




Enhancing machine learning classification of microstructures: A workflow study on joining image data and metadata in CNN

Marie Stiefel , Department of Materials Science, Saarland University, 66123 Saarbrücken, Germany
Martin Müller , and **Björn-Ivo Bachmann** , Department of Materials Science, Saarland University, 66123 Saarbrücken, Germany; Material Engineering Center Saarland, 66123 Saarbrücken, Germany
Maria Agustina Guitart , and **Ullal Pranav Nayak** , Department of Materials Science, Saarland University, 66123 Saarbrücken, Germany
Frank Mücklich , Department of Materials Science, Saarland University, 66123 Saarbrücken, Germany; Material Engineering Center Saarland, 66123 Saarbrücken, Germany

Address all correspondence to Marie Stiefel at marie.stiefel@uni-saarland.de

(Received 5 January 2024; accepted 21 March 2024)

Abstract

In view of the paradigm shift toward data-driven research in materials science and engineering, handling large amounts of data becomes increasingly important. The application of FAIR (findable, accessible, interoperable, reusable) data principles emphasizes the importance of metadata describing datasets. We propose a novel data processing and machine learning (ML) pipeline to extract metadata from micrograph image files, then combine image data and their metadata for microstructure classification with a deep learning approach compared to a classic ML approach. The ML model attained excellent performances with and without metadata and bears potential for performance improvement of further use cases within the community.

Introduction: motivation

Research in materials science and engineering (MSE) is subject to a paradigm shift with data-driven science emerging as a new field and big data being the new resource for scientific breakthroughs.^[1,2] In order to handle large amounts of data and, more importantly, enabling new discoveries through data exploration, good data management is crucial.^[3] In 2016, Wilkinson et al. have proposed the FAIR guiding principles for research data management and stewardship, designed by various stakeholders from academia, industry, publishing, and funding.^[3] Since then, the GO FAIR initiative was founded, aiming to implement the FAIR data principles (go-fair.org), and numerous others have decided to lead by example (e.g., FAIRmat^[4]). In order to satisfy the FAIR guidelines, research data must be Findable, Accessible, Interoperable, and Reusable (go-fair.org/fair-principles), as prerequisites for widespread usability in the scientific community. Scheffler et al. reinterpreted the acronym as ‘Findable and AI-ready,’ properties of research datasets they deem indispensable for future scientific research, especially regarding their use in AI or machine learning applications.^[4] In order for research data to be considered as FAIR, an important property is their richness in metadata, on which the community puts a special focus.^[5,6] Metadata provides information on the dataset itself and its main purpose is to facilitate organization and findability of the data, by supplying, e.g., details on the authors, purpose of the data, or the data acquisition.^[7] The thorough collection and usage of metadata aim to increase the reproducibility of research findings, also enabling other researchers within the community to reuse the data and

reproduce experiments, which unlocks the potential to greatly ameliorate collaboration within the community as well as to facilitate reviews and verification of research findings in order to ensure high-quality research.^[4,7] In addition, with the emergence of artificial intelligence (AI) and machine learning (ML) and the reuse of research data, it is of relevance to investigate the role metadata can play in machine learning applications.

A significant application of machine learning in materials science is the use of convolutional neural networks (CNN)—a type of artificial neural networks for image analysis—for micrograph analysis, especially microstructure characterization, as performed by,^[8–12] among others. In fact, this is fundamental in the investigation of the correlation between materials microstructures and mechanical properties. The microstructure, referring to the inner structure of a material, contains a myriad of information regarding, on one hand, its genesis and processing history, and, on the other, its chemical, physical, and mechanical properties.^[12] Therefore, the microstructure is considered the central information carrier of a material,^[13] and an in-depth analysis of the microstructure and the phases contained in it is decisive to understand process–microstructure–property relationships. Considering the paradigm change from empirical process–property correlation to microstructure-based development of new materials^{1,2} in MSE, an in-depth analysis and understanding of the microstructure are all more important for materials and process optimization. Similarly, for ML-based materials discovery and design, a focus on possible novel chemical compositions is not sufficient and a consideration of the microstructure is essential,^[14] since the microstructure acts

as the central pillar when establishing relationships between processing and material properties. Typically, microstructure is analyzed manually, which can be time-consuming and produce qualitative results only, often accompanied by a high degree in subjectivity, making the analysis a bottleneck in microstructure-based materials development. With ever-improving imaging techniques and the recent growth of AI use, we are faced with a new set of possibilities when it comes to machine learning applications for microstructure analysis.^[13,15–18] For example, ML-supported microstructure analysis has been conducted on steel,^[12,19,20] nickel-base alloys,^[21] ceramics,^[22] or metal powder.^[23] These image analysis tasks can be categorized into image classification, (semantic) segmentation, object detection, and classification.^[23] The most commonly used neural network for microstructure analysis are CNN (including, but not limited to VGG,^[24] Inception,^[25] and Xception^[26]), densely connected neural networks (DenseNet^[27]), and deep residual networks (ResNet^[28]).

