

## 機械学習を用いた金属組織写真から硬さの予測

材料技術室 石川 隆朗, 藤枝 正之

## Predicting Hardness from Metallurgical Microstructure Photographs Using Machine Learning

Takaaki ISHIKAWA, Masayuki FUJIEDA

現在, 人工知能分野で機械学習技術が発展している。機械学習技術を材料工学分野に応用するにあたり, 金属の硬さ, 靱性等の情報を豊富に含んでいる金属組織写真に適用することを考えた。金属の特徴の指標として金属硬さを採用し, 深層学習を用いた回帰分析により金属組織写真から硬さの予測を行った。深層学習モデルの選定, 画像前処理(画像拡張処理), 予測結果の統計処理が, 機械学習予測値に与える影響を調査した。複数画像の統計処理を行うことにより金属硬さ実測値と機械学習予測値は良い一致を見せた。機械学習により金属組織写真の特徴を捉えられたと考える。

## 1. はじめに

2012年にGeoffrey Hintonが画像認識コンテストILSVRCで深層学習技法を用い優勝したことは大きなインパクトを与えた。その後, 深層学習技術は様々な分野に応用されている。

深層学習の技術はKerasやPyTorch等のライブラリとして整理され, その情報は郊外の小さな書店でも容易に手に入るようになってきている。現在では, 計算機科学を修めていないものづくり分野の技術者でも手軽に扱えるようになってきている。計算に必要なPCも家電量販店で手軽に手に入るものとなった。

金属組織写真は, 焼入れを始めとする熱処理の情報を豊富に含んでいる。金属組織写真を見ることによって金属の硬さ, 靱性等を判断することができる。

機械学習技術の適用として一般的に広く行われている動物, 自動車等の物体の画像識別に用いられる写真と異なり, 金属組織写真は全体に一樣な模様が広がっている画像である。しかし, マルテンサイトとソルバイト等, 人間の目でも判別つかない微量な差異を認識しなければならない。また, 金属組織写真には, 金属組織の情報以外にも介在物や材料の欠陥も含まれている。そのため, 画像前処理, 画像拡張は物体の認識とは違った考え方が必要となる。

本研究は金属組織写真に深層学習を適用し, 有用な情報を得ることができるかどうか調査することを目的とした。金属組織写真から得る特性値の指標として金属硬さを採用した。マイクロビッカース硬さの実測値と深層学習により得られた予測値との比較を行った。

## 2. 実験方法

## 2.1 金属組織写真

測定で扱う金属材料は, ホームセンター等で販売されている工具等の鉄鋼試料を用いた。

試料を切断後, 樹脂に埋込し, 研磨を行った。株式会社ミットヨ製マイクロビッカース硬さ試験機HM-220Dにより硬さの測定を行った。その後2%ナイトル溶液でエッチングを行い, 株式会社キーエンス製三次元顕微鏡VHX-1000/1100により金属組織写真を取得した。顕微鏡の倍率は400倍で, 撮影領域は $800\mu\text{m} \times 600\mu\text{m}$ であった。画像は $3200 \times 2400\text{pixel}$ である。

それぞれの試料は研磨, エッチング後に20か所以上を撮影した。また, いくつかの試料は再研磨, エッチング後にさらに20か所以上の写真撮影を行い, 最終的に24試料, 874枚の金属組織写真を取得した。取得した金属組織写真の例を図1に示す。



図1 取得金属組織写真例

## 2.2 計算環境

深層学習計算は主記憶 128GiByte, GPU はメモリ 11GiByte の GTX1080Ti を搭載した計算機を用いた。OS は Ubuntu GNU/Linux を用いた。

計算は Python 言語を用いた。深層学習ライブラリは PyTorch ライブラリを用いた。

深層学習ネットワークは PyTorch ライブラリに標準で提供されているものを用いた。オーソドックスな畳み込みニューラルネットワークである VGG19 とトランスフォーマーネットワークを画像に応用した ViT\_B\_32 を利用した。画像の分類ではなく、回帰分析であるので最終出力層を1ユニットの線形出力層に変更し、損失関数は平均絶対誤差を採用した。

