Whitepaper

Evaluation of the Comparability Between Non-Destructive Eddy Current Testing with AI Prediction versus Destructive Microhardness Profile Testing for Determining Case Depth in Induction-Hardened Components

1. Introduction

In industrial quality assurance of heat-treated components, determining the case depth (e.g., in surface hardening) is crucial. The microhardness profile test (SHD) is the recognized reference method but is destructive, time-consuming, and very costly. As an innovative alternative, eddy current testing (ET) combined with artificial intelligence (AI) is proposed to non-destructively predict case depth.

This paper describes an evaluation methodology to verify whether the ET+AI method yields comparable results to the SHD measurement.

An additional advantage of the eddy current method is that, unlike destructive hardness testing, it inherently provides a normalized result across the entire eddy current volume of the test piece, making the test less sensitive to local microstructural deviations or preparation errors compared to the SHD method.

2. Overview of Methods

2.1 Case Depth Measurement (SHD)

- Destructive reference method according to DIN EN ISO 18203 - SHD
- Sample preparation by cutting, embedding, grinding, polishing
- Hardness measured at defined depth intervals
- Case depth defined by Hardness Limit value per Table 1, DIN EN ISO 18203

2.2 Eddy Current Testing with AI

- Non-destructive testing via electromagnetic interactions (multi-frequency testing, fundamental and harmonic frequencies, amplitudes, phase angles, etc. - as described in DIN EN ISO 15548-1)
- Frequency-dependent penetration depths into material
- Model training phase with representative dataset (ET signals + SHD values)
- Modeling using machine learning (e.g., neural networks)
- Prediction of case depth via hardness profile curve (based on ET signal characteristics)

45
SET HARDNESS LIMIT

1.5 3.0
SET MIN / MAX ECD

55 65
SET MIN / MAX SH

ECD SH GEN INDUCTED V2.4x
SELECT MODEL

1.0
CONFIG FACTOR

2.9 mm
RESULT ECD

58.4 HRC
RESULT SH

TEST PART: 3

Figure 1: Evaluation screen of the ibg eddyQlab AI software. The display shows the hardness profile determined destructive by the customer laboratory (red curve) compared with the eddy current-based AI prediction (blue curve) on an induction-hardened end face of a valve stem. The hardness values are shown in HRC (instead of HV), as the customer laboratory is using this hardness testing method in this particular case. The ibg eddyQlab AI software is capable of handling hardness values from both testing methods – HV and HRC – for both calibration input and the prediction output.

3. Evaluation Methodology

3.1 Comparison Criteria

The following methods are recommended for comparing two testing procedures:

- Scatter band analysis (according to [13] HORSCH A.)
- Bland-Altman plot for graphical assessment of systematic deviations
- Correlation analysis (e.g., Pearson coefficient of predicted vs. measured case depth)

To evaluate comparability, acceptable tolerances must be defined and aligned with quality management, production, and customer requirements, such as:

- All predictions lie within the scatter band of the SHD curves
- No systematic bias in the Bland-Altman diagram
- No values outside the tolerance range in the Bland-Altman diagram that cannot be attributed to measurement errors

A Measurement System Analysis (MSA) is not applicable for this comparison. The basic assumption of MSA – that the true value Xm of the reference standard does not vary – is not valid in this case.

It is also important to note that the referenced hardness test used to determine the microhardness profile is a destructive test. Unlike, for example, a length measurement, a destructive test cannot be repeated under identical conditions, as each microhardness indentation must, by design, be placed at a different position on the sample cross-section. Therefore, the measurement capability or suitability of the measuring device for evaluating case depth and hardness cannot be determined.

Furthermore, inhomogeneities in the material are always present, which lead to systematic deviations in SHD determinations. These are typically not taken into account (according to [13] HORSCH A.).

3.2 Initial Trials (Step 1)

For an initial trial to assess the extent to which the eddy current-based AI prediction delivers results comparable to the conventional laboratory testing of case depth, the following procedure is recommended.

It is essential to keep all influencing factors (material, hardening process, laboratory preparation, etc.) constant during the production of test parts. Only in this way can the outcome be influenced solely by the two testing methods being compared.

