Classifier Tuning of Automated Surface Inspection System

Digital technologies are transforming industry at all levels. Steel has the opportunity to lead all heavy industries as an early adopter of specific digital technologies to improve our sustainability and competitiveness. This column is part of AIST's strategy to become the epicenter for steel's digital transformation, by providing a variety of platforms to showcase and disseminate Industry 4.0 knowledge specific for steel manufacturing, from big-picture concepts to specific processes.

Authors

Liwei Zhang
Lead Engineer, ArcelorMittal Global R&D – East Chicago, East Chicago, Ind., USA
liwei.zhang@arcelormittal.com

Alisha Sawant
New York University, New York, N.Y., USA
ans698@nyu.edu

Brad Masters
Technical Specialist, Quality, HDGL, AM/NS Calvert LLC, Calvert, Ala., USA
bradley.masters@arcelormittal.com

Brian Feagan
Team Manager Inspection System, AM/NS Calvert LLC, Calvert, Ala., USA
brian.feagan@arcelormittal.com

Kyle Harmon
Quality Technical, AM/NS Calvert LLC, Calvert, Ala., USA
kyle.harmon@arcelormittal.com

While automated surface inspection has been continuously improved over time along with computer hardware and software advances, automated surface inspection systems (ASIS) have been widely deployed to many flat carbon steel processes, including hot rolling,[1] picking,[2] tinning, hot-dip galvanizing (HDG),[3] etc., for surface quality control and assurance.

ASIS deployed in the HDG process has a longer history than other processes as the coating line process has ambient temperature and the line speed is slower. There are quite a few commercial ASIS available on the market that can be deployed in the coating line process. They all typically use industrial line scan or area scan cameras to perform real-time image acquisition, the traditional image processing for defect detection and pattern recognition techniques for defect classification.[4] However, the image data processing, pattern recognition algorithms and software graphical user interface (GUI) have been quite different among these commercial systems.

AM/NS Calvert LLC installed an ISRA Parsytec ASIS in 2008. In order to provide higher-quality products, Calvert quality managers have set up initiatives to further improve all the ASIS performance and its uses since 2016. This paper mainly covers the ASIS classifier tuning work for improved ASIS classification performance. It starts with a general review of Parsytec ASIS configurations at AM/NS Calvert #3HDGL (hot-dip galvanizing line), then presents a decision tree type of classifier tuning method, and objectives and procedures using the classifier tuning for galvanized bright field (GI-BF) material at Calvert #3HDGL as an example.

