (σ_y: 572→459 MPa) and ultimate tensile strength (UTS: 826→801 MPa) while preserving comparable ductility, reflecting the grain growth and relief of cold-work hardening induced by the heat treatment.

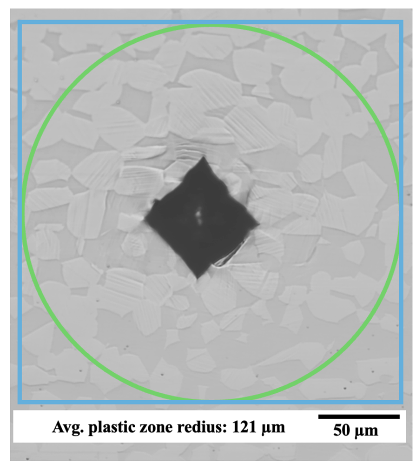


Fig. 4. Schematic of the ROI extraction strategy. The black diamond denotes the Vickers indentation, the green circle represents the average surface plastic zone radius ($\bar{C}_s \approx 121\,\mu$m) computed via the expanding cavity model, and the blue square indicates the 720×720 pixel ROI used as the model input.

4) Data Splitting Strategy: Model performance was estimated by five-fold cross-validation. The 94 samples were shuffled and partitioned into five folds of 18–19 samples; each fold served once as the test set, while the remaining four folds were further split into 85% training and 15% validation, yielding an approximate 67/13/20% train/val/test distribution. Every sample therefore receives exactly one held-out prediction per repetition, and the MAE is computed over all 94 held-out predictions. The procedure was repeated three times with different random seeds, and all configurations shared the same seeds, so they are compared on identical fold partitions.

C. Image Processing Pipeline for Phase Mask Extraction

To provide input for the second branch of IM2PROP (PhaseCNN), an automated phase mask extraction pipeline based on classical computer vision was developed. Classical computer vision (CV) methods were chosen over deep learning-based segmentation for two reasons: first, the two phases of homogenized DSS exhibit sufficiently clear contrast under OM, making reliable binarization achievable without complex semantic segmentation models; second, the dataset

TABLE I
DATASET DESCRIPTIVE STATISTICS ($n = 94$)

	Hardness (GPa)	**Plastic Zone Area** (μm^2)	**FCC Austenite** (%)	**BCC Ferrite** (%)
Mean	2.449	45,991	62.98	37.02
Std	0.182	3,040	10.69	10.69
Min	2.235	35,615	37.81	10.78
25%	2.327	44,930	56.15	31.19
Median	2.397	46,766	62.77	37.23
75%	2.495	48,174	68.81	43.85
Max	3.147	50,150	89.22	62.19

of only 94 images is insufficient for fine-tuning a dedicated segmentation network.

The pipeline consists of six sequential stages—median denoising, background normalization, CLAHE contrast enhancement, Otsu binarization, morphological refinement, and small-region filtering—producing a binary mask where pixel value 255 represents austenite (light phase) and 0 represents ferrite (dark phase). Phase ratios are computed directly as $N_{255}/N_{\text{total}} \times 100\%$ and $N_0/N_{\text{total}} \times 100\%$, respectively. A detailed description of each stage and the corresponding processing parameters are provided in Appendix VII-B.

D. IM2PROP Model Architecture

IM2PROP is a tri-branch deep learning regression framework whose core design principle is to embed physical domain knowledge into the feature extraction process. The overall architecture is illustrated in Fig. 5 the three branches process complementary levels of information and are fused before a shared regression head.

1) Branch 1: MicroNet-Encoded RGB Features: The first branch uses a ResNet50 encoder initialized with MicroNet pre-trained weights [2] (see Section II-B). The encoder is entirely frozen during training to prevent catastrophic overwriting of the domain-specific representations by the small 94-image dataset. The RGB input produces $2048 \times H' \times W'$ feature maps.