計算再現性の為、乱数のシード値は画像ファイルの切り出しには画像ファイルの MD5 ハッシュ値を、深層学習計算には 0 を採用した。

## 2.3 画像処理

全ての金属試料からそれぞれ 1~2 枚の画像を抜き取り、検証、評価用画像とした。838 枚を学習に、36 枚を検証及び評価用に用いた。

3200×2400pixel の画像より、ランダムに 224×224pixel の領域を切り出し、学習、評価画像として用いた。1 枚の金属組織写真より 100 枚の画像の取得を行った。

深層学習モデルの検証、評価用画像を分けるのとは別に、深層学習過程において、学習データと検証データは 8:2 に分けて学習を行った。

## 3. 計算結果

### 3.1 硬さ予測値の算出

事前学習済み ViT\_B\_32 モデルを 10 エポックファインチューニングした後、そのモデルを用いて、検証用画像から切り出した画像から硬さを予測した結果を図 2 に示す。それぞれの金属組織写真から 100 枚切り出したため、一つの組織写真につき 100 点のプロットがある。一見、プロットは大きく広がっているように見える。ある金属組織写真の予測硬さの分布をヒストグラムで表したものを図 3 に示す。

ほとんどの予測硬さは 230~240 の間に含まれ、平均から離れたものはごく少数であることがわかる。1 枚の金属組織写真中から得られた 100 枚の画像は、ほとんどは妥当な硬さ予測値を算出するが、一部、誤認識を起こす画像が存在してしまうためだと考

えられる。そこで、1 枚の画像から得られた 100 の予測硬さの平均を取り、その平均から標準偏差  $1\sigma$  以上離れているものを除外し、平均を取ることとした。そのように処理したもののグラフを図 4 に示す。