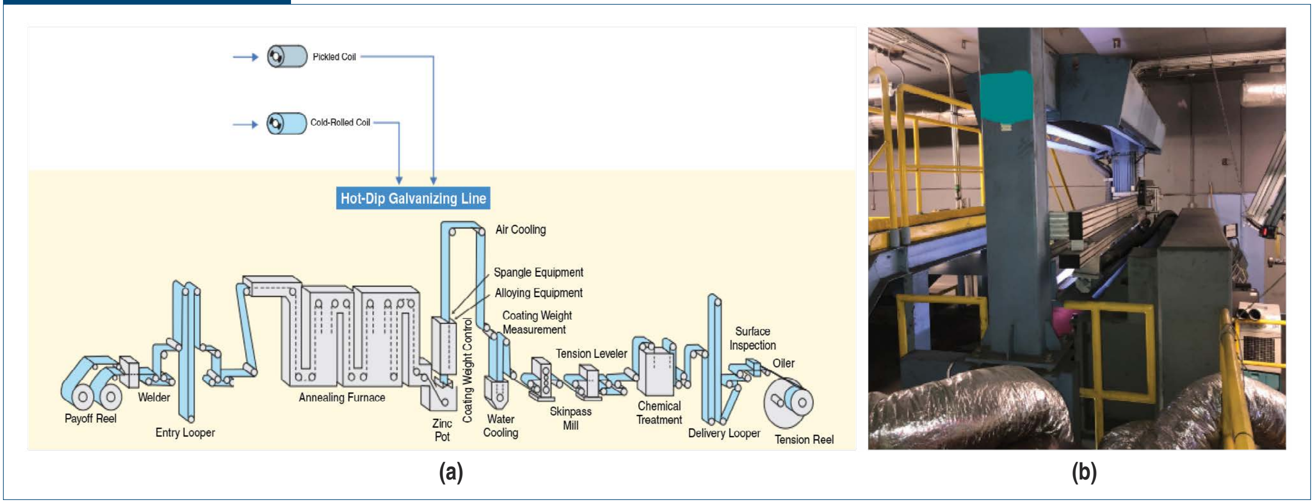
Classifier Tuning Method and Procedures

ASIS Configurations at AM/NS Calvert #3HDGL — ASIS is a camera-based vision system enabling the on-line detection, localization and classification of surface defects while the strip is running. A picture of the ASIS installation at Calvert #3HDGL is shown in Fig. 1. The system was installed in the end of the coating line prior to the sidetrimmer. It consists of bright field and dark field of views for each side. Each field of view has two 4K line scan cameras. The maximal speed of the line is 260 m/minute. This results in the image resolution of 0.25 mm x 0.5 mm/pixel.

Classifier Tuning Method — A general method of an ASIS tuning and performance optimization is shown in Fig. 2.[6] It includes defect detection tuning and classification tuning tasks. Both tasks require specifying the business objectives of system use. The business objectives of system use guide how the system should be configured and tuned along with the process information. After the system is initially tuned, the system performance can be evaluated through coil inspection maps and user feedback. The information is then fed into the system, fine-tuning until the system performance reaches an optimal status. The tuning chain is a closed loop, which normally requires two to three circles. In order to reach an optimal system performance, two types of knowledge are required, one of which is quality and process knowledge and the other of which is ASIS tuning knowledge.

Calvert #3HDGL has been producing galvanized (GI) and

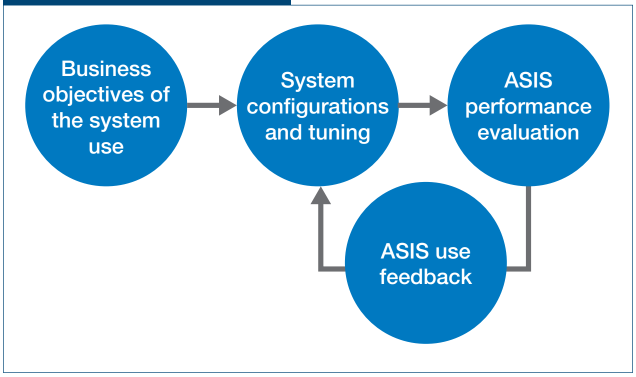
Figure 1



Schematic drawing of hot-dip galvanizing line (HDGL) process (a)[5] and automated surface inspection systems (ASIS) bottom side installation at AM/NS Calvert #3HDGL (b).

galvannealed (GA) materials for automotive and some of industrial non-automotive products. The main use of the system is the assistance tool of quality assurance related to surface defects and process troubleshooting. During the tuning process, the authors were able to consult finishing line quality experts and the line inspectors, collect the requirements of the system use, get daily coil rejection reports, draft a production list with process parameters, receive their feedback on the system use and so on. This has been one of the critical tasks to tune the system toward an optimal performance.

Figure 2

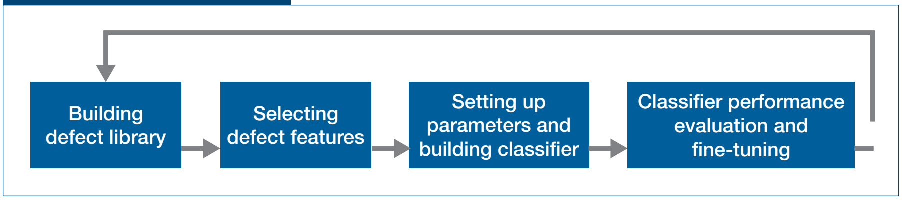


Flow chart of a typical ASIS tuning.

Classifier Tuning Objective and Procedure — The classification performance for a certain defect is evaluated by two parameters: accuracy rate and confidence rate. The accuracy rate defines the percentage of the defect correctly classified by the classifier. The confidence rate defines the percentage of true sample numbers in the classified samples for a certain defect.[7] To meet the business needs of a system use, the classification tuning objective is to tune a classifier to be able to more accurately classify all the real gross and important defects (with higher accuracy rate) and reduce false positive ones (as high classification confidence rate as possible). Confusions between gross and real defects, including important and less important ones, are more accepted than incorrectly classifying an important defect.

The classifier tuning procedure includes building a defect library, selecting defect features, setting up the decision tree parameters and creating a classifier model, and evaluating the classifier performance and classifier fine-tuning (shown in Fig. 3). Based on the classifier test results and trying to meet the classifier tuning objective, further improvement of the classifier can be done by enriching defect library, selecting optimal feature set and/or setting up different classifier

Figure 3



A flow diagram of classifier tuning procedure.