In this use case, research data are available in the form of image datasets, such as, for example, the ASM Micrograph Database,^[29] the NEU-DET surface defect database,^[30] or the Ultra-High-Carbon Steel Micrograph Database.^[31] For the case of micrographs, Kemmer et al. and Huisman et al. argue that imaging datasets become fully valuable and usable

only when they are accompanied by abundant metadata.^[6,32] In fact, this renders datasets reusable by other members of the community, as they have at their disposition all necessary complementary information allowing them to utilize the dataset and fully understand its significance. Ghiringelli et al. emphasize that the reusability aspect of FAIR data especially refers to repurposing the research data, which is only possible if the dataset is accompanied by rich metadata.^[33]

With respect to FAIR research data, we propose the following study on a data management workflow designed for usage of micrograph datasets for CNN classification tasks. An approach on the combination of image data and patient metadata for skin lesion images was proposed by Nunnari et al. in the context of the ISIC 2019 skin lesion classification challenge dataset, combining images and patient metadata which partially led to a significant improvement of the classification model performance^[34] and inspired the present work. In fact, with the recent debate on the cruciality of metadata in research datasets, we explore a novel data processing and machine learning pipeline whose central pillar is the concatenation of image (or pixel) data and metadata in a classification task, complemented by a study of the classification model performance assessing the effect of the added metadata on micrograph datasets (Fig. 1).

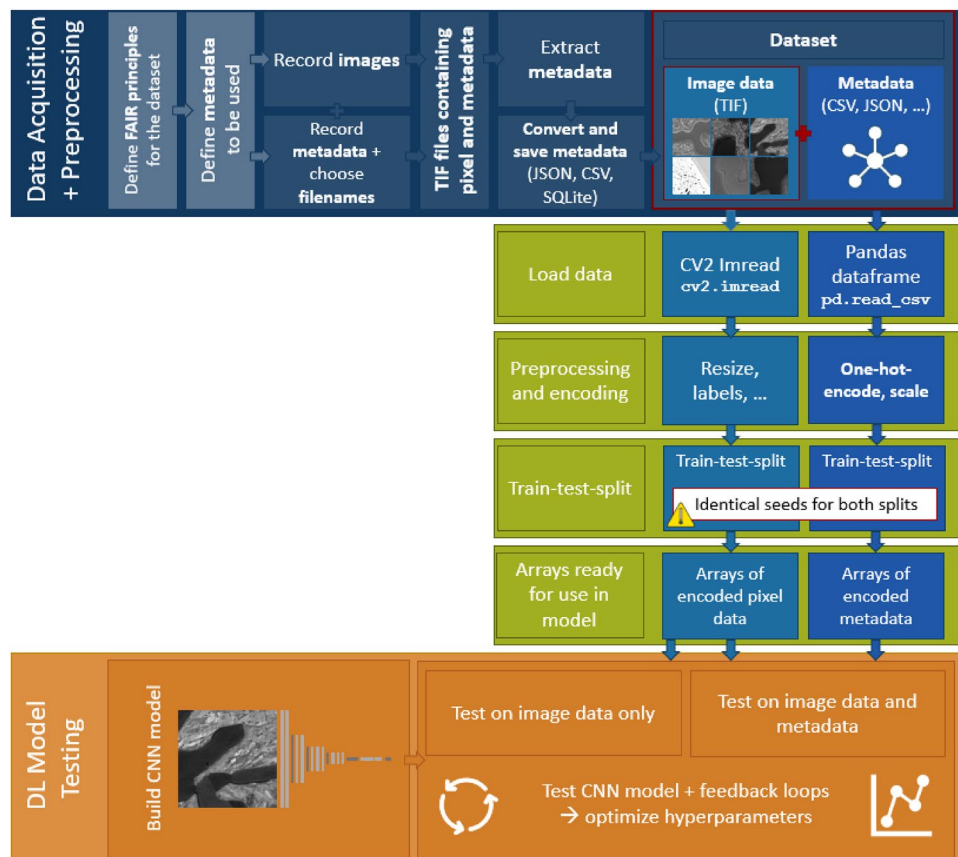


Figure 1. General workflow for the establishment of a machine learning model including a preliminary data treatment pipeline.

Materials and methods: FAIR datasets and AI in materials science

Workflow

To even get to the point where image and metadata can be combined in an ML model, a preliminary data processing pipeline is required, where all the necessary data are transferred into a processable format. Assuming that a complete dataset does not yet exist, the first step is to gather the required image data. In this case, we anticipate that a number of scanning electron microscopy (SEM) images was already recorded and exist in the TIF file format, which contains both image (or pixel) information and a range of additional information in the form of strings or numerical values, arranged in the form of a dictionary, automatically saved upon exporting recorded images in the TIF file format. Furthermore, we assume that any crucial information regarding the specimen the image is taken from is also present, *e.g.*, in tabular form, stored in the TIF file, or as substring in the filename of the image file. On that basis, metadata on the specimen, the imaging (*i.e.*, SEM parameters), and the image itself are extracted from the TIF file. The metadata extraction process differs with the version of the TIF file, the form of the dictionary used, and the manufacturer of the imaging system, but is generally enabled by the TiffFile Python library.^[35] For example, for Carl Zeiss SEM TIF images, metadata extraction is performed using the function `sem_metadata`, while for FEI Company (Field Electron and Ion Company, Hillsboro, OR, USA, subsidiary of Thermo Fisher Scientific Inc.) SEM, the library provides a function titled `fei_metadata`. The metadata is then converted to a tabular form which can be used in the machine learning algorithm, while the image data can remain in its original form (*i.e.*, TIF file format). A convenient file format for metadata storage is JSON; others may include SQLite or CSV, with JSON being, for example, suggested by Aversa *et al.* who propose a GUI-supported mapping service for Carl Zeiss SEM metadata.^[36]

Image data are loaded using the OpenCV `cv2.imread` module, and metadata is transformed from the CSV file in to a Pandas dataframe using `pd.read_csv`. The metadata is then pre-processed by scaling numerical values and one-hot encoding categorical values. For the ground truth assignment in the form of class labels, in our use case, the dataset is subdivided into folders, each containing one class. The classes are assigned to the images and then converted to categorical data. The metadata is assigned to the respective images via the image filename which is a unique identifier for the image.