In industrial production practice, for reasons of time and cost, SHD testing used for process monitoring often involves only a single line of microhardness indentations to determine the hardness profile. While this may seem sufficient for case depth requirements with rather broad tolerance limits (e.g., 550 HV between 2 and 4 mm depth) for production monitoring, it is inadequate for a method validation.

To assess reproducibility and method consistency, it is recommended to determine one hardness profile (at least two parallel microhardness indentation tracks) at three radially offset positions (each 120° apart) on the cross-section of the test parts.

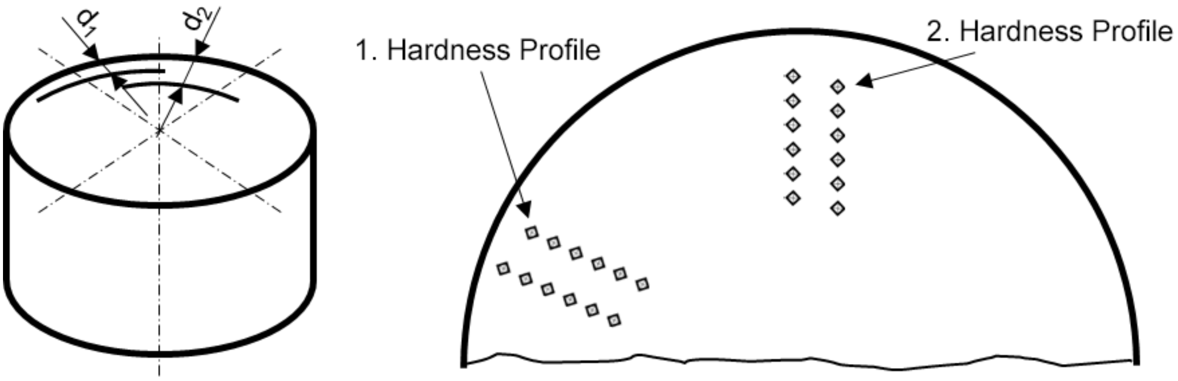

Figure 2: Arrangement of multiple hardness profiles (left, according to ISO 18203: Figure A.1). The positioning of the hardness test indentations should follow ISO 18203 Figure 3 (right). Source: [14] AWT, Guidelines for the Application of DIN EN ISO 18203 – Determination of the Thickness of Hardened Surface Layers

Important: If the case depths determined from two microhardness tracks differ by more than 0.1 mm or 10%, DIN EN ISO 18203 requires that the test be repeated.

Background: Due to sample preparation and especially scattering within the material, from processing and imperfections in the material (e.g. inhomogeneities like banding), variations in hardness across the cross-sectional area can occur. Such variations are completely normal and inherent to the material. Important: The deeper the induction hardening depth (SHD), the greater these fluctuations become. At depths of 3 to 4 mm, variations of up to 0.3 mm are to be expected on a round shaft.

In contrast, eddy current testing inherently provides results that represent the entire volume of material within the coil taking the signals. The magnetic flux density uniformly permeates the test area in all directions. As a result, any variations in microstructure distributed around the diameter are averaged out in the eddy current response, leading to more consistent results.

In general, the Vickers indentationer is well suited for determining average properties as Vickers hardness (HV) - used for calibration of the AI prediction - is not altered by the choice of the test force, from 25 gf to 1000 gf, because the indentation geometry is constant as a function of indentation depth.

Recommended Procedure for the Initial Trial

- Harden 15 parts using Setting A (production setting for nominal target hardness) in direct sequence. Prepare 1 (one) of these parts carefully for microhardness test (in accordance with section 7.3 of DIN EN ISO 18203) and perform destructive testing.
- For each test location, a hardness profile shall be determined from the surface under examination to the point where the hardness corresponds to the core hardness (total hardened depth, THD, as defined in ISO 18203). The spacing between successive indentations shall comply with ISO 18203, but shall not exceed 0.20 mm for HV1 and 0.50 mm for HV5.
- Harden 15 parts using an alternative Setting B (e.g., target hardness at the lower acceptance limit) in direct sequence. Perform destructive testing on one part.
- Calibrate the AI software using five parts from the remaining parts of Group A and the result of the destructive microhardness test of the initial one part of Group A.
- Use the AI to predict the case depth for the remaining 9 parts of Group A and for all remaining parts in Group B.
- In general, verify uniformity within the groups and check whether the expected minimal scatter of the eddy current results within each group is present. If any part shows repeatedly results with significant deviations from the overall group behavior, this part must be destructively tested to determine its actual values. The verification of the uniformity within each group understandably requires a sufficient number of tested parts. Therefore, a quantity of 15 parts per group is recommended for typical part families, such as constant velocity joints. For large and correspondingly expensive components, such as drive axles, a smaller sample size may also be considered.

