SURF CON INSPECT

WHITE PAPER



Better classification: blending rules and images to reduce false negatives

SurfConInspect White Paper
Series 7/10

About Us

WHO WE ARE

SurfConInspect (SCI) is a European project funded by the Research Fund for Coal and Steel (RFCS). Our consortium brings together leading steel producers, research institutes, technology developers and innovation experts.

The partners combine industrial expertise, academic knowledge and advanced digital solutions to transform the way steel quality is controlled.



OUR MISSION

The steel industry faces increasing pressure to become more sustainable, efficient and competitive—key goals of the European Green Deal.
Our mission is to enable zero-defect manufacturing in flat steel production by:

- Detecting surface defects earlier and more accurately.
- Reducing waste and improving resource efficiency.
- Supporting operators with advanced visualization tools.
- Providing real-time control actions directly in production lines.

MAIN ACTIVITIES

- Industrial trials in tin-plate and automotive steel production.
- Development of AI-based inspection methods.
- Application of AR for real-time operator support.
- Creation of open datasets to enhance research and industrial collaboration.
- Dissemination of results to industry, research and society through events, publications and digital platforms.

EXPECTED RESULTS

- Zero-defect production through optimized inspection and control.
- Reduced waste and energy consumption, contributing to the
  European Green Deal.
- Higher competitiveness for European steel producers.
- Transferable technologies with potential applications beyond the steel sector.
- Improved working conditions, with operators supported by innovative visualization and decision-making tools.

Executive Summary

In automatic surface inspection systems for flat steel production, classification is the step that determines whether a detected anomaly is assigned to a defect category and, if so, which one. The quality of this assignment has direct operational consequences: a misclassified defect either reaches downstream processing unchallenged, contributing to coil breaks, coating failures, or customer claims, or it triggers an unnecessary intervention, generating scrap and reducing line yield. Both types of error carry measurable economic costs, and the balance between them is the central performance parameter of any classification system.

For decades, rule-based classification has been the industrial standard in automatic surface inspection systems (ASIS). Rule-based systems assign defect categories by evaluating a set of predefined conditions against the measured attributes of each detected event: its size, shape, contrast, and position on the strip. These systems are transparent, auditable, and computationally efficient. They also have a fundamental structural limitation: their performance is bounded by the ability of handcrafted rules to capture the full variability of defect appearance under real production conditions. In environments where the same defect type can manifest with different visual characteristics depending on material grade, process conditions, and surface state, and where critical defects can appear visually similar to non-critical events, rule-based systems generate classification errors that neither rule refinement nor threshold adjustment can eliminate.

The source of these errors lies in the nature of the classification problem itself. Surface defects in flat steel production are not defined by fixed visual signatures; they are defined by physical properties, geometry, material state, position on the strip, that express themselves through variable visual appearances depending on context. WP1 and WP4 of this series establish that the geometric characterisation of defects such as dents, abrasions, and shells requires measurement of depth and profile sharpness, parameters that are not directly accessible to intensity-based imaging. A rule-based system that classifies based on appearance alone cannot reliably access the physical properties that determine defect severity, and the consequence is a persistent rate of false negatives for critical defects whose visual signature falls outside the range captured by the rule set.

Hybrid classification approaches address this limitation by combining rule-based logic with image-based analysis. Rather than replacing the rule set, image-based methods introduce an additional classification pathway that operates on the spatial structure of the defect image directly, learning associations between visual patterns and defect categories from labelled training data. The two pathways are complementary: rule-based logic provides stable performance for well-characterised defect types whose appearance is consistent, while image-based analysis extends coverage to defect instances whose visual variability exceeds the expressive capacity of predefined rules. Within the SurfConInspect project, this hybrid approach is developed and validated in direct relation to four industrial use cases – automated cutting control, process-speed optimisation, and tin-plate online grading at the thyssenkrupp Rassselstein (RAS) pickling and tinning lines in Andernach, and online grading at the the Salzgitter Flachstahl (SZFG) electrolytic galvanising line in Salzgitter – where the reduction of false negatives for critical defect categories is a primary operational objective.