ここで、決定係数

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

$y_i$ : 実測値

$\hat{y}_i$ : 予測値

$\bar{y}$ : 実測値の平均値

を計算したところ 0.9961 であった。外れ値の除外を行わず、単純に平均をとったものの決定係数  $R^2$  の値は 0.9931 であった。

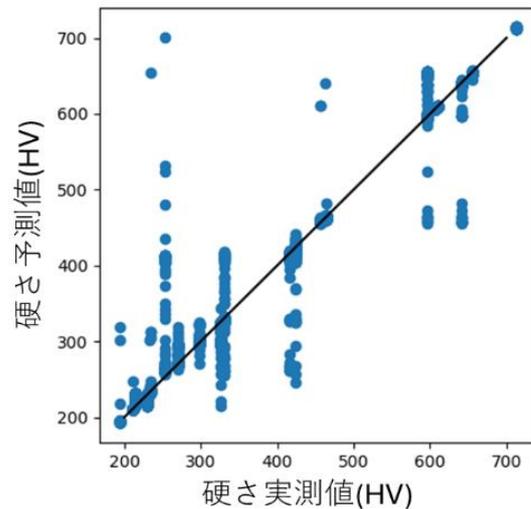


図 2 硬さ実測値と予測値の比較

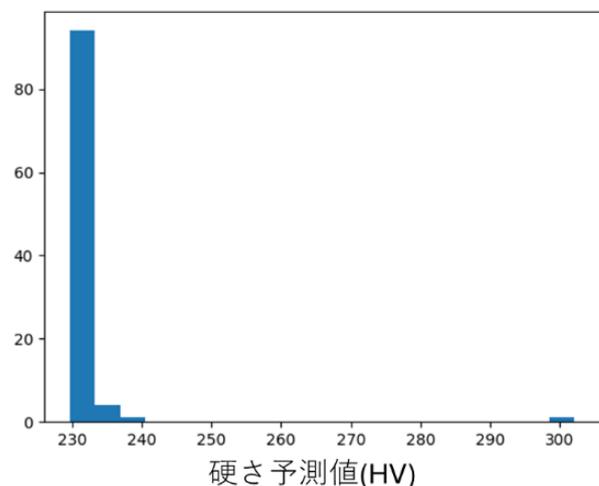


図 3 予測硬さのヒストグラム

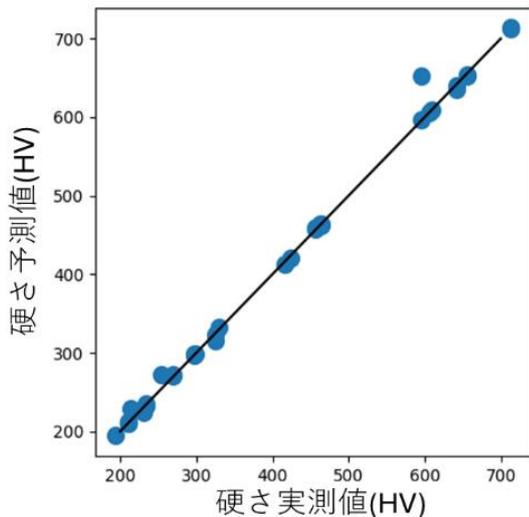


図4 外れ値を除外し、平均を取った硬さ実測値と予測値の比較

### 3.2 深層学習モデルの比較

深層学習モデル、VGG19 と ViT\_B\_32 の比較を行った。それぞれ、PyTorch 標準配布の事前学習済みモデルからファインチューニングを行ったもの、乱数で初期化されたモデルを学習したものの比較を行った。計算時間は、VGG19 を 100 エポック学習させるのに約 20 時間、ViT\_B\_32 を 100 エポック学習させるのに約 10 時間要した。

それぞれの学習過程における損失値、学習損失 (train loss), 検証損失 (val loss) の推移を事前学習なし VGG19 のものを図 5 (a) に、事前学習を済み VGG19 のものを図 5 (b) に、事前学習なしの ViT\_B\_32 のものを図 5 (c) に、事前学習を済み ViT\_B\_32 のものを図 5 (d) に示す。VGG19, ViT\_B\_32 両方とも事前学習済みのモデルを用いたものは初期の train loss, val loss とともに少ない値からスタートしているように見える。また、VGG19 の方が ViT\_B\_32 と比較して、滑らかに損失値が減少しているように見える。

決定係数  $R^2$  の値を VGG19 のものに対して表 1 (a) に、ViT\_B\_32 のものに対して表 1 (b) に示す。ViT\_B\_32 に関しては事前学習済みモデルを用いることにより、10 エポックでの予測値が大きく向上した。VGG19, ViT\_B\_32 双方とも、100 エポック学習させたものでは、事前学習なしモデルから学習を開始したものの方が予測値は正確だった。最も予測値が不正確だった事前学習なしで 10 エポック学習させた ViT\_B\_32 モデルの実測値と予測の対比のグラフを図 6 に示す。