The CNN classification model consists of two branches, one being fed the image data and the other the encoded tabular data. Both branches contain one or more convolutional layers, and the obtained feature vectors are concatenated and fed into a dense classification layer. Figure 2 shows a comparison of a single-branch (image data only) and a two-branch (image data and metadata) CNN classification model. The image classification is performed with a

ResNet50 backbone pre-trained on the ImageNet database (non-trainable weights), followed by a 2D global average pooling layer, a batch normalization layer, and two dense layers (128 and 32 units, respectively), respectively, including dropout layers in between. For the model with metadata, the metadata is treated using a dense layer (with a dropout layer) and then both branches of the model are concatenated. The last layer is the classification layer. The example model was created for the UHCSDB (Ultra-High-Carbon Steel Micrograph Database),^[31] which, after pre-processing the metadata (*i.e.*, one-hot encoding categorical data), comprises 14 distinct features fed into the metadata branch, compared to 32 convolutional features.

Metadata collection and integration

The rich metadata which is required for scientific datasets to satisfy the FAIR guidelines generally consists of information regarding the author, time, and context of the dataset creation. However, metadata can go beyond this—for example, micrograph datasets can be complemented by information regarding the specimen, such as its chemical composition, heat treatment, and pre-imaging preparation. Therefore, we gathered a wide range of metadata for our micrograph dataset. A comparable approach was made by DeCost *et al.* who proposed the Ultra-High-Carbon Steel Micrograph Database (UHCSDB) and provided various metadata to add to the micrographs in their dataset.^[31] They provided information on the magnification (including the micron bar) and the detector (imaging information) and on annealing time and temperature as well as the quenching method (specimen information). The primary microconstituent was used as class for the classification task.

For our dataset, consisting of SEM images of high chromium cast iron (HCCI), the metadata comprised (i) image metadata: brightness and contrast (as provided by the SEM metadata), (ii) imaging metadata: SEM detector, accelerating voltage, beam current, working distance, and physical pixel width, and (iii) specimen metadata: chromium content, heat treatment, temperature of heat treatment, quenching method, etchant for specimen preparation, and etching time. Metadata concerning authors and context has been collected as well in order to complement the dataset for publishing, but it is not taken into account by the machine learning model.

Following the concept of data repurposing as stated in the FAIR principles, we made use of the large number of SEM micrographs of HCCI specimens that had been previously recorded for other purposes^[37,38] and gathered them in a new dataset of images enriched with the above-mentioned metadata. HCCI, a cast iron of the Fe–C–Cr ternary system with a carbon content of 2.4–4 wt.% and chromium content of 15–30 wt.% according to ASTM A532^[39] is an interesting material for machine learning applications, as the microstructure is multi-scale and multi-phase, showing larger eutectic carbides (EC) as well as, for heat-treated specimens (as used in our dataset), smaller-scale secondary carbides (SC), which are visualized using different etchants and contrasting methods, resulting in a

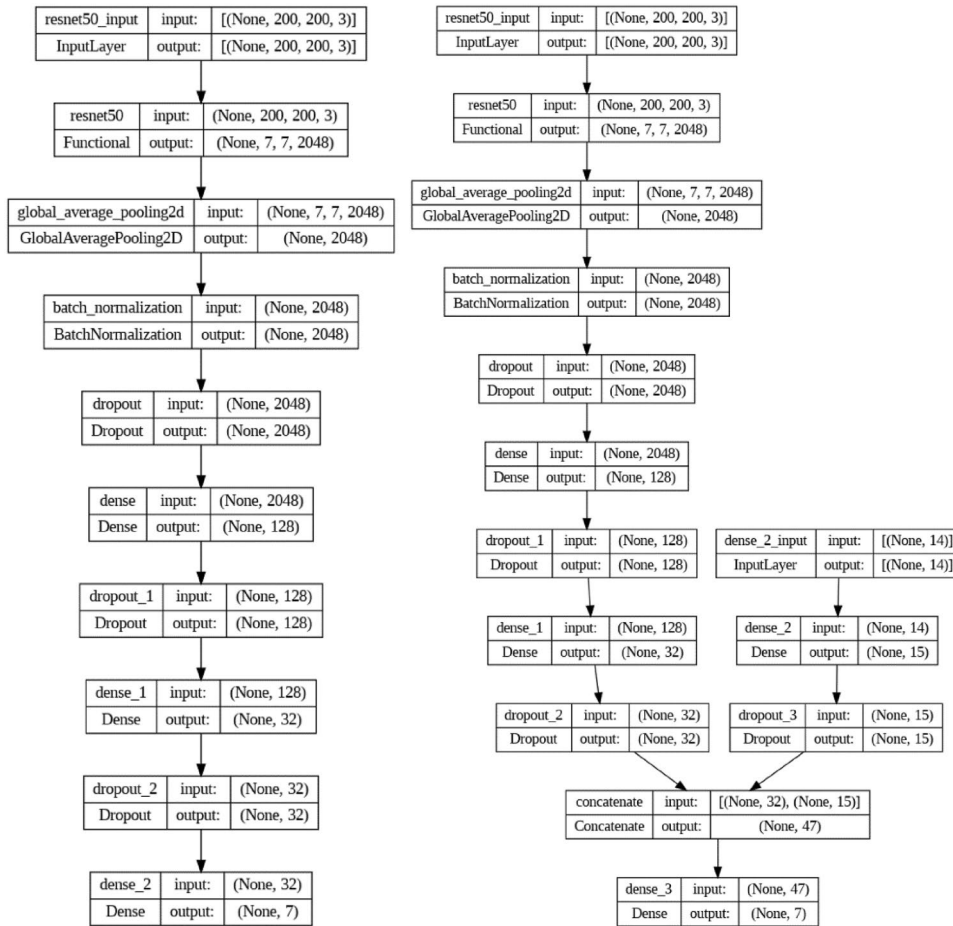


Figure 2. Classification models for pixel data only (left) and pixel and metadata (right) for the use case of the UHCSDB (14 input features for the metadata branch).

