X-RAY DIFFRACTION

AI-Assisted Phase Identification from Powder XRD Data

Application Note 633

Abstract

This laboratory report summarizes a joint Bruker AXS and KIT investigation into AI-assisted phase identification for powder X-ray diffraction (XRD) using iron ore as a proof-of-concept. We generated **synthetic teaching scans** from crystal-structure information and trained convolutional neural networks (CNNs) on mixtures drawn from a constrained phase list.

The models were stress-tested on harder mixtures (more phases and smaller fractions) and benchmarked on measured scans from a Bruker QC application package. Key findings: realistic **peak-position variability** is critical for robustness, the minimum taught **weight-%** limits trace-phase capability, and performance gains can saturate beyond a certain teaching-set size relative to training time.

Transfer to measured scans (within the same instrument/protocol assumptions) was demonstrated for the evaluated scope. Overall, the results support AI as a complementary method for defined QC scenarios rather than a replacement for classic search/match workflows.

AI can learn diffraction patterns—but successful phase identification depends less on AI architecture than on teaching the model the realities of XRD.

This study demonstrates that robust AI-assisted phase identification is possible when synthetic teaching data accurately reflects peak shifts, peak broadening, background effects, and measurement variability. The results highlight both the potential and the boundaries of AI: high-throughput screening and standardized decision support within a defined phase scope, while expert knowledge and established search/match methods remain essential for unexpected phases and challenging edge cases.

Objective and scope

Why use AI?

Phase identification typically starts with database search/match, but final calls often still require expert judgment—especially when peaks overlap or shift, or when phases are minor. For complex samples this is time-consuming and results can vary between users. We test whether AI can support a realistic QC workflow by standardizing phase-present/absent decisions within a defined scope, helping to set up the correct TOPAS quantification model while keeping an expert in the loop.

What is the scope of this study?

Because a large labeled set of measured scans was not available, we generated realistic synthetic teaching data from crystal-structure information. Using a fixed candidate set, we trained and tested a multi-label classifier on simulated mixtures and then evaluated transfer to measured scans collected under comparable instrument and protocol assumptions. The aim is to clarify feasibility, effort, and limitations for a QC-style task—not to replace general-purpose search/match methods.

Why iron ore?

Iron ore was selected because a commercially available quality-control (QC) application package for quantitative DIFFRAC.TOPAS refinement defines a practical and bounded test case. It provides well-defined instrument settings and measurement parameters and a limited phase list that is typical for routine work. The package covers several iron-ore types (low- and high-grade ores with oxidic, sulfidic, and carbonatic chemistry) plus typical gangue minerals (see Table 1). It includes 28 phases, which keeps the proof-of-concept manageable while still spanning a wide range of crystal symmetries, from high-symmetry α-Fe with few peaks to low-symmetry gangue minerals with many peaks. Iron ore exhibits common experimental complications such as disorder, differences in crystallinity that broaden peaks, solid-solution effects that shift peaks, and particle-shape effects (for example, plated minerals) that can cause preferred orientation.

Table 1

List of ore and gangue minerals (phase scope) covered in this study.
For readability, Muscovite-1M and Muscovite-2M1 are reported as a single "Muscovite" item.

Minerals			
Ore/Fe-species		**Gangue**	
α-Iron	Fe	Albite	$NaAlSi_3O_8$
Goethite	FeO(OH)	Anatase	TiO_2
Hematite	Fe_3O_4	Anorthite	$CaAl_2Si_2O_8$
Ilmenite	$FeTiO_3$	Antigorite	$(Mg, Fe)_3Si_2O_5(OH)_4$
Magnetite	Fe_2O_3	Calcite	$CaCO_3$
Pyrite	FeS_2	Chlorite IIb	$Na_{0.5}Al_4Mg_2Si_7AlO_{18(}OH)_{12}.5(H_2O)$
Siderite	$FeCO_3$	Dolomite	$CaMg(Co_3)_2$
Wueslite	FeO	Gibbsite	$Al(OH)_3$
		Hornblende	$Ca_2Mg_4(Al, Fe)(Si_7AlO_{22})(OH)_2$
		Hydroxylapatite	$Ca_5(PO_4)_3(OH)$
		Kaolinite	$Al_2Si_2O_5(OH)_4$
		Magnesite	$MgCO_3$
		Microcline	$KAlSi_3O_8$
		Muscovite 1M	$KAl_2(Si_3Al)O_{10}(OH)_2$
		Muscovite 2M1	$KAl_2(Si_3Al)O_{10}(OH)_2$
		Orthoclase	$KAlSi_3O_8$
		Pyrophylite	$Al_2Si_4O_{10}(OH)_2$
		Quartz	SiO_2
		Rutile	TiO_2
		Talc	$Mg_3Si4O_{10}(OH)_2$