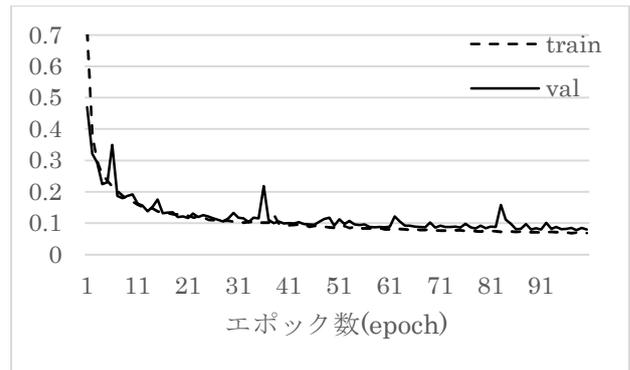


図5 (a) 事前学習なしの VGG19 モデルの学習時の損失値の推移

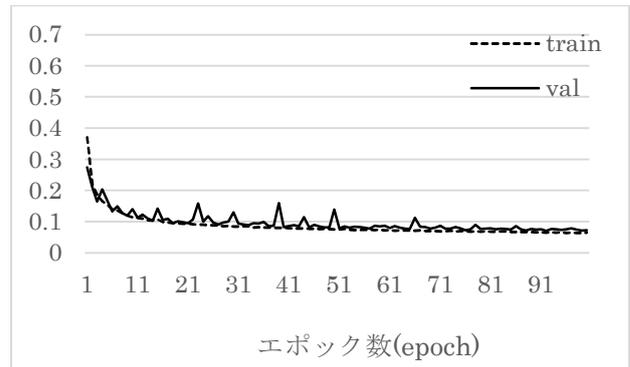


図5 (b) 事前学習済み VGG19 モデルのファインチューニング時の損失値の推移

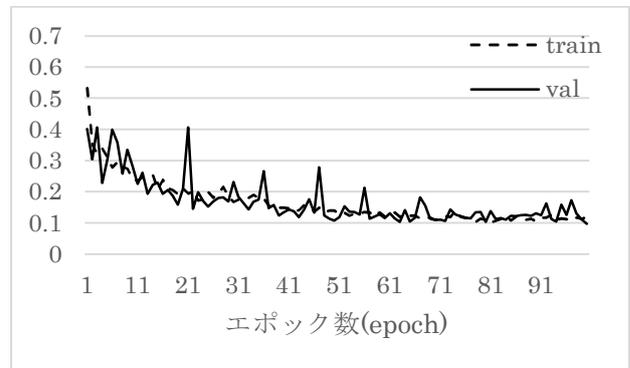


図5 (c) 事前学習なしの ViT\_B\_32 モデルの学習時の損失値の推移

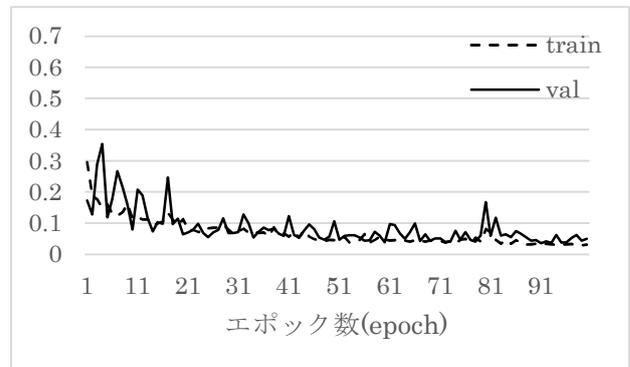


図5 (d) 事前学習済み ViT\_B\_32 モデルのファインチューニング時の損失値の推移

表 1 (a) VGG19 ネットワークの予測値の決定係数  $R^2$

	10 エポック	100 エポック
事前学習なし	0.9951	0.9995
事前学習済み	0.9903	0.9967

表 1 (b) ViT\_B\_32 ネットワークの予測値の決定係数  $R^2$

	10 エポック	100 エポック
事前学習なし	0.8202	0.9991
事前学習済み	0.9961	0.9977

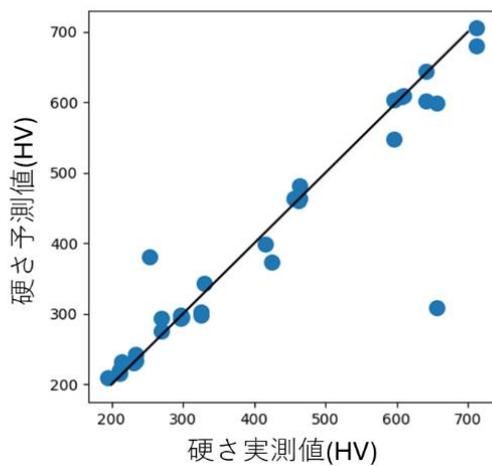


図 6 事前学習なし ViT\_B\_32 モデルを 10 エポック学習させたものの硬さ実測値と深層学習予測値との比較

### 3.3 学習画像領域サイズの影響

3 章 2 節までは  $800\mu\text{m} \times 600\mu\text{m}$  の領域を  $3200 \times 2400\text{pixel}$  で撮影した画像を、 $224 \times 224\text{pixel}$  の大きさに切り出して学習、予測を行った。これは  $60\mu\text{m}$  四方の領域の画像で学習、予測を行ったことになる。

大きな領域を切り出し、学習を行なった場合、微細な特徴を失う可能性がある。また、小さな領域で学習を行なった場合、大きな構造を学習に反映させられない可能性もある。1 枚の金属組織写真から大きな画像を切り出す場合、取得する画像のバリエーションが少なくなる問題も考えられる。

$224 \times 224\text{pixel}$  ( $60\mu\text{m}$  四方) ~  $1120 \times 1120\text{pixel}$  ( $300\mu\text{m}$  四方) まで学習、予測画像サイズを変化させ、硬さの予測を行わせた。事前学習済み ViT\_B\_32 モデルを用い、10 エポックファインチューニングを

行った。その結果を表 2, 図 7 に示す。学習、予測画像幅が  $672 \times 672\text{pixel}$  ( $180\mu\text{m}$ ) を超えたあたりから決定係数  $R^2$  の値が悪化することがうかがえる。

図 8 に最も決定係数  $R^2$  の値が悪かった画像幅  $1120 \times 1120\text{pixel}$  ( $300\mu\text{m}$ ) の硬さ実測値と深層学習予測値の比較のグラフを示す。硬さ実測値 400 未満の領域でのずれが見られる。結晶構造が微細な硬さの大きい領域は、一部外れ値があるものの、それ以外は大きなずれは見られなかった。