wide variety of contrasts.^[40] From the micrographs, using hand-assigned masks, we extracted smaller tiles of 200×200 pixels that could be clearly assigned to one of the two classes (EC or SC). The full metadata HCCI dataset comprises 460 full images, of which a total of 5462 tiles were extracted for classification.

Tests and methods

An assessment of the CNN classification model was performed with two different datasets:

- The UHCSDB as proposed by DeCost et al.^[31] a major openly available micrograph database with an accompanying metadata set in materials science. Since the classes in the dataset are highly unbalanced, it is particularly interesting for ML trainings, especially when evaluating a model’s performance. The micrographs were classified into 7 folders by their primary microconstituent class, as shown in Fig. 3. The provided metadata is especially interesting because the manufacturing param-

eters are richly described. Efforts were made to complete the metadata table and provide complete metadata for the entire dataset. Additionally, the metadata table was rebuilt by converting the scale into microns for all images, converting the heat treatment duration to hours for all images, and changing the order of the micrographs to correspond to the order of the micrographs in the folders. Both filename and the primary microconstituent were omitted.

- The above-mentioned self-curated dataset on HCCI specimens for binary classification with an extensive set of 14 metadata features. The above-mentioned high variance in contrast and brightness within the classes makes this dataset particularly interesting, as seen in Fig. 4. Additionally, in contrast with the UHCSDB where the metadata is richest for the manufacturing of the specimens, the metadata of the HCCI dataset is focused greatly on the imaging (*i.e.*, SEM) information. Before assessing the CNN classification model itself, for this HCCI dataset, a preliminary classification of EC and SC tiles was performed using Haralick texture parameters (spatial relationship/depend-

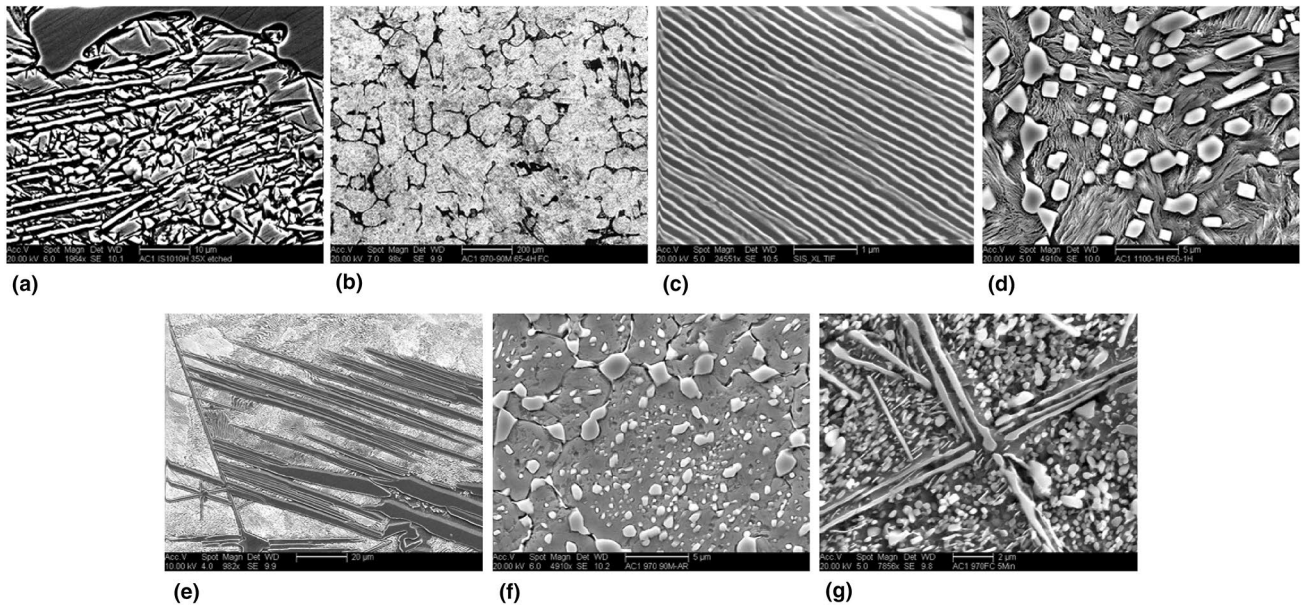


Figure 3. Example SEM micrographs from the Ultra-High-Carbon Steel Micrograph Database (UHCSDB). The seven classes were chosen according to the primary microstructure constituent of the images: (a) martensite and/or bainite, (b) proeutectoid cementite network microstructure, (c) pearlite, (d) pearlite containing spheroidized cementite, (e) pearlite containing Widmanstätten cementite, (f) spheroidized cementite, and (g) spheroidite and Widmanstätten cementite.^[31]