1. The classification problem in automatic surface inspection
2. Rule-based classification: principles and structural limitations
3. The role of visual variability in classification error
4. Image-based classification: operating on spatial structure
5. The hybrid classification architecture
6. Training data: domain adaptation and synthetic generation
7. Integration with the quality processor architecture
8. Operational considerations and system maintenance
9. The SurfConInspect perspective
10. Conclusions

The classification problem in automatic surface inspection

Automatic surface inspection systems for flat steel production operate as two-stage pipelines. In the first stage, detection algorithms identify regions of the strip surface that deviate from the normal background texture in terms of intensity, contrast, or, in systems equipped with three-dimensional sensing of the kind described in WP5 of this series, height profile. In the second stage, classification algorithms assign each detected region to one of a predefined set of categories: a specific defect type, a non-critical surface event, or an ambiguous case requiring human review. Detection determines what is found; classification determines what it means operationally.



The performance requirements for classification in an industrial inspection context are demanding in ways that are not always fully appreciated outside the production environment. Classification must be performed in real time, at the high strip speeds, with latency that allows results to be available for the quality processor and online quality supervisor described in WP6 within the time window relevant for process intervention. It must maintain consistent performance across the full range of material grades, surface finishes, and process conditions encountered in production, from thin tin-plate strip at RAS pickling line to automotive-grade steel at SZFG electrolytic galvanising line, without requiring manual reconfiguration each time a new product type is introduced. And it must achieve a rate of false negatives for critical defect categories that is low enough to prevent the systematic passage of defective material to downstream stages.

This last requirement is the most consequential when it fails. A false negative in surface inspection is not simply a classification error; it is a quality failure with direct downstream implications. A dent classified as a water droplet reflection at the pickling-line inspection stage does not disappear from the strip: it reaches the cold rolling mill with its full geometric severity intact, where it concentrates stress during thickness reduction and may initiate a crack that propagates to strip rupture. As documented in WP4, approximately 90 percent of coil break events at tandem cold rolling mills are traceable to topographic defects in the near-edge zone of the strip, and each coil break event carries substantial direct costs in the form of lost line availability, scrap, and maintenance and safety intervention, in addition to the downstream customer and reputational impact of the resulting quality failure. The classification problem in surface inspection is therefore not a matter of statistical optimisation in the abstract, but of specifically reducing false negatives for the defect categories that carry the highest downstream risk.

This asymmetry between the cost of false negatives and false positives defines the operational objective that classification system design must address. Rule-based systems, as the following section examines, are structurally ill-equipped to meet this objective for defect categories whose visual appearance is variable or context-dependent, a limitation that motivated the classification improvement component of the SurfConInspect project from its inception.

Rule-based classification: principles and structural limitations

Rule-based classification systems operate by evaluating a set of conditions against a vector of numerical features extracted from each detected event. These features typically include geometric descriptors, the length, width, area, and aspect ratio of the event's bounding box, and intensity descriptors, mean brightness, contrast relative to background, and texture statistics derived from the image region associated with the event. A rule is a logical expression that maps a combination of feature values to a classification outcome: if the event satisfies a specified set of conditions simultaneously, it is assigned to the corresponding defect category. A rule set can be automatically derived from training data by means of machine-learning algorithms or manually constructed in an empirical process driven by domain knowledge accumulated by inspection engineers and quality specialists over years of production experience at individual facilities.

The transparency and auditability of rule-based systems are among their most valued operational properties. Every classification decision can be traced back to the specific rule that produced it, and the rule can be inspected, verified, and modified without requiring access to training data or internal model representations. Production managers and quality engineers can audit the classification logic, verify that it reflects current quality requirements, and make targeted adjustments when new defect types emerge or product specifications change. The computational efficiency of rule evaluation is also significant: a rule-based classifier can process thousands of events per second with negligible computational overhead, a property well suited to the real-time processing constraints of high-speed inspection lines described in WP5.

The structural limitation of rule-based classification lies in the relationship between the feature space it operates on and the physical properties that determine defect severity. Features such as event size, contrast, and aspect ratio are derived from visual appearance under the specific illumination and surface conditions prevailing at the moment of acquisition. They are not direct measures of physical properties. A dent that appears as a compact, high-contrast feature under dry, stable illumination conditions may appear as a diffuse, low-contrast feature on a wet strip surface where specular reflections from the water film, described in detail in WP2 of this series, modify the local intensity distribution. A shell defect that exhibits a characteristic elongated shape at one material thickness may appear more compact and ambiguous at a different thickness where the surface deformation pattern changes. In both cases, the visual feature values that the rule set was designed to capture have shifted away from the expected range, and the rule that was correct under one set of conditions produces a misclassification under another.