Teaching concept and methodology

This work follows a QC-style teaching concept: the model is trained on many labeled diffraction patterns that reflect routine measurement variation (peak shifts, intensity changes, broadening, background, and noise). Because large, fully labeled measured datasets are rarely available, we generated synthetic scans from crystal-structure information. To keep the task repeatable, simulations assume a fixed instrument model and scan grid. For this study we used the Bruker D2 PHASER for a fundamental-parameters line-profile model and a 5–70° 2θ range at 0.01° steps (6500 data points per scan).

Step 1: Create base scans. For each phase, we used TOPAS to simulate single-phase patterns from the crystal-structure input. Each simulated single-phase pattern is a base scan, meaning one plausible realization of that phase within expected experimental variability. In these base scans we included peak-position shifts by varying unit-cell parameters, intensity changes by preferred orientation (March–Dollase type, typically 0.5–1.5), and peak-shape changes by varying crystallite size (typically 50–500 nm).

Step 2: Generate synthetic mixtures. Each mixture is created by randomly selecting up to five base scans from the 28 candidate phases, assigning random weights that sum to 100%, and linearly combining the scaled base scans into one mixed pattern.

Step 3: Add measurement-like effects. After mixing, the pattern is augmented with contributions that commonly appear in measured data, including air scattering at low angles, a polynomial background (Chebyshev, typically 3rd–5th order), specimen height offsets, and noise (Gaussian approximation of counting statistics). This post-processing reduces the gap between synthetic and measured scans from the same instrument class.

For the 28 phases, we generated about 50 base scans per phase and built training sets of roughly 100,000 synthetic mixtures, with additional runs up to 500,000 mixtures. Training time increased with data size, from hours for ~100k mixtures to tens of hours for ~500k mixtures on a GPU.

Model and evaluation

Model. Convolutional neural networks (CNNs) were trained as pattern recognizers for 1D diffractograms. The model outputs a present/absent decision for each candidate phase (multi-label classification).

Training and testing. Models were trained on synthetic mixtures where the phase labels are known by construction. Generalization was tested on separate, intentionally harder synthetic sets (smaller fractions and more phases than used during teaching).

Test criteria and outcome. We assessed whether the model finds truly present phases, avoids false claims, and remains reliable under less ideal patterns (shifts, broadening, stronger background/noise) and smaller phase fractions. The synthetic tests showed that peak-position variability is a key driver for robustness, that the minimum taught weight-% limits trace detection, and that gains level off once the training set is large enough (here, beyond ~100,000 mixtures) relative to added training time.

Validation performance (F1 score). Several CNN design variants were evaluated. Without discussing architectural details here, validation F1 scores exceeded 95%, with the best-performing setup reaching about 99% on the synthetic validation sets. The F1 score summarizes the trade-off between missed phases and spurious calls.

Once trained, the model can generate phase-present/absent predictions quickly, which may support objective screening of large QC batches (for example hundreds of scans) within seconds.

Application to measured scans

Beyond synthetic mixtures, we tested whether the trained models **transfer to measured powder XRD scans**. This comparison probes practical usefulness under real measurement conditions (background, noise, preferred orientation, peak shifts/broadening) and benchmarks the CNN against two references: (i) the package **expert-validated** phase model and (ii) a **fully automated** DIFFRAC.EVA search/match workflow.