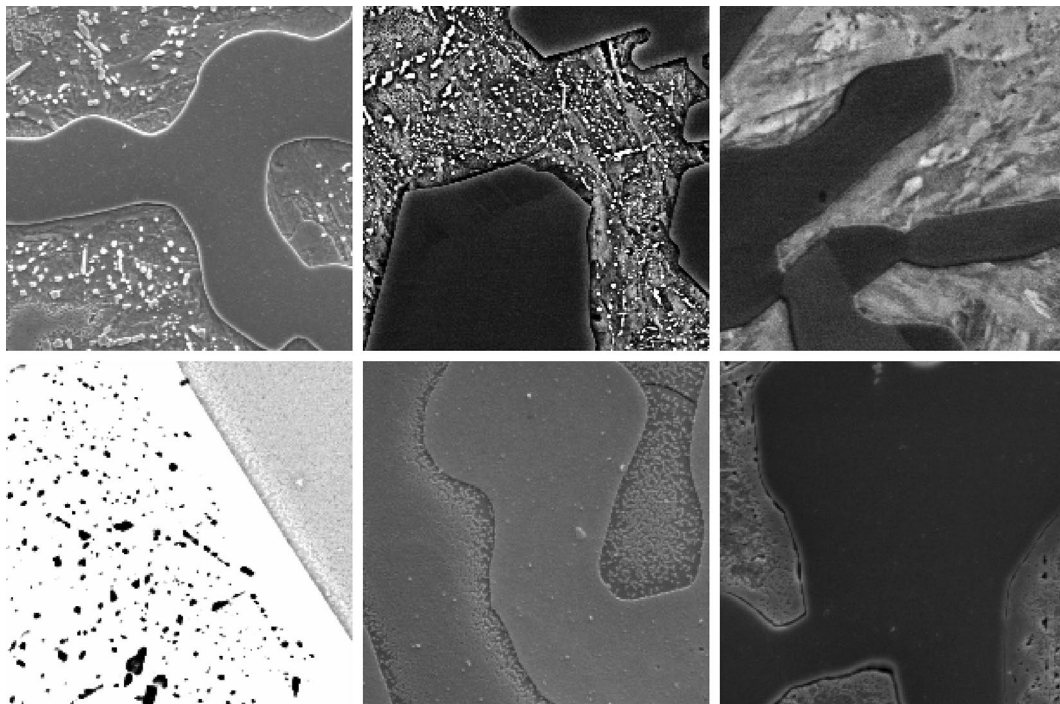


Figure 4. Example SEM micrographs from the High-Chromium Cast Iron (HCCI) database. All micrographs show the eutectic carbide class, showcasing the high variance of contrast and brightness of the images within the dataset.

ence of gray values in the image^[20] and metadata using the MathWorks MATLAB Classification Learner app with

its assistant feature ranking and selection tool and hyperparameter optimization.

The primary goal of the assessment was to confirm the proper functioning of the metadata-assisted CNN classification model by comparing it to the performance of the image-only model. The secondary aim was to assess the impact on the performance of the model by varying both the dataset and the model (*i.e.*, its backbone for feature extraction), as well as using an adversarial example. Therefore, firstly, the CNN model with a ResNet50 backbone was assessed with respect to its accuracy and loss for training and validation (*i.e.*, previously unseen) data, for each dataset. Secondly, the model performance was tested for different backbones, all with the UHCSDB dataset. Thirdly, two adversarial examples were used to assess the effect of wrongly used metadata, by first purposefully assigning metadata to the wrong images, rendering the metadata (and basically the entire dataset) unusable, followed by training with correct metadata and validation with images and purposefully wrongly assigned metadata, by randomly shuffling the metadata of the validation step upon loading.

The metrics used for model assessment were accuracy (number of correct predictions over total number of predictions) and categorical cross-entropy loss (cumulated error) for both training and previously unseen validation data, with the aim of increasing accuracy and reducing loss. Generally, validation loss is the most significant metric, as it reflects the model's performance on unseen data, and therefore qualifies its ability to generalize.

Results and discussion

Preliminary tests on hybrid image and metadata classification using conventional ML approaches, conducted with the HCCI dataset using the MATLAB Classification Learner application, have shown an increased accuracy when using a combination of features from image and metadata, as compared to image information only. Haralick textural features as well as local binary pattern were used as image features, resulting in 28 total image texture parameters.

A run with texture parameters only yielded an accuracy of 95.4%, after a reduction to 15 features (by MATLAB Classification Learner App automated feature selection) and optimizing the hyperparameters. The classification of the metadata only yielded an accuracy of 80.2%, with some features having no importance for the algorithm at all (*e.g.*, image size as that was identical for all images and irrelevant to the tile size), so they could easily be omitted. Not all feature ranking algorithms provided identical results but removing the universally low-ranked features allowed to reduce data from 15 to 7 features, resulting in an accuracy of 84.5% for metadata only. For the hybrid classification, a final accuracy of 97.5% was reached, using 15 texture and 8 metadata parameters after hyperparameter optimization.

This aligns with the results of hybrid ML models from biomedical applications, which reported a benefit from using patient metadata.^[34,41] These promising preliminary tests were the motivation to also test hybrid CNN classifications, since the performance of DL models often outperforms that of conventional ML in image processing.