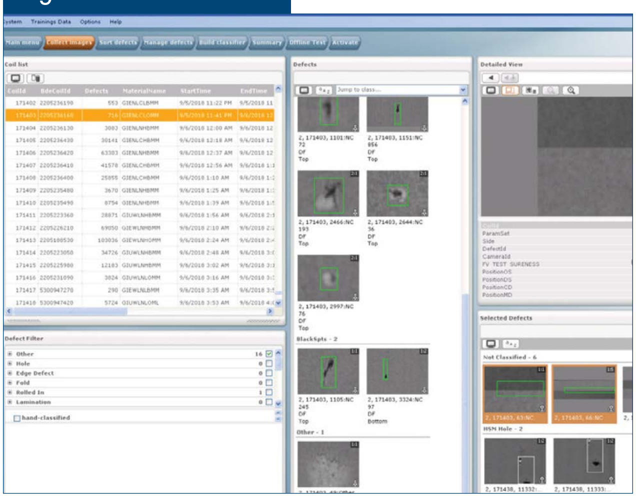
parameters. The whole procedure will be repeated several times until a satisfactory classifier result is reached.

Building Defect Library — One of the most important steps of creating a good classifier is to build a defect library, which can be representative to both important surface defects and pseudo defects in the normal production environment. When building the defect library, the best practices have been established based on the previous study[7] and experiences. These best practices are:

1. Group similar defects into one class.
2. Try to limit the total class categories (10 to 15 is better, but 20 to around 30 classes are good).
3. Get roughly equal sample numbers for each defect class, and the ratio between the largest to the smallest group/class should not exceed 5 to 1.
4. Try to make the defects in each class have similar appearance variety and avoid collecting the same defect images in one class; also avoid collecting the same or similar defect images from one class in two different classes.
5. Do not try to group defect classes based on defect severity.
6. Do not collect defect images in which feature values are not calculated.
7. Do not collect defects that were classified by pre-rule classifiers.
8. Do not collect many defect images from an unusual process into the defect library.

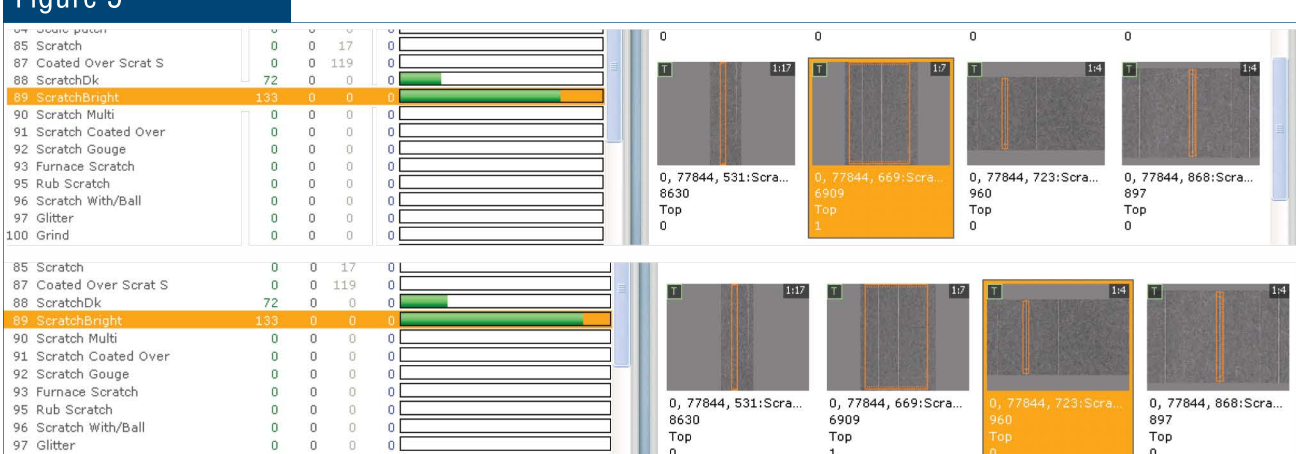
The ISRA Parsytec system includes software for users to easily complete the whole classifier tuning procedure (shown in Fig. 3). It has multiple functions to ease the classifier creation process and a user interface for fast image collection on multiple coils for certain defects; it also provided the similarity function to aid in the certain image collection and a user-friendly interface for users to quickly sort defect images. In addition, the classifier parameters can be easily accessed and set up through this software. One example of the GUI for the image collection on multiple coils is shown in Fig. 4. An example of defect image sorting best practice is shown in Fig. 5, where the images with the single scratch and multiple scratches are put in one class since they have similar similarity value.