Measured scans were taken from the **iron-ore application package**, which provides standardized instrument settings and representative ore types. The scans are grouped by material origin/type; the table below lists the shorthand labels used in this report.

Brazil	High-grade ore from Brazil
Scale	Mill scale (technical product) from Thailand
W Aus.	Mainly gangue from Western Australia ore
Carb	Magnetite ore with carbonatic gangue
Ven	High-grade ore from Venezuela
PY1, Py2	Pyritic material with different gangue

Groundtruth via human expert evaluation

The **ground truth** used for benchmarking is taken from the **Bruker iron-ore QC application package** itself. For each benchmark scan in that package, Bruker provides (and internally validated) a phase model and quantitative phase amounts obtained by a standard workflow: an initial phase identification in **DIFFRAC.EVA** (search/match against **PDF-5**, with automated background handling and an **element filter** based on the accompanying XRF information), followed by expert review of the proposed matches. The finalized phase set is then refined in **DIFFRAC.TOPAS** by Rietveld quantitative analysis. The resulting refined phase fractions are treated as the reference labels for this study and are reported in the result tables as follows: in **Tables 2 and 3, integer numbers** denote the **rounded TOPAS wt-%** for phases included in the final refinement model; entries marked **T** indicate **trace level** contributions below 1 wt-%; and an empty cell indicates that the phase is absent in the reference model.

Table 2
Automated **DIFFRAC.EVA** phase-identification results for the iron-ore benchmark scans (**1st-pass** automatic phase selection). **Integer values** in the cells are the package reference amounts (rounded **TOPAS** Rietveld wt-%); **T** denotes trace contributions <1 wt-%; and an empty cell indicates absence in the reference model. Candidates marked **(*)** were proposed by EVA but not selected automatically. The entry marked **(x)** refined to 0 wt-% in the subsequent PFRS step and was therefore removed before the 2nd pass. (Color code and TP/FP/FN abbreviations are defined in the Discussion).

	Brazil	Scale	W Aus	Carb	Ven.	PY1	PY2
Ore (Fe-species)							
α-Iron							
Goethite	3		3		11		
Hematite	89	6	5		56		
Ilmenite				T			
Magnetite	3	33		89	17		
Pyrite						3(*)	2
Siderite				T			
Wueslite		60					
Gangue							
Albite							
Anatase			T				
Anorthite							
Antigorite				T			
Calcite				4	13		
Chlorite IIb							
Dolomite				2(*)			
Gibbsite			T				
Hornblende							
Hydroxylapatite							
Kaolinite			5				9
Magnesite				2			
Microcline							5
Muscovite			32	1(*)			
Orthoclase							
Pyrophylite							15
Quartz	5	T	54		3	97	69
Rutile							(x)
Talc							

Automated conventional phase analysis using DIFFRAC.EVA

To benchmark the AI-based workflow against a classic approach, we also ran a **fully automated conventional** phase-identification workflow using **DIFFRAC.EVA**. EVA's search/match engine combines peak finding with background/noise handling and applies established heuristics to rank and select candidate phases. For a fair comparison, we created an EVA **User Database** containing **the same crystal-structure entries** that were used to generate the synthetic teaching data (i.e., the same 28-phase package scope). In the reporting and evaluation, the two muscovite polytypes (1M and 2M1) are treated as one combined "Muscovite" item for readability. We deliberately **did not apply additional chemistry filters**. We then used EVA's **automatic phase selection** and report the outcome after the **1st pass** in Table 2.

For completeness, we also applied EVA's **2nd-pass** workflow based on **PFRS** (Profile Fit Residual Search). In this step, the phases

Table 3
Predictions of the best-performing **CNN** on the same iron-ore benchmark m**easured** scans. Cell entries follow the same reference conventions as Table 1: **integer values** report the package reference amounts (rounded **TOPAS** Rietveld wt-%), **T** denotes <1 wt-% trace contributions, and an empty cell indicates absence in the reference model. Trace-level phases (**T**) were excluded from the quantitative scoring in this study. (Color code and TP/FP/FN abbreviations are defined in the Discussion.)