For deep learning classification for both HCCI and UHCSDB datasets using our own model, image-only classification yields excellent results with a very high validation accuracy of up to 98% and little error, surpassing the results from the conventional machine learning classification. The performance can be increased a little more using data augmentation (rescaling, zooming, and flipping of the images) before the training, which is especially relevant for smaller datasets. Using metadata results in slightly increased validation accuracy as well as a decrease of the error (validation loss). After identifying ResNet50 as the optimal CNN backbone, both the use of pixel data with data augmentation yielded accuracies of up to 100% for the HCCI dataset, compared to a maximum of 98% for pixel data only. For the UHCSDB, accuracies reached up to 99% for pixel data and 100% for hybrid data. However, with backbones as high performing as ResNet50, it is questionable whether the use of metadata is significant enough to be worth the additional steps of data treatment, especially when data augmentation can be done in fast and easy to implement step and yields similarly good results. A comparison of various backbones shows that ResNet50 overall performs best, with validation accuracies up to 98% for both image and combined features. In fact, the effect of the metadata addition is minimal and falls within the error range of the metrics. For in this showcase less high-performing backbones, such as VGG16, Inception, or Xception, the improvement of the accuracy when using metadata is generally higher (up to 7%), with an accompanying decrease in the loss, but the performance does not reach as high as ResNet50. Hence, the hyperparameters chosen for the deep learning model, especially the backbone, do influence its performance.

Although a slight increase in performance due to the use metadata was detectable, the substantial rise in accuracy as seen in biomedical CNN models^[34] failed to appear. This is most probably due to the fact that the micrographs, recorded with a high-performing SEM, are already data rich, *i.e.*, they are particularly high in details,^[23] compared to medical images, such as X-rays. Patient metadata therefore probably provides more important features in comparison to the features that can be extracted from medical images. Additionally, ResNet50 is very high performing in feature recognition, which results in an overall excellent performance for micrographs alone and an additional use of metadata cannot have any major effect. In fact, for other backbones, the overall performance was lower but in comparison, the effect due to the metadata use increased. In short, in this use case, metadata can be helpful if the most appropriate ML approach or the optimal hyperparameters were not chosen beforehand. However, with the proper choice of ML approach and hyperparameters, baseline models of image only classification can outperform other metadata-based classification. Thus, the effect of hyperparameter-tuning can outweigh that of using metadata when it comes to improvement of the ML model. Nonetheless, for applications where users cannot or do not want to invest as much time and effort into building and optimizing a ML model, using metadata can be a useful approach. In fact, while model optimization has a vast impact on its performance, it can become complex and may require some guesswork. Therefore, a

ML model simply including metadata could be beneficial, robust, and easier to build from scratch in this use case. Once a pipeline for metadata pre-processing and inclusion into the ML model has been established, an incorporation of metadata by default can definitely prove to be useful.

The general impact of metadata use was tested using a first adversary example, where a change in the train-test split seed resulted in metadata getting wrongly assigned to the image data, rendering it unusable. The resulting performance of the hybrid model was similar to that of the image model, showing that without useful metadata, a hybrid model serves no purpose. At the same time, the equally high performance of both models despite faulty metadata shows how performant the image data model is and that the effect of the metadata in our showcase is minimal. The limited effect of the metadata features may also be attributed to the smaller number of metadata features, as compared to convolutional features. A second adversarial example, consisting in training the model with a correct dataset but performing a validation with external data where incorrect metadata was assigned to the images, showed no difference in the model's performance in comparison to not using metadata at all, which aligns with the previous results that wrongly assigned metadata serve no purpose, which thus applies for both training and validation. Rather, there might be a potential bias of the model generated by the metadata (e.g., production parameters or magnification) which could contradict an objective microstructure classification—however, this would have to be subject to further evaluation.

The general good functioning of the hybrid model could be beneficial for the incorporation of further data in addition to image data, such as material properties, especially when aiming for a deeper understanding of processing–microstructure–properties relationships for materials design and optimization. Other simulation or data generation methods for material property prediction were proposed by Herriott et al.^[42], who suggested data-driven modeling for mechanical property prediction of a simulated microstructural dataset using a macroscale finite-volume model for thermal history prediction during direct laser deposition, paired with a 3D cellular automata (CA) and a solid mechanics model or by Acar et al.,^[43] presenting a ML approach to study the linkage between deformation processing and microstructural texture evolution supported by a single-crystal plasticity model. Other examples comprise finite elements (FE) simulations for mechanical response prediction from the microstructure,^[44] the Materials Genome Integration System Phase and Property Analysis (MIPHA),^[45] or the use of electron backscatter diffraction (EBSD) data.^[46–48] Zhang and Shao propose other data acquisition techniques for image-based materials property prediction, including numerical (bandwidth, structural geometry), textual (composition, structure, properties), and image data (various types of microscopy and spectroscopy), emphasizing, for example, Fourier transformed infrared (FTIR) spectra or molecular images.^[17]

Conclusion and outlook

We have shown how to extract and curate metadata in order to use them in ML applications, in the context of the FAIR guidelines. We have proposed a machine learning algorithm based on a CNN which can combine image (*i.e.*, pixel)-based classification with metadata and compare the performances. While the CNN yields excellent results for both datasets (HCCI dataset with high variance and UHCSDB with unbalanced classes), the significance of metadata was minor because of the high initial performance. For the assessed classification task, it is questionable whether the implementation of metadata is worth its cost, as the metadata curation can be time-consuming. Considering the targeted systematic collection of metadata from the FAIR principles, the effort of the metadata curation would decrease, making their use more convenient, and experimental verifications could show use cases for which the use of metadata could be more significant. Furthermore, we encourage to use the concatenating architecture to import other features, not necessarily metadata, with the aim to improve CNN model performance. In conclusion, we have a proposed a novel step in the emerging topic that is metadata usage in materials science by combining it with a deep learning approach and look forward to exploring its potential.