2) Branch 2: PhaseCNN and Phase-Driven Spatial Attention: PhaseCNN, a lightweight three-layer convolutional network, encodes the binary phase mask into 128-channel feature maps capturing phase boundary morphology and island distribution. These features are spatially aligned to Branch 1 resolution via adaptive average pooling and passed through a PhaseAttentionHead (1×1 conv + sigmoid), generating a $[0, 1]$ attention map. Element-wise multiplication of the attention map with the Branch 1 features integrates phase-derived spatial cues into the encoder's feature representation. Both the attended RGB features and raw phase features are then compressed via global average pooling into 2048-dim and 128-dim vectors, respectively.

3) Branch 3: Macroscopic Phase Ratios: The light/dark phase ratios are directly fed as a two-dimensional scalar vector, supplying global phase composition information complementary to the local pixel-level features of the first two branches.

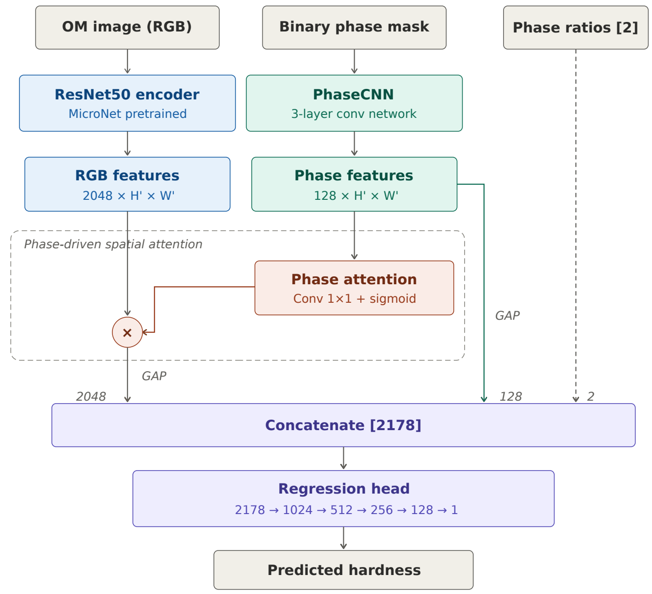


Fig. 5. Detailed architecture of IM2PROP. Branch 1 extracts 2048-dim features from the RGB image via a frozen MicroNet-pretrained ResNet50; Branch 2 derives a spatial attention map from the binary phase mask through PhaseCNN and a 1×1 convolution head; Branch 3 provides the two-dimensional phase ratio vector. The three representations are concatenated (2178-dim) and mapped to the predicted hardness by a five-layer regression head.

4) Feature Fusion and Regression Head: The three branch outputs ($2048 + 128 + 2 = 2178$ dim) are concatenated and mapped through a five-layer fully connected head ($2178 \rightarrow 1024 \rightarrow 512 \rightarrow 256 \rightarrow 128 \rightarrow 1$) with batch normalization, ReLU, and dropout (0.5 / 0.3) to the predicted Vickers hardness in GPa.

E. Training Strategy and Ablation Study Design

1) Loss Function and Optimizer: Training employs Smooth L1 Loss (Huber Loss), which behaves like L2 loss for small residuals (providing smooth gradients) and like L1 loss for large residuals (reducing the influence of outliers). The optimizer is Adam with a learning rate of 1×10^{-3}.

2) Evaluation Metric: Mean absolute error (MAE) is adopted as the primary evaluation metric, defined as

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|, \tag{2}$$

where y_i and $\hat{y}_i$ denote the measured and predicted hardness of the i-th sample, respectively. In hardness prediction, MAE carries a directly interpretable physical meaning: it represents the average absolute deviation in GPa. By contrast, mean squared error and root mean squared error,

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \quad \text{RMSE} = \sqrt{\text{MSE}}, \tag{3}$$

involve squaring the residuals, which compresses already small error magnitudes (e.g., 0.1–0.3 GPa) into values on the order

TABLE II
ABLATION STUDY: COMPONENT COMBINATIONS

Combo	R	A	F	Description
1	0	0	0	Baseline (RGB only)
2	0	0	1	+ Phase features
3	0	1	0	+ Attention
4	0	1	1	+ Attention + features
5	1	0	0	+ Phase ratios
6	1	0	1	+ Ratios + features
7	1	1	0	+ Ratios + attention
8	1	1	1	Full IM2PROP

of 10^{-2}, obscuring the physical scale. Accordingly, MSE and RMSE are reported only as supplementary metrics.