This variability is not a deficiency of the rule set itself but a consequence of the physical diversity of the inspection problem. No fixed set of feature thresholds can reliably capture this diversity without either becoming so permissive that it generates frequent false positives or so restrictive that it generates frequent false negatives. The performance ceiling of rule-based classification is therefore determined not by the skill with which the rules are constructed, but by the intrinsic variability of defect appearance relative to the feature space on which the rules operate. In practice, this ceiling manifests most clearly in the categories of defects that are both visually variable and operationally critical, precisely the categories that, as established in WP4, are most directly associated with coil break risk and coating adhesion failures.

The performance limitation of rule-based systems in industrial ASIS has been documented in the research literature. The reliable quantification of recall for relevant defect classes under running production conditions remains one of the most significant open challenges in surface inspection, a challenge that persists despite the substantial improvements in camera resolution and processing speed achieved over the preceding two decades. VDI/VDE/VDMA standard 2632 Part 4.2, which defines methods for evaluating the classification performance of surface inspection systems in flat steel production, explicitly addresses the difficulty of measuring recall under real production conditions as distinct from controlled dataset evaluation, a difficulty that reflects precisely the gap between rule-based performance on static benchmarks and rule-based performance in continuous production.

The role of visual variability in classification error

The variability of defect visual appearance in flat steel inspection is not random noise; it is a structured phenomenon with identifiable physical causes. Understanding these causes is a prerequisite for designing classification approaches that are robust to them, and it provides the physical basis for the hybrid approach developed in the following sections.

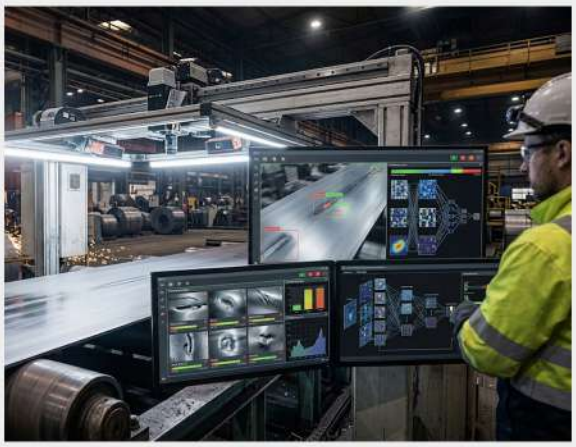
The most direct source of visual variability is the surface state at the moment of inspection. WP2 shows that even micrometric variations in film thickness can produce significant changes in the local reflectance of the surface through modifications of the interference and absorption condition. A defect whose visual signature is defined primarily by its reflectance contrast relative to the surrounding surface will therefore appear differently depending on whether the surrounding surface uniform or structured. This surface state modulation is one of the primary contributors to the systematic gap between false negative rates measured on static validation datasets and those observed in continuous production.

Material grade and surface roughness introduce a second source of variability. Different steel grades, including the tin-plate and automotive products that constitute the SurfConInspect use cases at RAS and SZFG respectively, have different surface textures arising from their rolling history and metallurgical composition. These textures interact with the illumination system in ways that modulate the background against which defects must be detected and classified. A defect of a given type and geometry will appear with higher contrast on a smooth, low-roughness surface than on a textured one where the background itself contains spatial frequency components similar to those of the defect. Classification rules calibrated on one surface type may therefore underperform on another, even when applied to geometrically identical defects.

Process condition variability, changes in pickling acid concentration, temperature fluctuations, and speed variations during coil transitions, introduces a third source. These conditions affect the chemical and physical state of the strip surface at the time of inspection, modifying both the intrinsic reflective properties of the steel and the distribution of surface residues that contribute to the visual appearance of the inspection image. Classification performance in machine-learning-based imaging systems is significantly affected by surface texture variation across material types, a finding consistent with the operational experience of inspection engineers at flat steel production facilities. The interaction between these sources of variability is not additive but multiplicative: a defect at the boundary of classification reliability under one source of variability may become systematically misclassified when a second source acts simultaneously, which explains why false negative rates measured on controlled datasets routinely underestimate those observed in production.

Image-based classification: operating on spatial structure

Image-based classification methods approach the classification problem from a fundamentally different starting point than rule-based systems. Rather than evaluating handcrafted features against predefined thresholds, they operate directly on the spatial structure of the image region associated with each detected event, learning to associate patterns of pixel values with defect categories from collections of labelled examples. The critical distinction is not between learned and non-learned methods per se, but between feature-based and image-based representations: the former reduces the defect image to a small number of scalar descriptors before classification, while the latter preserves the full spatial information content of the image and uses it as the basis for the classification decision.