	Brazil	Scale	W Aus	Carb	Ven.	PY1	PY2
Ore (Fe-species)							
α-Iron							
Goethite	3		3		11		
Hematite	89	6	5		56		
Ilmenite				T			
Magnetite	3	33		89	17		
Pyrite						3	2
Siderite				T			
Wueslite		60					
Gangue							
Albite							
Anatase			T				
Anorthite							
Antigorite				T			
Calcite				4	13		
Chlorite IIb							
Dolomite				2			
Gibbsite			T				
Hornblende							
Hydroxylapatite							
Kaolinite			5				9
Magnesite				2			
Microcline							5
Muscovite			32	1			
Orthoclase							
Pyrophylite							15
Quartz	5	T	54		3	97	69
Rutile							
Talc							

selected in the 1st pass are first used for an automated profile-fit refinement (background, lattice parameters, peak-shape parameters, and scale factors). The remaining **fit residual** is then searched again to identify additional phases that were missed initially, with the aim of completing the phase model.

Automatic AI based phase identification

For the AI baseline, we applied the trained **convolutional neural network (CNN)** to the same **measured** iron-ore scans from the QC package. Each scan is processed on the fixed 5–70° 2θ grid used during teaching, and the network outputs a probability for each phase in the package scope (28 phases; for readability, *Muscovite-1M* and *Muscovite-2M1* are bundled and reported as a single "Muscovite" item). These probabilities are converted into a phase-present/absent decision using the model's predefined decision threshold (kept identical for all scans). The resulting phase calls are summarized in **Table 3** and compared against the package reference values using the same table conventions (numbers, **T**, blank) described above.

Discussion: limitations and why classic methods remain essential

We compare **Auto EVA** and the **CNN** by evaluating their phase-present/absent calls against the package **ground truth** (TOPAS reference model; Tables 2 and 3). Each entry is classified as **TP** (present and reported), **FP** (reported but absent), or **FN** (present but missed). Tables 2 and 3 visualize these outcomes with **TP = light blue, FP = plum**, and **FN = orange**.

From these counts we compute standard classification metrics: **Recall** = TP/(TP+FN), **Precision** = TP/(TP+FP), and **F1 score** = 2·(Recall·Precision)/(Recall+Precision). The aggregated results across all benchmark scans are summarized in Table 4.

Table 4 Performance metric results for Auto EVA and CNN based phase ID (values in %).

	Auto EVA	CNN
Recall	63	90
Precision	100	76
F1 score	78	83

Overall, the **CNN** achieved higher **recall** than **Auto EVA**, i.e., it missed fewer ground-truth phases. This higher sensitivity comes with lower **precision**: the CNN more often flags additional phases not present in the reference model. In QC terms, the CNN behaves as a stronger screening tool, while Auto EVA is more conservative. Both methods captured the **major phases**; differences mainly affect minor components and edge cases where the measured pattern deviates from a single stored reference entry.

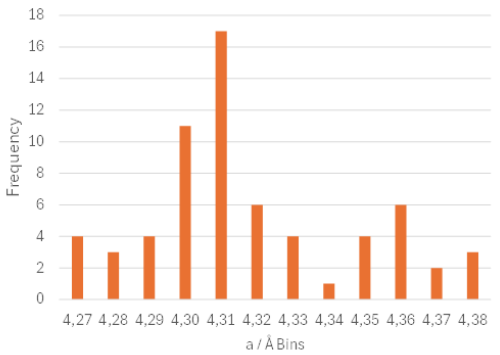

Figure 1
Spread of wüstite lattice parameters reported in the PDF 5 database.

A prominent edge case is **wüstite** ($Fe_{1-x}O$) in the Scale group: it is dominant in the ground truth but was **missed by Auto EVA** in the 1st-pass selection. As a non stoichiometric solid solution, wüstite spans a **wide lattice-parameter range** and its peaks can shift measurably. In our restricted EVA user database, the single wüstite entry used for matching had a = 4.307 Å, while the measured pattern corresponds to ~4.289 Å. With this offset, EVA's conservative selection did not accept wüstite as a close match despite its high intensity.