3) Ablation Study Design: The three switchable components — phase ratios (R), phase-driven spatial attention (A), and PhaseCNN features (F) — yield 8 combinations (Table II). Combo 1 (R=0, A=0, F=0) is the baseline using only MicroNet-encoded RGB features; Combo 8 (R=1, A=1, F=1) is the full IM2PROP model. Disabled components are replaced with zero tensors during the forward pass so that the network topology remains identical across all configurations. Each combination was trained at three durations (30, 50, and 80 epochs) and evaluated with the repeated five-fold cross-validation described in Section III-B, yielding $8 \times 3 \times 3 \times 5 = 360$ training runs.

IV. RESULTS

This section presents the results of the full-factorial ablation study comprising 360 training runs (8 combinations $\times$ 3 epoch settings $\times$ 3 repetitions $\times$ 5 folds). The cross-validated MAE (GPa), averaged over the three repetitions, serves as the primary metric, supplemented by learning curve analysis and generalization gap diagnostics.

A. Overall Ablation Performance

Fig. 6 and Fig. 8(b) summarize the cross-validated MAE across all configuration–epoch combinations. Combo 4 (R=0/A=1/F=1) trained for 80 epochs achieves the lowest MAE of **0.175 GPa**, representing a 12.7% reduction relative to the baseline (Combo 1) at the same epoch setting (0.201 GPa), which is also the baseline's best result.

Phase ratios provide the most consistent gain. Configurations that include ratios (Combos 5–8) achieve an average MAE of 0.209 GPa at 50 epochs, compared to 0.220 GPa for those without ratios (Combos 1–4)—a 4.7% reduction—and the ratio-inclusive average is lower at all three training durations, by 3–5%. Even in isolation, adding only phase ratios (Combo 5) yields 0.183 GPa at 80 epochs, outperforming most dual-branch configurations that lack ratio input. This aligns with the physical expectation that macroscopic phase composition directly governs the hardness baseline.

Attention alone (Combo 3) does not improve consistently over the baseline (MAE: 0.203 / 0.219 / 0.226 vs. 0.231 / 0.223 / 0.201 at 30/50/80 epochs), suggesting that spatial weighting without phase-specific features can be misleading. However, when attention is combined with phase features (Combo 4), the MAE drops to 0.187 at 30 epochs and 0.175 at 80 epochs — the best among ratio-free configurations at both settings, and the best overall at 80 epochs. The full model (Combo 8) is one of only three configurations, with Combos 4 and 5, that improve on the baseline at every training duration (0.195 / 0.202 / 0.181 GPa), but it does not surpass Combo 4 or Combo 7 (0.176 GPa) at 80 epochs: under cross-validation, combining all three components brings no further gain over the best two-component configurations.

B. Training Dynamics and Generalization

The learning curves (Fig. 7) reveal the same pattern in all configurations: the validation loss spikes transiently, peaking at epoch 8, and settles by epoch 16, after which the train–validation gap remains negligible. Longer training nevertheless helps: seven of the eight configurations reach their lowest MAE at 80 epochs, and 80 epochs outperforms 30 epochs in 20 of the 24 configuration–repetition pairs.

The generalization gap heatmap (Fig. 8(a)) confirms that overfitting is suppressed: no configuration exceeds the 0.15 threshold, and the largest gap is 0.042 (Combo 7 at 30 epochs). The frozen MicroNet encoder, which locks 23M+ parameters, effectively serves as a built-in regularizer across all configurations.