The application of convolutional neural networks (CNNs) to surface defect classification in steel production has been demonstrated in literature showcasing that a max-pooling CNN trained on steel defect images achieved classification performance competitive with handcrafted feature-based methods on standard benchmark datasets, with the additional property of learning discriminative representations directly from pixel data without domain-specific feature engineering. This is true also for multiple defect types and surface conditions, confirming that CNN-based approaches generalise better across surface texture variations than feature-based classifiers trained on the same data. Within the SurfConInspect project, these findings motivate the adoption of image-based classification as a complementary pathway to rule-based logic, particularly for the defect categories, dents and shell defects in wet pickling environments, where surface state variability most severely degrades rule-based performance.

The training data requirement is central to the performance of image-based classifiers and constitutes their primary engineering challenge in industrial deployment. A classifier trained exclusively on defect images acquired under dry conditions will not generalise reliably to wet conditions, not because image-based methods are inherently limited in this respect, but because the training distribution does not represent the production distribution. This misalignment between training and production conditions is a well-documented source of performance degradation in deployed image-based inspection systems more broadly, a risk in the context of automated composite inspection, where an unrepresentative and undersized training set led to model overfitting rather than reliable generalisation. This motivates the data strategies developed within SurfConInspect that are addressed in detail in section six. An additional property of image-based classification that is operationally relevant is the ability to provide classification confidence estimates alongside category assignments, enabling selective escalation of borderline cases for human review, a capability that the quality processor architecture described in WP6 is designed to exploit.

The hybrid classification architecture

The limitations of rule-based classification and the complementary strengths of image-based methods motivate a hybrid architecture in which both approaches contribute to the classification decision. The hybrid architecture is not simply an ensemble of two independent classifiers whose outputs are combined by a fixed voting rule; it is a structured integration in which the two approaches address different aspects of the classification problem according to their respective capabilities and the operational requirements of the specific use case.

In the architecture developed within SurfConInspect, rule-based classification retains its primary role for defect categories whose visual signatures are sufficiently consistent and well-characterised that handcrafted features provide reliable discrimination. For these categories, the rule-based pathway produces correct classifications across the range of production conditions typically encountered, and the computational efficiency and transparency of rule evaluation are preserved without compromise. The image-based pathway is invoked selectively for events that the rule-based pathway identifies as ambiguous, either because the feature values fall near the boundary between two categories, or because the event does not match any rule with sufficient confidence. This selective activation limits computational overhead and maintains the auditability of the classification logic for the majority of events processed by the system.

For events routed to the image-based pathway, the classifier receives the image region of the event as input and produces a classification output that is combined with the rule-based result according to a defined integration logic. The integration can take several forms depending on the operational requirements and the cost structure of false positives and false negatives for the defect categories in question: a confidence-weighted combination, a sequential fallback in which the image-based result overrides the rule-based one only when the latter falls below a defined confidence threshold, or a parallel evaluation in which both outputs are presented jointly to the operator through the inspection interface described in WP5. The choice between these strategies is not fixed but reflects the specific operational context, the downstream risk profile of the material being inspected and the intervention options available at the relevant process stage.

A critical design principle of the hybrid architecture is the preservation of interpretability. One of the operational risks of introducing image-based methods into industrial classification systems is the opacity of the classification decision: unlike a rule that can be read and verified by a quality engineer, a learned classifier does not provide an explicit justification for its outputs. In the SurfConInspect hybrid architecture, this risk is managed by confining the image-based pathway to the ambiguous cases that the rule-based system cannot resolve with sufficient confidence, and by logging the activation of the image-based pathway alongside the classification result for every event. This logging provides the audit trail that industrial quality management requires, enabling quality engineers to monitor the performance of the hybrid system over time and to identify systematic patterns in the cases where image-based arbitration is invoked, patterns that may indicate drift in production conditions or a need for classifier recalibration.

The expected outcome of this architecture for the critical use cases of SurfConInspect, automated cutting control for heavy defects at the RAS pickling line and online quality grading at the SZFG electrolytic galvanising line, is a measurably lower false negative rate for dents, shell defects, and severe abrasions in near-edge strip zones, without requiring replacement of the rule-based infrastructure that industrial operators and quality engineers at both facilities have developed and validated over years of production experience.