For further Advanced Trials:

- Produce and test additional groups of 15 parts each, using additional Settings C, D, … (e.g., near lower and upper acceptance limits Min IN, Min OUT, Max IN and Max OUT)
- Produce parts with parameter settings which simulate known error influences in the hardening process or that lead to out-of-tolerance results.

3.3 Process Validation (Step 2)

According to the widely used CQI-9 Standard "Special Process: Heat Treat System Assessment", the process owner is fundamentally responsible for the control and execution of their hardening process. The conventional destructive laboratory testing using microhardness profile analysis has been the established and widely accepted method for in-process quality monitoring for decades.

Therefore, when transitioning to an alternative test method, it can be expected that end customers – particularly in industry sectors such as automotive, bearing, or aerospace – will require involvement in this decision-making process. In such cases, the process owner may be obligated to perform a validation of the new eddy current-based, non-destructive method against the established destructive test procedure prior to its implementation.

For this type of practical and application-oriented process validation of the AI-supported eddy current testing, the following procedure is recommended.

Recommended Procedure for Process Validation

- The remaining 14 parts from one or more groups (A, B, C …) manufactured for the Initial Trials / Advanced Trials (as described in section 3.2) are to be sectioned at the identical positions of the prior eddy current testing and prepared in the laboratory. At three radially offset positions (each 120° apart), two (or more) microhardness profiles shall be determined on the specimens in accordance with applicable standards.
- Subsequently, the mean hardness values at each defined depth increment (e.g., 0.5, 1.0 … up to 4.0 mm) are calculated across all specimens.
- Based on these mean values and the corresponding absolute maximum and minimum values at each depth, a scatter band diagram is created in accordance with Figure 3. This diagram displays a mean hardness profile curve along with the associated maximum and minimum boundary curves.
- The AI-predicted values per part are then plotted into this diagram. If all AI predictions lie within the minimum-maximum scatter band of the conventionally determined curves, the validation – or method comparison – is deemed successfully completed.
- As an alternative, the mean values and differences between the AI-predicted SHD values and the values obtained through destructive SHD testing can be plotted and evaluated in a Bland-Altman plot (as shown in Figure 4 - Appendix 1). If any values fall outside the defined tolerance limits, they must be examined for potential measurement errors, and the test should be repeated in the relevant area if necessary.