V. CONCLUSION

This paper presented IM2PROP, a tri-branch deep learning regression framework that predicts Micro-Vickers hardness of S32205 duplex stainless steel directly from real optical microscopy images. By embedding physical domain knowledge into the architecture—through a MicroNet-pretrained encoder, a phase-driven spatial attention mechanism, and macroscopic phase ratio inputs—IM2PROP reaches a cross-validated MAE of 0.175 GPa in its best configuration (attention with PhaseCNN features), a 12.7% improvement over the MicroNet-only baseline. The full-factorial ablation study across 360 cross-validated training runs reveals three key findings: (1) macroscopic phase ratios provide the most consistent gain, lowering the average MAE at every training duration, albeit by a modest 3–5%; (2) the phase-driven attention mechanism is most effective when coupled with phase-specific features, confirming that by-design interpretability requires complementary information sources; and (3) the full tri-branch model improves on the baseline at every training duration but does not surpass the best two-component configurations. The frozen MicroNet encoder further acts as a built-in regularizer, suppressing overfitting despite the limited 94-sample dataset.

VI. FUTURE WORK

Several directions remain for future investigation. First, incorporating sample-specific C_s values rather than the current uniform ROI size could better capture local deformation fields. Second, the transient validation-loss spike that all configurations show early in training suggests that advanced learning rate schedules such as warm-up or cosine annealing

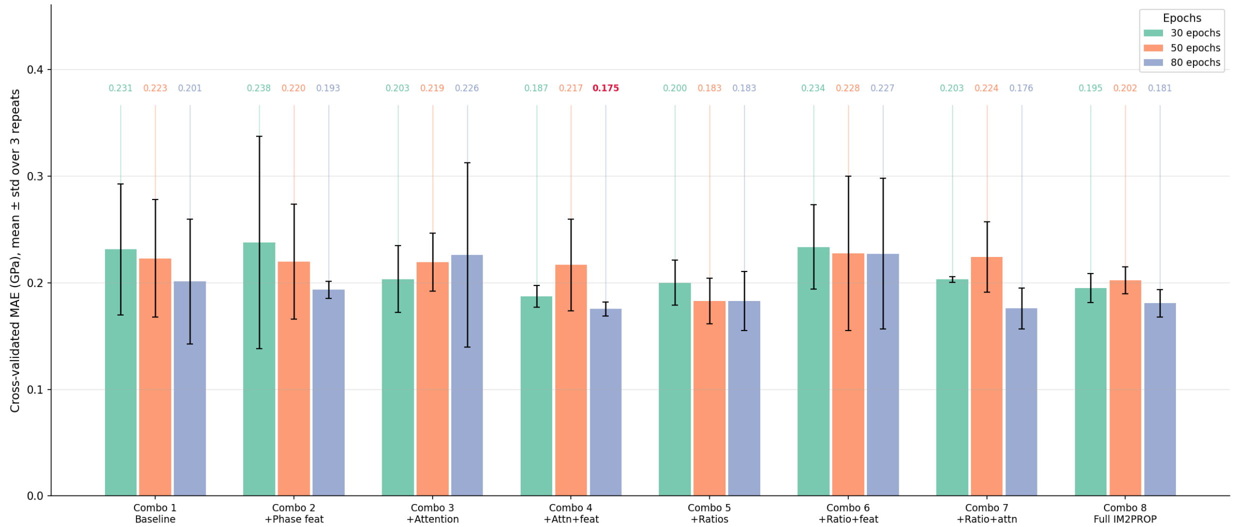


Fig. 6. Cross-validated MAE (mean ± std across 3 repetitions) for all eight ablation configurations at 30, 50, and 80 epochs. Combo 4 (attention + phase features) at 80 epochs achieves the overall best MAE of 0.175 GPa (highlighted in bold red). Seven of the eight configurations reach their lowest MAE at 80 epochs.