表 2 学習、予測画像サイズと予測値の決定係数  $R^2$  の関係

学習、予測領域	決定係数 $R^2$
$224 \times 224$ ( $60\mu\text{m}$ )	0.9961
$448 \times 448$ ( $120\mu\text{m}$ )	0.9944
$672 \times 672$ ( $180\mu\text{m}$ )	0.9964
$896 \times 896$ ( $240\mu\text{m}$ )	0.9702
$1120 \times 1120$ ( $300\mu\text{m}$ )	0.9160

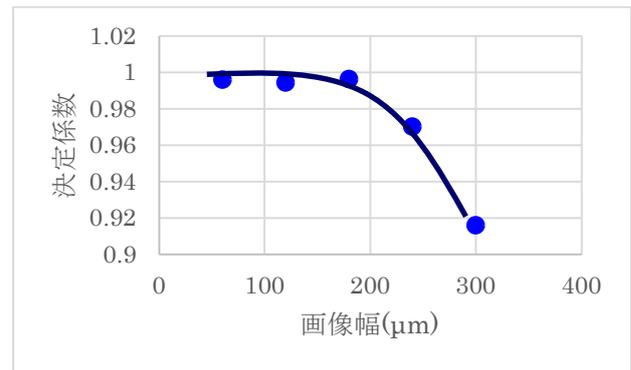


図 7 学習、予測画像幅と決定係数  $R^2$  の関係

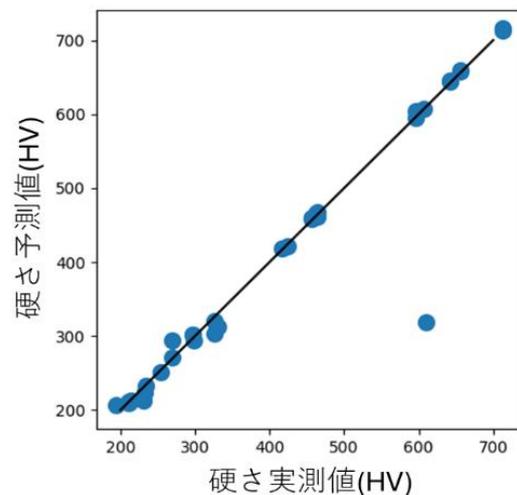


図 8 画像幅  $300\mu\text{m}$  で学習を行なったものの硬さ実測値と深層学習予測値との比較

### 3.4 画像回転処理

撮影された金属組織写真のうちいくつかは、鍛造等の影響により結晶の形状に異方性があるものがある。また、マルテンサイトの筐状の微細構造は統計的に十分な数が得られない場合、異方性と捉えられる可能性もある。

そこで、切り出す画像に回転処理を加え、異方性の影響を軽減することを考えた。

正方形の対角線の長さ比である $\sqrt{2}$ を超える1.5を学習、予測画像サイズに乗じた領域を金属組織写真から切り出し、ランダムな角度に回転させ、中心から学習、予測画像を切り出し、学習を行なった。深層学習モデルは事前学習済み ViT\_B\_32 モデルを用い 10 エポックファインチューニングを行った。結果を表3に示す。

表3 硬さ実測値と予測値の決定係数  $R^2$  の関係に与える画像回転処理の影響

学習, 予測領域	回転なし	回転あり
224×224(60 $\mu$ m)	0.9961	0.9974
448×448(120 $\mu$ m)	0.9944	0.9404
672×672(180 $\mu$ m)	0.9964	0.9938

### 3.5 未学習試料の硬さ予測

3章4節までは全試料を対象に学習用画像と検証・評価用画像の両方を得て、検討を行ってきた。そこで次に、検証・評価用画像と同一試料の画像を学習に一切使用せず深層学習モデルを得て、未学習試料の硬さ予測を行った場合、学習されているモデルを使用した場合に比べて予測精度にどのような違いがあるか確認を行った。学習には19試料 658枚画像を用いて深層学習モデルを得て、5試料 216枚を検証・評価画像として使用した。

未学習試料の硬さ予測では、モデルのファインチューニング量によって結果が大きく異なり、100エポックファインチューニングすることで結果が改善された。学習、予測領域、回転処理の有無を変え、事前学習済み ViT\_B\_32 を100エポックファインチューニングしたものの硬さ実測値と深層学習予測値の決定係数  $R^2$  の関係を表4に示す。また、画像領域幅 60 $\mu$ m の、回転なしの条件で学習を行なったものの硬さ実測値と深層学習予測値の対比を図9に示す。