Figure 4



Graphical user interface (GUI) of image collection from ISRA Parsytec software.

Figure 5



An example of defect image sorting best practice.

Table 1

Statistics for the Four Defect Libraries

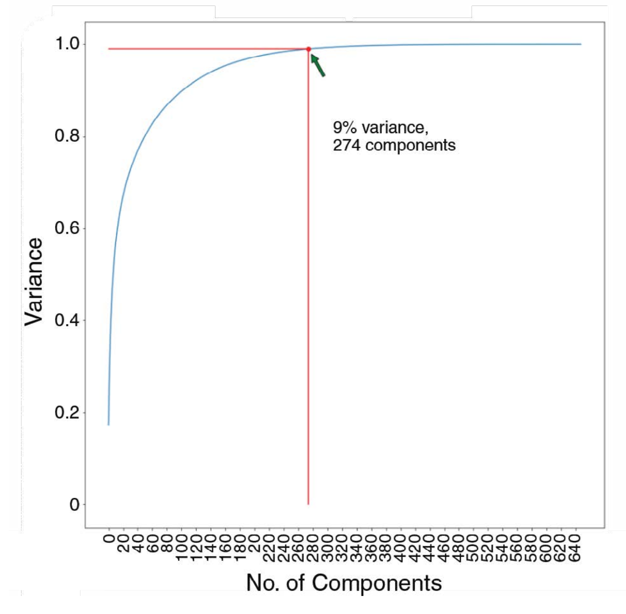
Defect information \ Material group	GI-BF	GI-DF	GA-BF	GA-DF
Total defect number	4,076	3,957	3,979	4,154
Total defect classes	32	30	33	32

Based on the Calvert system's configurations and system setup, four defect libraries were built: GI-BF, GI dark field (GI-DF), GA bright field (GA-BF) and GA dark field (GA-DF). The statistics of the defect libraries are shown in Table 1.

Selecting Defect Features — The second step of building a classifier is to select defect features. The ISRA Parsytec system provides more than 800 defect features. An interesting question is if the high number of features could cause the classifier overfitting, which presents a poor generalization ability to the unforeseen/new defect image data. This study aimed to verify if the feature number for the decision tree classifier needed to be reduced. Feature reduction includes two components: feature selection and feature extraction. The feature selection tries to find a subset of the original set of variables, or features, to get a smaller subset, which can be used to model the problem. It involves three ways, which are filter, wrapper or embedded.[8] The feature extraction reduces the data in a high dimensional space to a lower dimensional space.[8] The methods of dimensionality reduction include principle component analysis (PCA), linear discriminant analysis (LDA) and generalized discriminant analysis (GDA).[8] This section shows a study of PCA-based feature reduction algorithm and its effect to the final classifier performance.

PCA is a projection-based method that transforms the data by projecting it onto a set of orthogonal axes.[9] It finds the best linear combinations of the original variables so that the variance or spread along the new variable is optimized.[9] For this study, the original refined GI-BF defect library was used. The defect features for this defect library were extracted using ISRA Parsytec ASIS software. These features are fed into the PCA function in the Python development environment. The best 274 components/features with equal to and over 99% variance were selected (Fig. 6) for the classifier performance test. To compare the effectiveness of the top 274 features with the original classifier, the classifier using the 274 features was created using the same GI-BF defect library. The two classifier test results using new inspection image data (total 914 defects from four semi-exposed GI-BF coils) are shown separately in Tables 2 and 3. It was also noticed that the classifier training time using more than 800 features takes about 5 times longer than the one using the top 274 features. Based on the results, it was concluded that the classifier using the PCA-based feature reduction algorithm can bring similar performance to the one using the full set of

Figure 6



Feature variance chart using principle component analysis (PCA) algorithm.