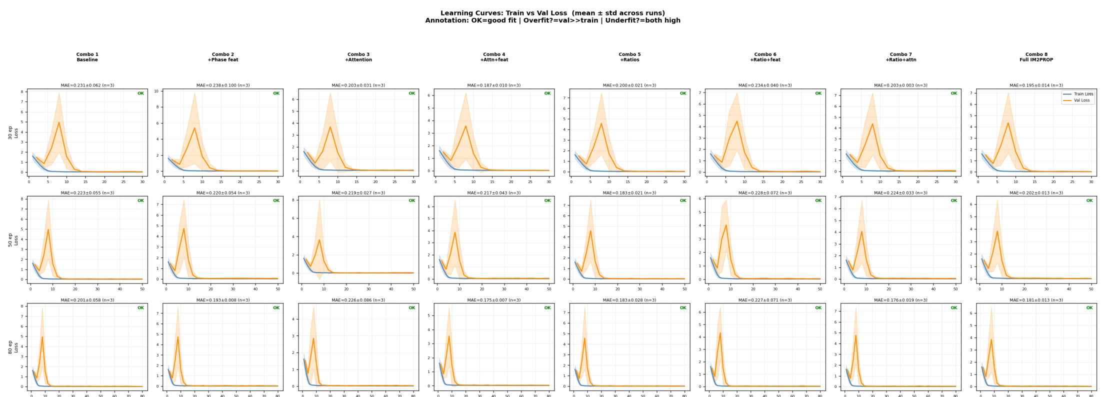
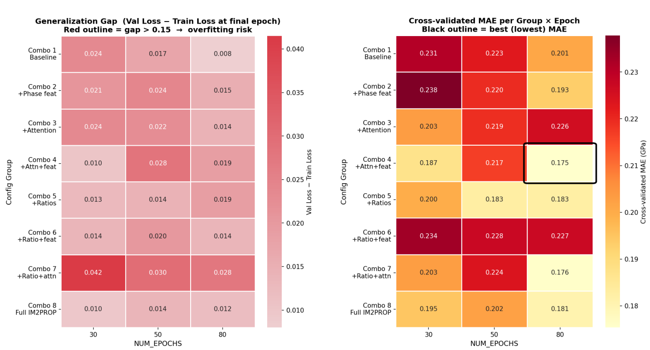


Fig. 7. Learning curves (train vs. validation loss, mean ± std across the 15 runs of each cell) for all 24 configuration–epoch combinations. In every configuration the validation loss spikes transiently, peaking at epoch 8, and settles by epoch 16, after which train and validation losses remain close.



(a) Generalization gap (b) Cross-validated MAE heatmap

Fig. 8. (a) Generalization gap (Val Loss − Train Loss at final epoch); cells with gap > 0.15 would be outlined in red as an overfitting risk, but no cell exceeds this threshold. (b) Cross-validated MAE per configuration and epoch; the global minimum (Combo 4, 80 epochs) is outlined in black.

may stabilize convergence of the multi-branch architecture. Third, and perhaps most significantly, the present model operates exclusively on two-dimensional OM images, which inherently lack crystallographic orientation information. In duplex stainless steels, the hardness of individual grains varies with crystal orientation, leading to a distribution of hardness values within each phase that follows an approximately normal distribution rather than the idealized bimodal (BCC vs. FCC) pattern. Integrating orientation data—for example, from EBSD maps—as an additional input branch could enable the model to disentangle phase-level and grain-level contributions to local hardness, potentially yielding finer-grained predictions and deeper physical insight into the microstructure–property relationship.

Acknowledgment

This research was supported by the College Student Research Grant, National Science and Technology Council (NSTC), Taiwan, under Grant No. 114-2813-C-110-052-E.

VII. Appendix

A. Plastic Zone Estimation via the Expanding Cavity Model

This appendix details the procedure for estimating the surface plastic zone radius C_s induced by a Vickers indentation, following the expanding cavity model of Mata et al. [30]. The model draws an analogy between sharp indentation and the inflation of a spherical cavity in an infinite elasto-plastic medium [31], [32], enabling closed-form estimation of the plastic zone size from measured hardness and indentation geometry.