Training data: domain adaptation and synthetic generation

The performance of any image-based classifier is bounded by the quality and representativeness of its training data. In surface inspection for flat steel production, the assembly of training datasets that adequately represent the full range of production conditions is a practical constraint that fundamentally shapes what image-based classification can achieve in industrial deployment, and it is one of the central challenges that the SurfConInspect project addresses through dedicated research tasks.

The principal challenge is the rarity of critical defect events relative to the volume of strip inspected. In a well-controlled production process of the kind operated at RAS and SZFG, the frequency of heavy dents, shell defects, and other critical anomalies is deliberately minimised through upstream process optimisation. The consequence is that the training dataset assembled from routine production data is strongly imbalanced in favour of non-critical events. A classifier trained on an imbalanced dataset tends to develop a bias toward the majority class, producing exactly the kind of false negatives, critical defects classified as non-critical, that the hybrid approach is designed to reduce. This dataset imbalance problem is well documented in the surface inspection literature: the NEU surface defect dataset, one of the most widely used public benchmarks, contains only 1,800 images across six defect classes, a scale that is not representative of the diversity encountered in productive inspection systems trained on more than fifty defect classes.

Domain adaptation addresses the imbalance by exploiting inspection data collected by other systems, on the same production line at different periods, or on similar lines at other facilities, to augment the training set for underrepresented defect categories. Domain adaptation methods compensate for the systematic differences between the imaging characteristics of different systems and surface conditions, transferring classification knowledge across domains while preserving category-specific information relevant to the target production environment. Within SurfConInspect, approaches based on the Sliced Wasserstein Discrepancy are investigated to quantify and minimise the distributional distance between source and target domains, enabling the transfer of classification knowledge between the RAS and SZFG inspection installations, which differ in material grade, surface finish, and process conditions, without requiring full retraining from labelled data at each site.

Synthetic data generation provides a complementary strategy. Generative approaches including Variational Auto-Encoders (VAE) and Generative Adversarial Networks (GAN) can produce synthetic defect images whose statistical properties match those of real production data, augmenting the training set with representations of underrepresented conditions including wet-surface variants of critical defect types. Within SurfConInspect, these approaches are applied specifically to the generation of wet-surface variants of dent and shell defect images, targeting the surface state conditions, variable liquid film thickness, acid residues, illumination instability, that most directly drive the performance gap between rule-based and hybrid classifiers.

Both strategies reflect a broader principle that underpins the SurfConInspect approach to classification improvement: the performance of image-based methods in industrial inspection is not primarily a function of algorithmic sophistication but of data quality and representativeness. This principle also motivates the open data initiative developed within the project, through which a curated dataset of labelled surface defect images, targeting a minimum of 30,000 examples across standardised defect categories is made available to the research community as a shared benchmark for classifier development and evaluation. By creating a public dataset at a scale and with a category structure representative of industrial inspection requirements, the project aims to accelerate research progress in surface defect classification beyond the boundaries of individual production facilities and inspection system vendors.

Integration with the quality processor architecture

The hybrid classification system does not operate in isolation. In the SurfConInspect framework, it is a component of the broader quality control architecture described in WP6 of this series, and its outputs are consumed by the quality processor and online quality supervisor that translate classification results into quality indices and process actions. The interface between the classifier and the quality processor defines the operational context in which classification performance is ultimately evaluated.



The quality processor receives, for each detected event, a classification label and, in systems that provide it, an associated confidence measure. The quality model that maps classification results to quality indices is sensitive not only to the category assigned to each event but also to the spatial distribution of classifications across the coil. As described in WP6, a coil that generates isolated events classified as dents in the central strip zone presents a different risk profile from one that generates a cluster of dent classifications near the strip edge, where the elevated tensile stress during cold rolling amplifies the stress concentration effect of each individual defect, as established in WP4. This spatial and statistical context is lost if classification is evaluated as an isolated accuracy metric, and it is preserved only when the classifier output is assessed within the operational workflow that uses it.