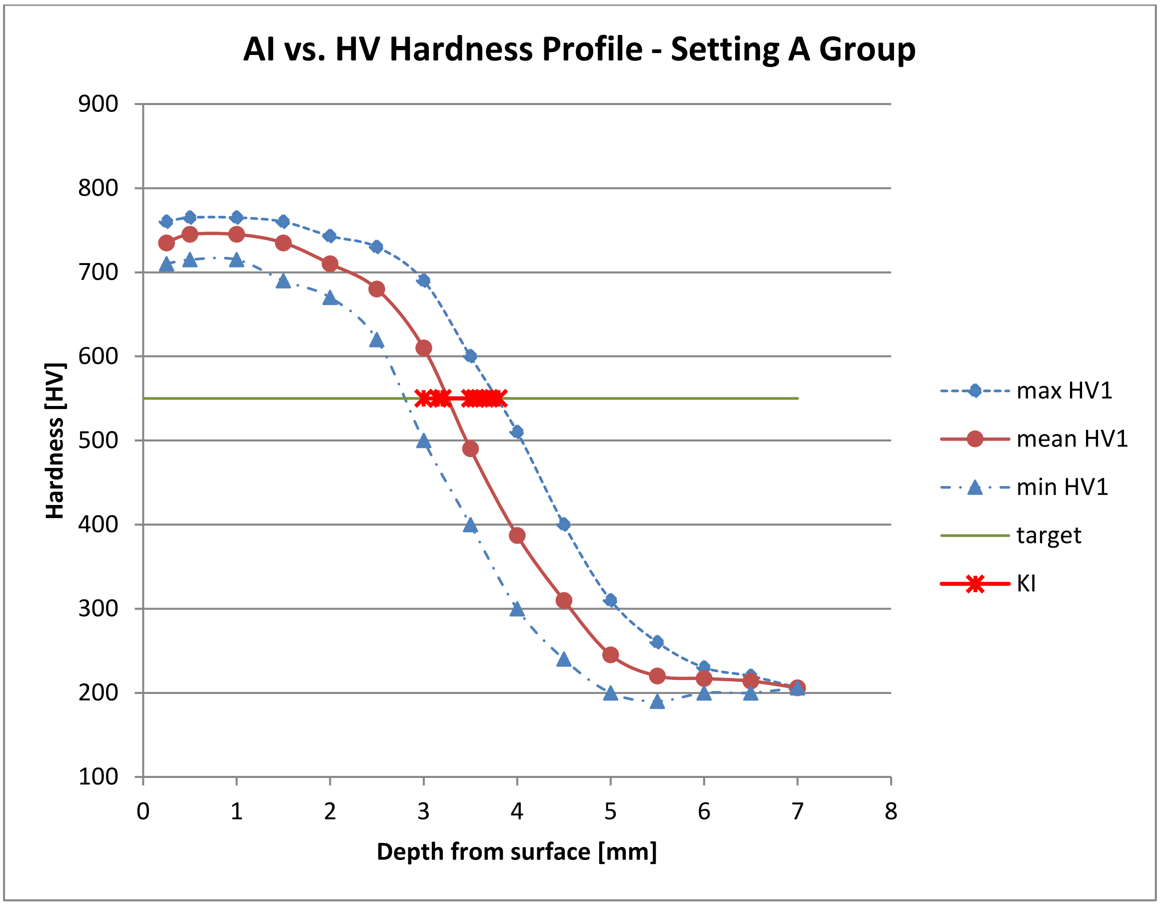

Figure 3: Scatter band diagram (according to [13] HORSCH A.*) comparing the AI prediction with the microhardness test (SHD) on 15 test specimens from a uniformly manufactured, consecutively induction-hardened batch of workpieces, as described in Section 3.2.*

4. Practical Considerations

Given the importance of selecting an appropriate testing method, it is advisable to conduct the SHD tests required for process validation in an independent laboratory, and thus outside of a production-oriented lab environment.

Following the introduction of eddy current-based AI prediction as the standard method for in-process monitoring, regular comparative testing with randomly sampled SHD tests should continue. This could include, for example, once daily at the beginning of a project, later once per week, or at the start and end of a production order, in order to verify long-term process stability and ensure method reliability.

Furthermore, comparative testing is mandatory in the event of changes in material, geometry, or hardening process parameters. Appropriate work instructions and procedural guidelines must provide the necessary framework for this.

5. Conclusion

A systematically executed process validation enables a reliable assessment of whether an AI-supported eddy current testing method can replace the destructive case depth measurement (SHD). This requires a robust data foundation, carefully trained AI models, and the definition of objective comparison criteria.

If successfully validated, the AI-based eddy current approach offers significant potential for improving efficiency – both in laboratory workflows and in achieving 100% in-line testing in production environments.

In addition to increasing throughput and reducing testing costs, this method supports predictive quality control, early detection of process deviations, and minimized scrap rates with its connected CO2 reductions. It also opens the door for digital documentation and traceability, further aligning with Industry 4.0 standards and modern quality assurance strategies.

Wolfgang Korpus
CTO - ibg NDT Technology

2025-04-08
Rev 06

Appendix 1: Bland-Altman Analysis

The Bland-Altman analysis is a well-established statistical tool for assessing the agreement between two test methods – particularly when evaluating a new method (e.g., AI-supported eddy current testing) against a reference method (e.g., microhardness profile testing). Compared to pure correlation analysis, it offers several advantages by explicitly revealing systematic deviations between methods.

Key aspects of Bland-Altman analysis:

- The differences between the two methods are plotted against their mean values.
- The mean of the differences represents the so-called “bias” – the systematic deviation between the methods.
- The 95% confidence limits (Limits of Agreement, LoA) indicate the range of variation in the deviations.
- An ideal outcome is achieved when the bias is approximately zero and the deviations are randomly distributed within the confidence limits, without any systematic trend.