Author contributions

Conceptualization: M.S., M.M.; Methodology: M.S., M.M., B.I.B., M.A.G., U.P.N.; Writing—original draft preparation: M.S., M.M.; Writing—review and editing of the manuscript: M.S., M.M., B.I.B., M.A.G., U.P.N.; Funding acquisition: M.A.G., U.P.N.; Supervision: F.M.

Funding

Open Access funding enabled and organized by Projekt DEAL. M.A.G. and U.P.N. received funding support from the Deutsche Forschungsgemeinschaft (DFG, Project: GU 2102/2-1).

Data availability

Both the ML code and the HCCI dataset are available from the corresponding author on reasonable request.

Declarations

Conflict of interest

The authors declare no conflict of interest.

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and

indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- C. Draxl, M. Scheffler, Big data-driven materials science and its FAIR data infrastructure, in *Handbook of Materials Modeling*, ed. by W. Andreoni, S. Yip (Springer International Publishing, Cham, 2020)
- L. Himanen, A. Geurts, A.S. Foster, P. Rinke, Data-driven materials science: status, challenges, and perspectives. *Adv. Sci.* **6**(21), 1900808 (2019)
- M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak et al., The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**(1), 160018 (2016)
- M. Scheffler, M. Aeschlimann, M. Albrecht, T. Bereau, H.J. Bungartz, C. Felser et al., FAIR data enabling new horizons for materials research. *Nature* **604**(7907), 635–642 (2022)
- F.X. Coudert, Materials databases: the need for open, interoperable databases with standardized data and rich metadata. *Adv. Theory Simul.* **2**(11), 1900131 (2019)
- M. Huisman, M. Hammer, A. Rigano, U. Boehm, J.J. Chambers, N. Gaudreault, et al., A perspective on microscopy metadata: data provenance and quality control (2021). <http://arxiv.org/abs/1910.11370>
- A. Medina-Smith, C.A. Becker, R.L. Plante, L.M. Bartolo, A. Dima, J.A. Warren et al., A controlled vocabulary and metadata schema for materials science data discovery. *Data Sci. J.* **29**(20), 18–18 (2021)
- R. Cohn, E. Holm, Unsupervised machine learning via transfer learning and k-means clustering to classify materials image data. *Integr. Mater. Manuf. Innov.* **10**(2), 231–244 (2021)
- V.H.C. de Albuquerque, P.C. Cortez, A.R. de Alexandria, J.M.R.S. Tavares, A new solution for automatic microstructures analysis from images based on a backpropagation artificial neural network. *Nondestruct. Test. Eval.* **23**(4), 273–283 (2008)
- B.L. DeCost, B. Lei, T. Francis, E.A. Holm, High throughput quantitative metallography for complex microstructures using deep learning: a case study in ultrahigh carbon steel. *Microsc. Microanal.* **25**(1), 21–29 (2019)
- B. Bachmann et al., Efficient reconstruction of prior austenite grains in steel from etched light optical micrographs using deep learning and annotations from correlative microscopy. *Front. Mater.* **9**, 1033505 (2022)
- S.M. Azimi, D. Britz, M. Engstler, M. Fritz, F. Mücklich, Advanced steel microstructural classification by deep learning methods. *Sci. Rep.* **8**(1), 2128 (2018)
- D. Britz, J. Webel, J. Gola, Y. Steimer, F. Muecklich, Practical metallography of low carbon steels—new approaches in preparation, imaging and analysis of microstructures. *Microsc. Microanal.* **24**(S1), 2226–2227 (2018)
- W.D. Callister Jr., D.G. Rethwisch, *Fundamentals of Materials Science and Engineering: An Integrated Approach* (Wiley, New York, 2020)
- K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn et al., Recent advances and applications of deep learning methods in materials science. *NPJ Comput. Mater.* **8**(1), 1–26 (2022)
- J. Na, S.J. Kim, H. Kim, S.H. Kang, S. Lee, A unified microstructure segmentation approach via human-in-the-loop machine learning. *Acta Mater.* **255**, 119086 (2023)
- L. Zhang, S. Shao, Image-based machine learning for materials science. *J. Appl. Phys.* **132**(10), 100701 (2022)
- M. Laub, B.I. Bachmann, E. Detemple, F. Scherff, T. Staudt, M. Müller et al., Determination of grain size distribution of prior austenite grains through a combination of a modified contrasting method and machine learning. *Pract. Metallogr.* **60**(1), 4–36 (2023)
- B.I. Bachmann, M. Müller, D. Britz, T. Staudt, F. Mücklich, Reproducible quantification of the microstructure of complex quenched and quenched and tempered steels using modern methods of machine learning. *Metals* **13**(8), 1395 (2023)
- M. Müller, D. Britz, L. Ulrich, T. Staudt, F. Mücklich, Classification of bainitic structures using textural parameters and machine learning techniques. *Metals* **10**(5), 630 (2020)
- V.H.C. de Albuquerque, C.C. Silva, T.I. Menezes, J.P. Farias, J.M.R.S. Tavares, Automatic evaluation of nickel alloy secondary phases from SEM images. *Microsc. Res. Tech.* **74**(1), 36–46 (2011)
- R. Kondo, S. Yamakawa, Y. Masuoka, S. Tajima, R. Asahi, Microstructure recognition using convolutional neural networks for prediction of ionic conductivity in ceramics. *Acta Mater.* **1**(141), 29–38 (2017)
- E.A. Holm, R. Cohn, N. Gao, A.R. Kitahara, T.P. Matson, B. Lei et al., Overview: computer vision and machine learning for microstructural characterization and analysis. *Metall. Mater. Trans. A* **51**(12), 5985–5999 (2020)
- K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition (2014). <https://arxiv.org/abs/1409.1556>
- C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., Going deeper with convolutions. <http://arxiv.org/abs/1409.4842>
- C.F. Xception, Deep learning with depthwise separable convolutions (2017). <http://arxiv.org/abs/1610.02357>
- G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely connected convolutional networks (2023). <http://arxiv.org/abs/1608.06993>
- K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition (2015). <http://arxiv.org/abs/1512.03385>
- A. Lupulescu, T. Flowers, L. Vermillion, S. Henry, ASM micrograph database™. *Metallogr. Microstruct.* **4**(4), 322–327 (2015)
- K. Song, Y. Yan, A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Appl. Surf. Sci.* **15**(285), 858–864 (2013)
- B.L. DeCost, M.D. Hecht, T. Francis, B.A. Webler, Y.N. Picard, E.A. Holm, UHCSDB: ultrahigh carbon steel micrograph database. *Integr. Mater. Manuf. Innov.* **6**(2), 197–205 (2017)
- I. Kemmer, A. Keppler, B. Serrano-Solano, A. Rybina, B. Özdemir, J. Bischof, Building a FAIR image data ecosystem for microscopy communities. *Histochem. Cell Biol.* (2023). <https://doi.org/10.1007/s00418-023-02203-7>
- L.M. Ghiringhelli, C. Carbogno, S. Levchenko, F. Mohamed, G. Huhs, M. Lüders et al., Towards efficient data exchange and sharing for big-data driven materials science: metadata and data formats. *NPJ Comput. Mater.* **3**(1), 1–9 (2017)
- F. Nunnari, C. Bhuvaneshwara, A.O. Ezema, D. Sonntag, A study on the fusion of pixels and patient metadata in CNN-based classification of skin lesion images, in *Machine Learning and Knowledge Extraction (Lecture Notes in Computer Science)*, ed. by A. Holzinger, P. Kieseberg, A.M. Tjoa, E. Weippl (Springer International Publishing, Cham, 2020), pp. 191–208
- C. Gohlke, cgohlke/tiffle: v2023.8.30. Zenodo; 2023. <https://zenodo.org/record/6795860>. Accessed 13 Sep 2023
- R. Aversa, R. Joseph, E. Vitali, A. Kirar, Extracting, mapping, editing SEM metadata. <https://publikationen.bibliothek.kit.edu/1000158866>. Accessed 2 Aug 2023
- M.A. Guitar, U.P. Nayak, D. Britz, F. Mücklich, The effect of thermal processing and chemical composition on secondary carbide precipitation and hardening in high-chromium cast irons. *Int. Metalcast.* **14**(3), 755–765 (2020)
- U.P. Nayak, F. Mücklich, G.M.A. Time-Dependant, M. Evolution, T. Behaviour, Time-dependant microstructural evolution and tribological behaviour of a 26 wt% Cr white cast iron subjected to a destabilization heat treatment. *Met. Mater. Int.* (2022). <https://doi.org/10.1007/s12540-022-01276-8>
- A04 Committee. Specification for abrasion-resistant cast irons. ASTM International, West Conshohocken. <http://www.astm.org/cgi-bin/resolver.cgi?A532A532M-93AR03>. Accessed 14 Sep 2023
- U.P. Nayak, M.A. Guitar, F. Mücklich, Evaluation of etching process parameter optimization in the objective specific microstructural characterization of as-cast and heat treated HCCI alloy. *Pract. Metallogr.* **57**(10), 688–713 (2020)
- Y.N.T. Vu, R. Wang, N. Balachandar, C. Liu, A.Y. Ng, P. Rajpurkar, MedAug: contrastive learning leveraging patient metadata improves