The false negative rate of the classifier therefore translates directly into the false negative rate of the cutting decision that the quality processor supports. A dent misclassified as a water droplet reflection does not generate a severity score in the quality processor; it does not contribute to the cutting threshold evaluation; and it does not trigger a quality event in the online quality supervisor. It passes through the entire quality control chain undeclared, reaching the downstream rolling mill with the same geometric risk it would have carried if no inspection system had been present. In this sense, classification is not a preparatory step toward quality control but a constitutive component of it: the quality of the classification is inseparable from the quality of the process decision. The integration of the hybrid classifier within the SurfConInspect quality control architecture is therefore not an optional enhancement but a structural requirement for the system to meet its operational objectives.

This integration perspective also defines how the performance of the hybrid classification system should be validated in the SurfConInspect industrial demonstration phase. Standard classification metrics, accuracy, precision, recall, measured on held-out validation datasets provide a necessary but not sufficient basis for performance assessment. What ultimately matters is the rate at which critical defects proceed to downstream stages without triggering a quality event, and this rate can only be measured through operational monitoring that connects classification outputs at the pickling stage with quality outcomes at the cold rolling mill or coating line. This longitudinal validation, tracing the downstream consequences of classification decisions made upstream, is one of the key activities of the SurfConInspect demonstration at RAS and SZFG and provides the empirical foundation on which the performance claims of the hybrid classification approach are grounded.

Operational considerations and system maintenance

The deployment of a hybrid classification system in continuous production introduces maintenance requirements that differ in important respects from those of a purely rule-based system. A rule set, once validated, is static: its performance on a given defect category remains constant as long as the production conditions and material specifications assumed in its construction remain stable. An image-based classifier is a learned model whose performance may degrade over time if the distribution of production conditions shifts away from the distribution represented in its training data. This degradation can occur gradually as material grades, process conditions, or equipment characteristics evolve over the operational lifetime of the inspection system, a particularly relevant concern in long-running production facilities such as the RAS Andernach site, where ASIS have been in continuous operation for an extended period, or it can occur abruptly following a significant change such as the installation of new rolls, a modification of the acid bath chemistry, or the introduction of a new product range. Monitoring for classifier performance degradation requires a mechanism that tracks classification outputs against downstream quality outcomes on a continuous basis, rather than relying on periodic validation campaigns against static datasets. In the SurfConInspect architecture, the traceability record generated by the quality processor and online quality supervisor provides the data infrastructure for this continuous monitoring: every classification result is logged alongside the downstream quality outcome for the corresponding coil section, enabling statistical process control methods to detect systematic shifts in classification performance before they accumulate into operational quality failures.

Recalibration of the image-based component, when required, involves updating the training dataset with recently acquired labelled examples that reflect current production conditions and retraining the classifier on the updated dataset. In practice, the labelling effort is reduced by targeting annotation at the ambiguous cases identified by the hybrid system's confidence monitoring, the events for which the image-based pathway was invoked and the classification confidence was low, rather than applying it uniformly across the full inspection output. This selective labelling strategy, consistent with active learning principles established in the statistical learning literature, concentrates the annotation effort where it contributes most to classifier improvement, reducing the per-unit cost of dataset maintenance.

The transparency requirements of industrial quality management, including compliance with VDI/VDE/VDMA standard 2632 Part 4.2, which defines performance evaluation procedures for ASIS classification in flat steel production, also apply to the maintenance of the hybrid system. Changes to the classification logic, whether in the rule set or in the image-based model, must be documented, validated against a defined benchmark dataset, and approved through the quality management process before deployment in production. The logging architecture described in section five, which records the activation of the image-based pathway alongside the classification result for every event, provides the audit trail that this validation process requires.

The SurfConInspect perspective

Within the SurfConInspect project, the development of hybrid classification approaches is positioned as a necessary evolution of the ASIS capability that underpins all of the project's industrial use cases. Reliable classification is not a standalone objective; it is the prerequisite for every downstream function of the quality control architecture. The quality processor described in WP6 cannot generate meaningful quality indices from unreliable classification outputs. The online quality supervisor cannot produce operationally justified control actions if the classification results it receives systematically misrepresent the severity distribution of defects on the coil. And the automated cutting control that is the most directly measurable output of the SurfConInspect project at the RAS pickling line, targeting a 20 percent reduction in coil break frequency at downstream tandem rolling mills, cannot achieve this target if the dents and shell defects responsible for those breaks are misclassified before they reach the quality processor.