Step 1: Vickers Hardness

Given a fixed load F_{kgf} and the two measured diagonals d_1, d_2 of the indentation (in μm), the mean diagonal is $d = (d_1 + d_2)/2$. The Vickers hardness in GPa is

$$H_{\mathrm{GPa}} = \frac{18.1848523 \cdot F_{\mathrm{kgf}}}{1000 \cdot d_{\mathrm{mm}}^2}, \tag{4}$$

where $d_{\mathrm{mm}} = d/1000$ converts to millimeters.

Step 2: Yield Stress Estimate

For elastic–perfectly plastic solids, the constraint factor relating hardness to yield stress is approximately 2.57 [30], [32]:

$$\sigma_{ys} = \frac{H}{2.57}. \tag{5}$$

Step 3: Projected Contact Area and Equivalent Radius

The projected contact area of a Vickers indentation is related to the mean diagonal by

$$A_c = \frac{d^2}{2}, \tag{6}$$

where d is in μm and A_c in μm^2. The equivalent contact radius, which maps the pyramidal imprint to a circular contact of equal area, is

$$a_s = \sqrt{\frac{A_c}{\pi}}. \tag{7}$$

Step 4: Indenter Shape Parameter

Following the parametrical analogy established in [30], the equivalent spherical cavity radius R is obtained by matching the plastic zone size between a conical indenter (semi-apical angle $\theta = 70.3°$, equivalent to Vickers geometry) and a spherical indenter at indentation strain $a_s/R = 0.635$:

$$R = \frac{a_s}{0.635}. \tag{8}$$

Step 5: Plastic Zone Depth

For the Vickers indenter, the cavity pressure is related to projected hardness through the Ludwick hardness $H_L = H/f$, where the projection factor $f = 1.079$ [30]. For elastic–perfectly plastic solids ($n = 0$), the relationship between hardness and plastic zone size along the symmetry axis is given by [30, Eq. 30]:

$$\frac{H}{\sigma_{ys}} = f\left[\frac{2}{3} + 2\ln\left(\frac{c}{R}\right)\right], \tag{9}$$

where $c = Z_{ys} + 1.217\, a_s$ is the total plastic zone parameter along the z-axis [30, Eq. 29]. Rearranging Eq. (9) for c:

$$c = R \cdot \exp\left[\frac{1}{2}\left(\frac{H}{f\,\sigma_{ys}} - \frac{2}{3}\right)\right]. \tag{10}$$

The plastic zone depth below the indentation is then

$$Z_{ys} = c - 1.217\, a_s. \tag{11}$$

Step 6: Surface Plastic Zone Radius

The plastic zone radius at the indented surface is obtained from the geometric relation [30, Eq. 37]:

$$C_s = R\sqrt{\left(\frac{c}{R}\right)^2 - 0.5868}. \tag{12}$$

The corresponding plastic zone area at the surface is $A_{pz} = \pi\, C_s^2$.

Numerical Example

For a representative indentation with $d_1 = 87.35$ μm, $d_2 = 87.53$ μm, and $F = 1$ kgf:

- Mean diagonal: $d = 87.44$ μm
- Hardness: $H = 2.378$ GPa
- Yield stress: $\sigma_{ys} = 0.925$ GPa
- Projected area: $A_c = 3{,}822.9$ μm^2
- Equivalent contact radius: $a_s = 34.88$ μm
- Indenter shape parameter: $R = 54.94$ μm
- Plastic zone depth: $Z_{ys} = 87.05$ μm
- Parameter c: $c = 129.51$ μm
- **Surface plastic zone radius:** $C_s = 122.5$ μm
- **Surface plastic zone area:** $A_{pz} = 47{,}127$ μm^2

This C_s value defines the radius of the circular region around each indentation whose microstructure physically governs the measured hardness, and is used to determine the ROI crop size for dataset construction (Section III-B).

B. Phase Mask Extraction Pipeline Detail

The six-stage classical CV pipeline used to extract binary phase masks from OM images is described below, with all processing parameters listed in Table III.

Stage 1 — Denoising. A median filter with kernel size $k = 5$ is applied to the grayscale image to suppress salt-and-pepper noise while preserving phase boundary sharpness.