- representations for chest X-ray interpretation, in *Proceedings of the 6th Machine Learning for Healthcare Conference (PMLR)*, pp. 755–69. <https://proceedings.mlr.press/v149/vu21a.html>. Accessed 21 Feb 2023
42. C. Herriott, A.D. Spear, Predicting microstructure-dependent mechanical properties in additively manufactured metals with machine- and deep-learning methods. *Comput. Mater. Sci.* **1**(175), 109599 (2020)
 43. P. Acar, Machine learning approach for identification of microstructure-process linkages. *AIAA J.* **57**(8), 3608–3614 (2019)
 44. S.G. Ammasai, Prediction of two-phase composite microstructure properties through deep learning of reduced dimensional structure-response data. *Compos. B* **225**, 109282 (2021)
 45. Z.L. Wang, Y. Adachi, Property prediction and properties-to-microstructure inverse analysis of steels by a machine-learning approach. *Mater. Sci. Eng. A* **28**(744), 661–670 (2019)
 46. Y. Han, R.J. Griffiths, H.Z. Yu, Y. Zhu, Quantitative microstructure analysis for solid-state metal additive manufacturing via deep learning. *J. Mater. Res.* **35**(15), 1936–1948 (2020)
 47. I.D. Jung, D.S. Shin, D. Kim, J. Lee, M.S. Lee, H.J. Son et al., Artificial intelligence for the prediction of tensile properties by using microstructural parameters in high strength steels. *Materialia* **1**(11), 100699 (2020)
 48. B.I. Bachmann, M. Müller, D. Britz, A.R. Durmaz, M. Ackermann, O. Shchyglo et al., Efficient reconstruction of prior austenite grains in steel from etched light optical micrographs using deep learning and annotations from correlative microscopy. *Front. Mater.* (2022). <https://doi.org/10.3389/fmats.2022.1033505>

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