The project addresses this dependency at multiple levels. At the sensor level, the introduction of three-dimensional topographic measurement described in WP1 and WP5 provides a depth channel that is largely independent of the surface state variability that modulates the optical signal, reducing the ambiguity in the feature space on which classification operates. The availability of depth data for each detected event, depth, profile sharpness, and boundary geometry, extends the feature set available to both the rule-based and image-based classification pathways, providing access to physical properties of the defect that are directly relevant to its downstream risk and are not accessible through intensity imaging alone. At the algorithmic level, the hybrid architecture combines rule-based and image-based pathways to extend classification coverage beyond what either approach achieves in isolation. At the data level, domain adaptation and synthetic generation strategies address the training data constraint that limits image-based generalisation to the full range of production conditions at RAS and SZFG. And at the community level, the open data initiative creates a shared benchmark resource that enables systematic benchmarking of classification approaches across the broader research community.

The expected performance improvement from this multi-level approach is expressed in terms of the recall of critical defect categories, dents, shells, and severe abrasions in near-edge strip zones, relative to the baseline performance of the rule-based systems currently deployed at the industrial partner sites. Quantitative targets are defined in the SurfConInspect project plan as part of the assessment criteria specified for the quality processor and classification improvement deliverables, and their achievement is validated through the longitudinal monitoring mechanism that connects pickling-stage classification outputs with quality outcomes at downstream cold rolling and coating facilities. This validation methodology, which grounds performance claims in operational evidence rather than benchmark accuracy, reflects the project's consistent emphasis on industrial relevance as the standard against which inspection system advances must ultimately be measured.

Conclusions

Classification is the step in the ASIS pipeline that determines whether detection results translate into quality control actions. Its performance is a direct determinant of inspection system effectiveness: classification errors that allow critical defects to pass through the quality control chain produce exactly the downstream failures, coil breaks, coating defects, customer claims, that the inspection system is designed to prevent. The asymmetric cost of false negatives relative to false positives in critical defect categories defines the performance objective that classification system design must address, and it is the criterion against which the hybrid approach developed within SurfConInspect is evaluated.



Rule-based classification provides a transparent, auditable framework that has served as the industrial standard in ASIS for decades. Its structural limitation is the inability to capture the full variability of defect visual appearance under real production conditions, variability driven by surface wetness, material grade diversity, and process condition fluctuations that are intrinsic to the pickling-line environment described in WP2, and whose consequences for defect detectability and classification are established in WP4. This variability produces a systematic gap between false negative rates measured on static validation datasets and those observed in continuous production, a gap that neither rule refinement nor threshold adjustment can close.

Hybrid classification architectures address this limitation by combining rule-based logic with image-based analysis in a structured integration that preserves the transparency and operational reliability of the former while extending classification coverage through the variability tolerance of the latter. The training data strategies developed within SurfConInspect, domain adaptation through Sliced Wasserstein Discrepancy methods, synthetic data generation through VAE and GAN architectures, and the open data initiative targeting 30,000 labelled production images (available under https://ckan-surfconinspect.cetic.be) , address the data quality constraint that limits image-based generalisation to the full range of production conditions at the industrial partner sites.

The integration of hybrid classification within the quality processor and online quality supervisor architecture of WP6 ensures that improvements in classification performance translate directly into improvements in quality control decisions. In the SurfConInspect industrial demonstration phase at RAS and SZFG, the longitudinal validation methodology that connects pickling-stage classification outputs with downstream quality outcomes provides the empirical foundation on which the operational case for hybrid classification rests. Classification improvement is not an isolated technical objective but a constitutive element of the zero-defect manufacturing vision that the SurfConInspect project is designed to advance in European flat steel production.

SURF CON INSPECT

The Consortium

SurfConInspect unites a diverse group of partners:

- **Steel producers:** thyssenkrupp Rasselstein (RAS), Salzgitter Flachstahl (SZFG)
- **Research & Academia:** VDEh-Betriebsforschungsinstitut (BFI), Scuola Superiore Sant'Anna (SSSA), CETIC
- **Technology developers:** Clecim France (CLECIM)
- **Communication & dissemination experts:** yourscienceEDU (YSE)



STAY CONNECTED

Follow our updates, publications and events on:
🌐 Project Website: **surfconinspect.eu**
💼 LinkedIn: **SurfConInspect**

Together, we are shaping the future of **sustainable steel production**.



thyssenkrupp · SALZGITTER FLACHSTAHL · BFI Excellence in Applied Research · Sant'Anna Scuola Universitaria Superiore Pisa · cetic Your Connection to ICT Research · CLECIM · YSedu yourscienceEDU