By contrast, the CNN flagged wüstite, suggesting the **teaching distribution** (unit-cell variation and peak-position augmentation) covered the observed offset. The case highlights a core limitation of restricted search/match: if variability is not represented by the stored entry—and heuristics are tuned to avoid false positives—dominant phases can be missed. Consequently, Auto EVA has been made more robust by adding **multiple candidates** spanning the known range, adapting tolerances/heuristics for variable phases, and using the **PFRS** second pass to recover missed phases. The higher recall of the AI algorithm for minor phases paired with lower precision might be due to the different treatment of intensities in the CNN compared to Auto EVA. This observation is subject to further investigation.

More broadly, both approaches offer clear **improvement potential**. For classic workflows, this includes better automated **heuristics** (selection thresholds and variability-aware tolerances) and search spaces that represent known phase variability rather than a single "best" entry. For AI, the key levers are more realistic teaching data (peak shifts and common artifacts), calibrated thresholds to reduce false positives, and targeted coverage of edge cases. A near-term **hybrid workflow** is therefore attractive: use the CNN as a fast, sensitive screener to propose a shortlist, then validate and quantify with EVA/TOPAS, with expert review for ambiguous cases.

Even with these results, **human experts and classic methods** remain essential. EVA-style workflows are general-purpose and transparent (peak-based evidence) and can broaden the search when unexpected phases occur. AI models are constrained by their teaching scope and can degrade when instruments, protocols, or materials differ from what was represented during teaching.

- Teaching-set coverage: a model cannot identify phases it has never been taught, or edge cases that are missing from the teaching distribution.

- Trace phases and detection limits: if training mixtures enforce a minimum weight-% threshold, phases below that level are inherently underrepresented; this directly limits trace-phase detection.

- Transferability: changes in instrument configuration, radiation, optics, detector settings, or sample preparation can shift the data distribution. Without retraining or explicit augmentation, performance may degrade.

- Maintenance effort: teaching sets, model versions, and acceptance criteria must be managed like any other method in the lab (documentation, change control, periodic revalidation).

- Explainability and trust: classic methods provide direct peak-based evidence; AI outputs need additional tooling (e.g., highlighted 2θ regions) to support confident decisions in ambiguous cases.

Conclusions and outlook

This work demonstrates that AI-assisted phase identification can be engineered and validated in a controlled way for a realistic XRD use case (iron ore) by combining: (i) constrained candidate sets, (ii) TOPAS-based synthetic data that includes realistic variability, and (iii) rigorous testing on harder-than-training mixtures and on measured scans. The outcome is best understood as an **additional tool** for defined QC problems, not as a replacement for general-purpose search/match workflows. From the synthetic evaluation, three practical design rules stand out: **realism of peak shifts** is critical for transfer, the **minimum phase fraction** used during teaching sets the effective detection limit, and **teaching-set size** shows diminishing returns once the core variability is covered.

Lessons learned
(practical rules for CNN-based phase identification)

For QC deployment: cover **peak-position variability** during teaching, set the **minimum phase fraction** to match the desired detection limit, expect some generalization to more complex mixtures if variability is represented, and plan for **diminishing returns** once key variability sources are covered.

Collaboration note

This investigation was performed as joint research by Bruker AXS R&D and the Karlsruhe Institute of Technology (KIT), Institute for Automation and Applied Informatics (IAI). We thank Jan Schützke for documenting and consolidating the work in the references listed below.

Schuetzke, J., Benedix, A., Mikut, R. & Reischl, M. (2021). Enhancing deep-learning training for phase identification in powder X-ray diffractograms. IUCrJ, 8(3), 408–420. https://doi.org/10.1107/S2052252521002402

Schützke, J. (2019). Evaluation of Machine Learning Approaches for Crystalline Phase Identification (Master's thesis, Mechanical Engineering). Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.

Bruker AXS is continually improving its products and reserves the right to change specifications without notice.
Order No. DOC-A88-EXS633 © 2026 Bruker AXS SE.

Bruker AXS
info.baxs@bruker.com

bruker.com

Worldwide offices
bruker.com/baxs-offices

Online information
bruker.com/xrd-software