Significance of an Existing Bias:

- A positive bias indicates that the AI prediction systematically yields higher values than the SHD test.
- A negative bias suggests a systematic underestimation by the AI.
- A bias may be acceptable, provided it is consistent and small – or if it can be compensated for through correction.

Possible Causes of Bias:

- Differences in surface condition (affecting ET signal response)
- Material inhomogeneities or geometric variations
- Inaccuracies in the AI training dataset
- Systematic errors in sample preparation or in the SHD test itself (e.g., incorrect placement of the first microhardness indentation)
- Imbalance in training data across different hardness zones

The Bland-Altman analysis thus provides not only an assessment of measurement accuracy, but more importantly, a clear evaluation of the systematic reliability of the new testing method. This makes it particularly well-suited for the validation of AI-based inspection techniques.

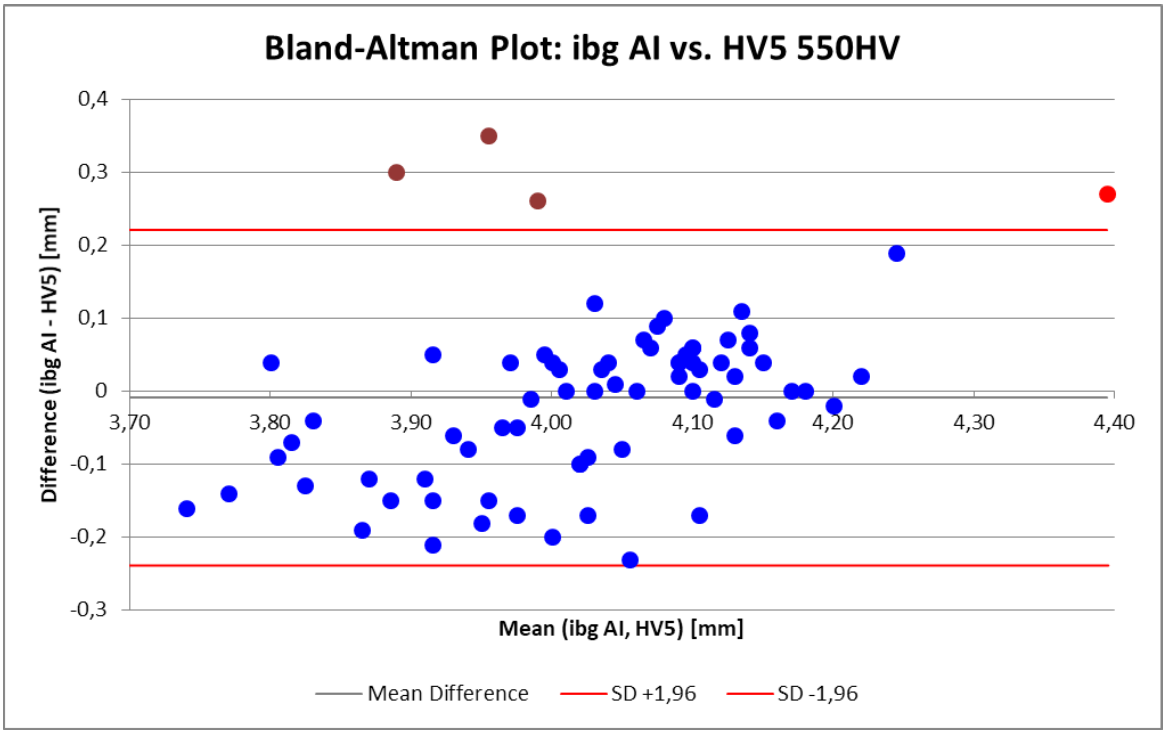

Figure 4: Bland-Altman diagram comparing the AI prediction with the microhardness profile test (SHD) from a multi-week comparison study conducted by a customer laboratory. The diagram shows a slight negative bias as well as four values outside the confidence limits. Upon later review, three of these values were identified as measurement errors (dark red) and one as a deviation in the eddy current test (light red).