Stage 2 — Background normalization. A large-sigma Gaussian blur ($\sigma = 80$) estimates the slowly varying background illumination pattern. The original image is divided pixel-wise by this estimate and rescaled to a reference intensity of 128, correcting uneven illumination across the field of view.

Stage 3 — Contrast enhancement. Contrast limited adaptive histogram equalization (CLAHE) with clip limit 3.0 and tile grid 8×8 is applied to enhance local phase contrast.

Stage 4 — Smoothing and binarization. A Gaussian blur (5×5, $\sigma = 1$) smooths residual high-frequency noise, after which Otsu's adaptive thresholding produces the initial binary mask.

Stage 5 — Morphological refinement. Morphological opening (elliptical kernel 3×3, 2 iterations) removes small isolated noise regions, followed by morphological closing (elliptical kernel 5×5, 2 iterations) to fill voids within phase regions.

Stage 6 — Small region filtering. Connected components in the dark phase with area below 200 px are reclassified as light phase via contour-based filtering, eliminating residual segmentation artifacts.

TABLE III
CV Pipeline Processing Parameters

Stage	Parameter	Value
Median filter	Kernel size	5
Background norm.	Gaussian σ	80
CLAHE	Clip limit	3.0
	Tile grid	8×8
Gaussian smooth	Kernel size	5×5
	σ	1
Morph. opening	Kernel (ellipse)	3×3
	Iterations	2
Morph. closing	Kernel (ellipse)	5×5
	Iterations	2
Region filter	Min. area	200 px

References

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[2] J. Stuckner, B. Harder, and T. M. Smith, "Microstructure segmentation with deep learning encoders pre-trained on a large microscopy dataset," *npj Comput. Mater.*, vol. 8, p. 200, 2022.

[3] M. V. Pathan, S. A. Ponnusami, J. Pathan, R. Pitisongsawat, B. Erice, N. Petrinic, and V. L. Tagarielli, "Predictions of the mechanical properties of unidirectional fibre composites by supervised machine learning," *Sci. Rep.*, vol. 9, p. 13964, 2019.

[4] Y. Sun, I. Hanhan, M. D. Sangid, and G. Lin, "Predicting mechanical properties from microstructure images in fiber-reinforced polymers using convolutional neural networks," *J. Compos. Sci.*, vol. 8, no. 10, p. 387, 2024.

[5] D. W. Abueidda *et al.*, "Explainable prediction of the mechanical properties of composites with CNNs," *arXiv preprint*, 2025, arXiv:2505.14745.

[6] M. Larmuseau *et al.*, "Towards accurate processing-structure-property links using deep learning," *Scripta Mater.*, vol. 211, p. 114490, 2022.

[7] R. Kondo, S. Yamakawa, Y. Masuoka, S. Tajima, and R. Asahi, "Microstructure recognition using convolutional neural networks for prediction of ionic conductivity in ceramics," *Acta Mater.*, vol. 141, pp. 29–38, 2017.

[8] G. Lambard, T. Yamazaki, and M. Demura, "Generation of highly realistic microstructural images of alloys from limited data with a style-based generative adversarial network," *Sci. Rep.*, vol. 13, p. 441, 2023.

[9] A. Cheloee Darabi, S. Rastgordani, M. Khoshbin, V. Guski, and S. Schmauder, "Hybrid data-driven deep learning framework for material mechanical properties prediction with the focus on dual-phase steel microstructures," *Materials*, vol. 16, no. 1, p. 447, 2023.

[10] L. Wang *et al.*, "Deep learning-driven microstructure characterization and Vickers hardness prediction of Mg-Gd alloys," *arXiv preprint*, 2024, arXiv:2410.20402.

[11] P. B. Yekta *et al.*, "Machine learning of microstructure–property relationships in materials leveraging microstructure representation from foundational vision transformers," *Comput. Mater. Sci.*, 2025, arXiv:2501.18637.

[12] K. Alrfou, T. Zhao, and A. Kordijazi, "CS-UNet: A flexible segmentation algorithm for microscopy images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2024.