Table 2

Classifier Result Using All the Features

Predicted class \ Assigned class	Gross defect	Pseudo defect	Total	Confidence rate
Gross	43	—	43	100%
Pseudo	—	772	772	100%
Non-classified defects	2	97	99	—
Total defect number	45	869	914	—
Accuracy rate	95.6	88.9%	—	—

Table 3

Classifier Result Using the Top 274 Features

Predicted class \ Assigned class	Gross defect	Pseudo defect	Total	Confidence rate
Gross	42	2	44	95.5%
Pseudo	—	771	771	100%
Non-classified defects	3	96	99	—
Total defect number	45	869	914	—
Accuracy rate	93.3%	88.7%	—	—

features. It could also be estimated that the classifier may take less time to classify new defects. This would be extremely useful if the system was configured to have a higher sensitivity of detection parameters and process a large amount of data in real time.

Setting Up Classifier Parameters and Building Classifier — The third step of classifier tuning is to configure the classifier parameters and create a classifier model. The ISRA Parsytec system has provided a decision tree classifier algorithm, which appears to have used the pruning technique and boosting algorithm for good classifier generalization ability. The boosting parameter is the total number of concurrent classifiers. The default number is 10. The user is able to put in a maximum value of 12. If a higher number is chosen, it takes more processing time with potential better classifier generalization ability. In this case, there were 12. There are two pruning parameters: pre-pruning and post-pruning. The default values were used for these two parameters. The "vote" parameter appears to be applied during the single decision tree classifier process. The smaller number of the vote parameter can reduce unclassified defects. The "confidence" parameter appears to be applied among different concurrent classifiers. Reducing this value can also bring fewer unclassified defects. During the initial classifier tuning stage when there are not enough defect samples, these two parameters are set up as the default value of 0.6. After the defect library is fully developed, both parameters can be reduced to 0.35 so as to reduce total unclassified defect number.

After the classifier parameters are set up, the classifier can be automatically trained. But the first classifier is not normally the one meant to be put on-line and further fine-tuning of the classifier is needed.

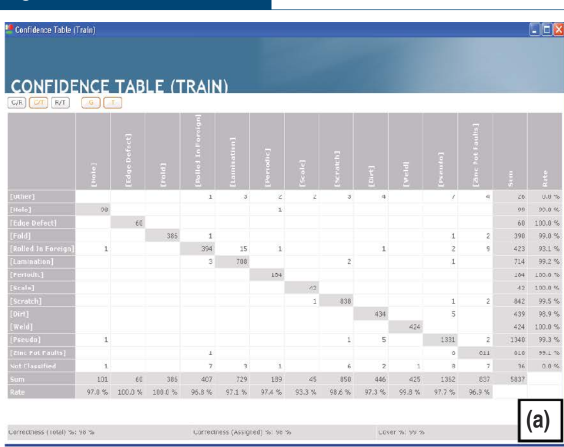
Classification Performance Evaluation and Classifier Fine-Tuning — Classifier tuning is a repeated process (shown in Fig. 3). After the first classifier is created, it can be tested against the training defect library. The confusion matrix of the classifier performance can then be used to guide the defect library refinement. An example of the confusion matrix is shown in Fig. 7a, where C/T means the assigned classes (classified by the experts) against classes classified by the classifier. Using the first classifier results, the uncertain defects can be checked between the gross defects and important defects and the defects grouped with similar appearance in one class; one can check if there are duplicated defect samples in one class and if less representative defects are in the defect class by looking at the appearance of unclassified defects. The verification process is intended to further improve defect library quality for an improved classifier. In this way, it takes a few circles for a classifier to be finely tuned and used in the on-line system.

After the off-line tuning of a classifier and its defect library, if the classifier results are satisfied, the classifier can be promoted to the on-line system. It may take a couple more circles to further improve the on-line classifier performance by collecting unclassified and falsely classified important defects and re-sorting them into the defect library. After this procedure, the finely tuned classifier for GI-BF field is shown in Fig. 7b. The classifier results on unseen image data are shown in Fig. 8.

Conclusions and Discussion

This paper presented a decision tree–based classifier tuning method and procedures, which includes the

Figure 7



Assigned Class / Predicted Class	Hole	Lamn	Rolled In Foreign	Dent	Dross	Baffle buildup	WeldHole	Others	Total	Confidence
Hole	76								76	100%
Lamn		134							134	100%
Rolled In Foreign			237						237	100%
Dent				131					131	100%
Dross					274			3	277	98.9%
Baffle buildup						198		1	199	99.5%
WeldHole							180		180	100%
Others		2	4		11	2		2816	2835	99.3%
NC					2			5	7	
Total	76	136	241	131	287	200	180	2825	4076	
Accuracy	100%	98.5%	98.3%	100 %	95.5%	99%	100%	99.7%		