1回の研磨、エッチングを行った試料から複数の金属組織写真を取得している。図9のプロット

一つ一つはある試料の一つの金属組織写真に対応している。同一研磨、エッチング後の金属組織写真を1グループとして扱い整理した。これまで一つの金属組織写真から切り出した100枚の画像からの予測値それぞれに対して、棄却、平均を行ったが、同グループの金属組織写真から得られた画像をまとめて上記統計処理を行った。結果を表5に示す。また、画像領域幅 60 $\mu$ m、回転なしの条件で学習を行なったものの硬さ実測値と深層学習予測値の対比を図10に示す。

表4 学習に用いなかった試料の硬さ実測値と予測値の決定係数  $R^2$  の関係

学習, 予測領域	回転なし	回転あり
224×224(60 $\mu$ m)	0.8831	0.9076
448×448(120 $\mu$ m)	0.9241	0.9094
672×672(180 $\mu$ m)	0.8352	0.8736

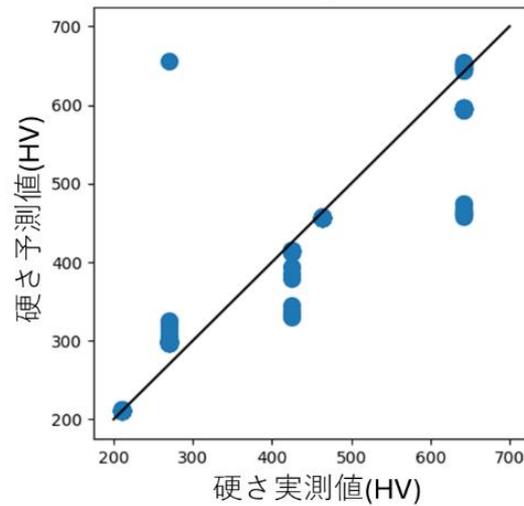


図9 学習データに用いられない試料の硬さ実測値と金属組織写真からの硬さ予測値との比較

表5 複数の画像の統計処理値を深層学習予測値とした場合の決定係数  $R^2$

学習, 予測領域	回転なし	回転あり
224×224(60 $\mu$ m)	0.9822	0.9497
448×448(120 $\mu$ m)	0.9832	0.9731
672×672(180 $\mu$ m)	0.8765	0.9378

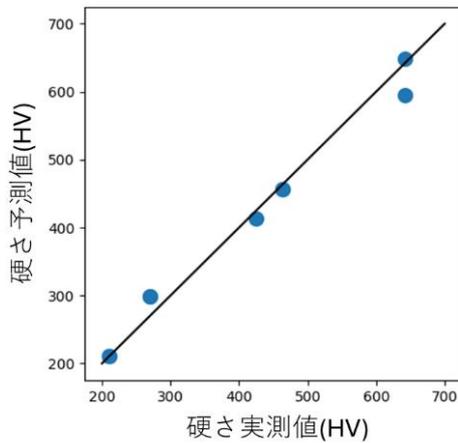


図10 複数の金属組織写真をグループ化し、統計処理を行った場合の硬度実測値と硬度予測値の比較

#### 4. まとめ

本研究は金属組織写真を機械学習技法で学習を行なうことにより、金属組織の特徴を抽出できるかどうかを確認することを目的とした。その指標

として金属硬さを用いた。深層学習モデルの選定、様々な画像前処理が機械学習に与える影響を調査した。また、統計処理を行うことによる予測値の変化を示した。複数画像から得られた結果を統計処理することにより、金属硬さ実測値と機械学習による予測値は良い一致を見せた。

同一試料の組織写真が学習データに入っていれば、学習に用いられなかった組織写真から精度よく硬さを予測することができた。これは、差がわかりづらい金属組織の特徴を認識し、判別しているからである。学習データに用いられない試料の硬さ予測を行った場合、結果はばらついたが、学習する試料数を増やし、金属組織の特徴を網羅することにより、未知試料の硬さの予測も精度よく行うことが期待できる。

金属組織写真の特徴を認識し捉えることができたことが確認された。今後、硬さ以外の物性値の予測、故障等のトラブル解析等に应用することが期待できる。