[13] Ł. Marcjan, S. Gajoch, D. Wilk-Kołodziejczyk, M. Małysza, K. Jaśkowiec, and G. Gumienny, "Enhancing segmentation of irregular microstructural elements using extended channel information and transfer learning," in *Database and Expert Systems Applications (DEXA 2025)*, ser. Lecture Notes in Computer Science, vol. 16047. Springer, 2026.

[14] Y. Xue *et al.*, "A deep learning convolutional neural network and multi-layer perceptron hybrid fusion model for predicting the mechanical properties of carbon fiber," *Composites Part A*, vol. 171, p. 107561, 2023.

[15] A. Y.-T. Wang, S. K. Kauwe, R. J. Murdock, and T. D. Sparks, "Compositionally restricted attention-based network for materials property predictions," *npj Comput. Mater.*, vol. 7, p. 77, 2021.

[16] S. Takeshita *et al.*, "Interpretable deep learning for XRD pattern analysis in lithium-ion batteries: CNN and attention-based feature extraction," *J. Power Sources*, 2025.

[17] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Rev. Phys.*, vol. 3, pp. 422–440, 2021.

[18] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 128, pp. 336–359, 2020.

[19] W. Chen *et al.*, "Explainable machine learning in materials science," *npj Comput. Mater.*, vol. 8, p. 204, 2022.

[20] F. Oviedo, J. L. Ferres, T. Buonassisi, and K. T. Butler, "Interpretable and explainable machine learning for materials science and chemistry," *Accounts Mater. Res.*, vol. 3, no. 6, pp. 597–607, 2022.

[21] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, "Sanity checks for saliency maps," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 31, 2018, pp. 9525–9536.

[22] T. A. DeBold, "Duplex stainless steel—microstructure and properties," *JOM*, vol. 41, pp. 12–15, 1989.

[23] R. N. Gunn, *Duplex Stainless Steels: Microstructure, Properties and Applications*. Cambridge, U.K.: Woodhead Publishing, 1997.

[24] K. R. Gadelrab, G. Li, M. Chiesa, and T. Souier, "Local characterization of austenite and ferrite phases in duplex stainless steel using MFM and nanoindentation," *J. Mater. Res.*, vol. 27, pp. 1573–1579, 2012.

[25] P. Tao, J.-M. Gong, Y.-F. Wang, Y. Jiang, Y. Li, and W. Cen, "Characterization on stress-strain behavior of ferrite and austenite in a 2205 duplex stainless steel based on nanoindentation and finite element method," *Results Phys.*, vol. 11, pp. 377–384, 2018.

[26] H. Xie *et al.*, "Micro-deformation evolutions of the constituent phases in duplex stainless steel during cyclic nanoindentation," *Sci. Rep.*, vol. 8, p. 6478, 2018.

[27] E. S. de Argandoña *et al.*, "Effect of the temperature in the mechanical properties of austenite, ferrite and sigma phases of duplex stainless steels using hardness, microhardness and nanoindentation techniques," *Metals*, vol. 7, no. 6, p. 219, 2017.

[28] E. S. de Argandoña *et al.*, "Mechanical characterization at nanoscale of austenite, ferrite, and sigma phases via hardness measurement and fretting wear behavior of a duplex stainless steel," *Metals*, vol. 13, no. 5, p. 864, 2023.

[29] G. Fargas, M. Anglada, and A. Mateo, "Effect of the annealing temperature on the mechanical properties, formability and corrosion resistance of hot-rolled duplex stainless steel," *J. Mater. Process. Technol.*, vol. 209, no. 4, pp. 1770–1782, 2009.

[30] M. Mata, O. Casals, and J. Alcalá, "The plastic zone size in indentation experiments: The analogy with the expansion of a spherical cavity," *Int. J. Solids Struct.*, vol. 43, pp. 5994–6013, 2006.

[31] K. L. Johnson, *Contact Mechanics*. Cambridge, U.K.: Cambridge University Press, 1985.

[32] D. Tabor, *Hardness of Metals*. United Kingdom: Clarendon Press, 1951.