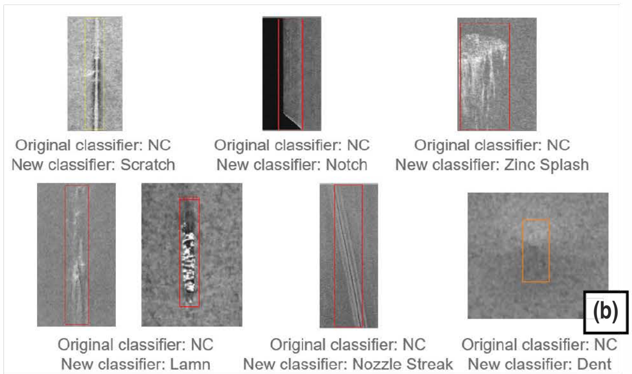
(b)

Classifier performance: confusion matrix for the first classifier (a) and confusion matrix for the finely tuned classifier (b).

Figure 8

Predicted Class \ Assigned Class	Hole	Lamn	Rolled In Foreign	Dent	Baffle buildup	WeldHole	Others	Total	Confidence
Hole	22							22	100%
Lamn		1						1	100%
Rolled In Foreign			20					20	100%
Dent				14				14	100%
Baffle buildup					6			20	100%
WeldHole						22		22	100%
Others							736	736	100%
NC	2						91	93	
Total	24	1	20	14	6	22	827	914	
Accuracy	91.8%	100%	100%	100 %	100%	100%	89.5%		

(a)



Classifier results on unseen image data (four GI-BF semi-exposed coils in the normal production environment): classifier performance (a) and examples of classifier results (b).

defect library building best practices, feature selection, classifier parameter setup, and classifier performance evaluation and its fine-tuning. Using the GI-BF classifier building process at Calvert #3HDGL as an example, it highlighted the defect library building best practices. In the feature selection procedure, it studied the PCA-based feature reduction method and its effect to the final classifier performance. It can be concluded that the reduced feature sets (274 features) can achieve similar classifier performance with reduced classifier training speed and potentially reduced on-line classifier execution speed than the ones using the whole feature set.

While it is tedious and takes time to build a representative defect library, future research topics in this area could be developing methods and algorithms for creating fake representative defect images, auto defect class sorting, and higher generalization ability of a classifier.

Acknowledgments

The authors would like to acknowledge AM/NS Calvert's quality and operations departments for their support. The authors would also like to thank ArcelorMittal Global R&D – East Chicago management for their great support and encouragement in this project. Without their help and efforts, the authors would not have been able to complete this work.

References

1. M. Bailleul, "Dynamic Surface Inspection at the Hot Strip Mill," *AISTech 2005 Conference Proceedings,* 2005.
2. T. MacDougall and E. Dillon, "Implementation of Surface Inspection Technology on Noisy Surfaces," *AISTech 2007 Conference Proceedings,* 2007.
3. M. Gonzalez, A. Trevino, M. Bailleul and F. Nlome-Nze, "Automatic Surface Inspection for Galvanized Products at IMSA MEX," *AISTech 2006 Conference Proceedings,* 2006.
4. L. Zhang, H. Long and J. Chapko, "Tuning and Deployment of a Surface Inspection System," *Iron & Steel Technology,* Vol. 7, No. 1, 2010, pp. 38–46.
5. "Hot-Dip Galvanized Steel Sheets," https://www.jfe-steel.co.jp/en/products/sheets/c30.html.
6. L. Zhang, "Automated Surface Inspection System Deployment in the Flat Carbon Steel Process," *Surface Inspection Summit 2015,* Pittsburgh, Pa., USA, 2015.
7. Z. Chen, J. Li, R.J. Pavlosky, L. Galey and L. Zhang, "Performance Optimization of Automated Surface Inspection System," *Steel Times International,* August 2019.
8. "Introduction to Dimensionality Reduction," A Computer Science Portal for Greeks, https://www.geeksforgeeks.org/dimensionality-reduction.
9. A.K. Pal, "Understanding Dimension Reduction With Principal Component Analysis (PCA)," 11 December 2017, https://blog.paperspace.com/dimension-reduction-with-principal-component-analysis. ✦

This paper was published in the AISTech 2020 Conference Proceedings. AIST members can access the AISTech 2020 Conference Proceedings in the AIST Digital Library at digital.library.aist.org.