References

[1] ASTM E92, *Standard Test Methods for Vickers Hardness and Knoop Hardness of Metallic Materials*

[2] ASTM E384, *Standard Test Method for Microindentationation Hardness of Materials*

[3] CQI-9 - 4th Edition, *Special Process: Heat Treat System Assessment - AIAG*

[4] DIN EN ISO 4885, *Ferrous materials – Heat treatments – Vocabulary*

[5] DIN EN ISO 6507-1, *Metallic materials – Vickers hardness test – Part 1: Test method*

[6] DIN EN ISO 6507-2, *Metallic materials – Vickers hardness test – Part 2: Verification and calibration of testing machines*

[7] DIN EN ISO 14577-1, *Metallic materials – Instrumented indentation test for hardness and materials parameters – Part 1: Test method*

[8] DIN EN ISO 15548-1, *Non-destructive testing – Equipment for eddy current examination – Part 1: Instrument characteristics and verification*

[9] DIN EN ISO 15548-2, *Non-destructive testing – Equipment for eddy current examination – Part 2: Probe characteristics and verification*

[10] DIN EN ISO 15549, *Non-destructive testing – Eddy current testing – General principles*

[11] DIN EN ISO 15787, *Technical product documentation – Heat-treated ferrous parts – Presentation and indications*

[12] DIN EN ISO 18203:2025, *Steel – Determination of the thickness of surface-hardened layers*

[13] HORSCH A., *Reproducibility of the hardness depth determination CHD – SHD – NHD, ECHT 2020 – European Conference on Heat Treatment MARCH 25 – 27 2020 – Antwerp, Belgium*

[14] LIEDTKE D., WALDENMAIER T., HORSCH A., *Hinweise für die Anwendung der DIN EN ISO 18203 – Bestimmung der Dicke gehärteter Randschichten – AWT – July 2023*

Glossary

Metrological Terms in Testing and NDT (Eddy Current) Context

Based on ISO/IEC Guide 99, ISO 9000, ISO 17025, ISO 10012, and ISO 15548-1:2013

Verification

Verification is the process of providing objective evidence that a measuring instrument, system, or method conforms to defined technical and functional requirements.

It typically includes checking a system's performance against its specifications using reference artifacts, known inputs, or standard conditions.

Standards References:
- ISO/IEC Guide 99:2007 (VIM), Clause 2.44
- ISO 15548-1:2013, Clause 4.3

Application in NDT:
Verification ensures that eddy current instruments reliably perform within their defined operating limits. It does not determine measurement error (as calibration does), but rather checks functional integrity and measurement validity under specific test conditions.

Verification Levels According to ISO 15548-1:2013

Level	Purpose	Performed by	When
Level 1	Routine verification to confirm basic instrument functionality. Typically includes function checks using reference blocks.	Operator/User	Before or during regular use (e.g. daily or per shift)
Level 2	Periodic or post-service verification of specific functional parameters. Involves confirmation against known standards.	Qualified entity/ Manufacturer	At least annually, and after any repair, adjustment, or software update
Level 3	Comprehensive verification by the manufacturer, including full performance evaluation against all technical specifications.	Manufacturer	Once, before first delivery or commissioning

Calibration

Calibration is a documented process that establishes the relationship between the indication of a measuring instrument and the known value of a standard, under specified conditions.
It includes the determination of measurement uncertainty and may involve adjustments to align the instrument with traceable standards.

Standards References:
- ISO/IEC Guide 99:2007 (VIM), Clause 2.39
- ISO/IEC 17025:2017, Clauses 6.4 & 7.6
- ISO 10012:2003

Application in NDT:
Calibration ensures traceability to national or international standards and establishes confidence in measurement accuracy. It is foundational for meaningful verification and validation, and a calibration certificate typically includes uncertainty data and calibration intervals.

Validation

Validation is the process of demonstrating, through objective evidence, that a method, process, software, or system consistently performs as intended under real or simulated conditions.
It assesses the method’s ability to produce reliable and accurate results for its specific application.

Standards References:
- ISO 9000:2015, Clause 3.8.13
- ISO 10012:2003, Clause 7.3

Application in NDT:
Validation is especially relevant for non-standardized test methods, AI-based prediction models, or automated signal evaluation systems. It verifies that a given process produces results that meet the required accuracy, reproducibility, and relevance for the application. In regulated industries, validation is essential prior to method adoption.