

Streaming Machine Learning and Online Active Learning for Automated Visual Inspection.

Jože M. Rožanec*
Jožef Stefan International
Postgraduate School
Ljubljana, Slovenia
joze.rozanec@ijs.si

Elena Trajkova
University of Ljubljana, Faculty of
Electrical Engineering
Ljubljana, Slovenia
trajkova.elena.00@gmail.com

Paulien Dam
Philips Consumer Lifestyle BV
Drachten, The Netherlands
paulien.dam@philips.com

Blaž Fortuna
Qlector d.o.o.
Ljubljana, Slovenia
blaz.fortuna@qlector.com

Dunja Mladenić
Jožef Stefan Institute
Ljubljana, Slovenia
dunja.mladenic@ijs.si

ABSTRACT

Quality control is a key activity performed by manufacturing companies to verify product conformance to the requirements and specifications. Standardized quality control ensures that all the products are evaluated under the same criteria. The decreased cost of sensors and connectivity enabled an increasing digitalization of manufacturing and provided greater data availability. Such data availability has spurred the development of artificial intelligence models, which allow higher degrees of automation and reduced bias when inspecting the products. Furthermore, the increased speed of inspection reduces overall costs and time required for defect inspection. In this research, we compare five streaming machine learning algorithms applied to visual defect inspection with real-world data provided by *Philips Consumer Lifestyle BV*. Furthermore, we compare them in a streaming active learning context, which reduces the data labeling effort in a real-world context. Our results show that active learning reduces the data labeling effort by almost 15% on average for the worst case, while keeping an acceptable classification performance. The use of machine learning models for automated visual inspection are expected to speed up the quality inspection up to 40%.

CCS CONCEPTS

• Information systems → Data mining; • Computing methodologies → Computer vision problems; • Applied computing;

KEYWORDS

Intelligent manufacturing systems; Artificial intelligence; Machine learning; Quality assurance and maintenance; Fault detection; Intelligent manufacturing; Human centred automation

ACM Reference Format:

Jože M. Rožanec, Elena Trajkova, Paulien Dam, Blaž Fortuna, and Dunja Mladenić. 2021. Streaming Machine Learning and Online Active Learning for Automated Visual Inspection.. In . ACM, New York, NY, USA, 6 pages.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

IMS '22, March, 2022, Tel Aviv, Israel

© 2021 Copyright held by the owner/author(s).

1 INTRODUCTION

Quality control is a critical activity performed in manufacturing, which ensures products conform to specific requirements and specifications ([25, 38, 39]). Quality can be considered a competitive advantage and a business strategy, reducing rework in the production process, avoiding disruptions in the supply chain, and increasing consumer satisfaction ([3, 7]). Early defects detection enables tracing their root causes, take actions to mitigate them, enhancing the manufacturing process, and preventing the further manufacturing of defective parts.

While in many cases, quality inspection is visually performed by inspectors, such an approach has several drawbacks. First, the resources devoted towards training inspectors can grow proportionally to the scale of production. Second, the inspectors can work for a limited amount of time, and the quality of their work can be affected by the job design, environment, their experience, well-being, and motivation ([9]). Third, due to the factors mentioned above, the quality of inspection can vary over time for a single inspector ([24]). Finally, the quality of inspection can be affected by the inspector-to-inspector consistency.

The decreased cost of sensors and connectivity has enabled an increasing digitalization of manufacturing ([3]). The increasing democratization and adoption of Artificial Intelligence (AI) have been leveraged to enhance defects detection in manufacturing. Artificial Intelligence models provide virtually unlimited scalability capabilities, amend the drawbacks associated with the manual inspection ([8, 11, 12]), and have been successfully applied to quality inspection in multiple scenarios ([2, 3, 31, 32]).

To realize supervised machine learning models for defect detection, data must be collected and annotated. While a first dataset is used to train a quality inspection model, increasing the number of labeled samples can enhance the performance of such a model. It is important to consider how informative the new samples are to the existing model. This principle is taken into account by active learning, a subfield of machine learning, which aims to identify which instances are most informative to the model and ask an *oracle* to annotate them ([36]). The *oracle* can be a human annotator. The active learning strategy discards the non-informative samples.

We frame the defect detection problem as a supervised learning problem. In an optimal scenario, we consider that a production line would run without stopping (except for planned maintenance).



Figure 1: Samples of each class in the dataset: good, double print, and interrupted print.

Thus, the machine learning model should not require frequent deployments while delivering an acceptable defect discrimination performance. In addition, it is desired that manual labeling is minimized, reducing the manual effort and saving on data storage resources. We thus adopt a streaming active learning strategy and streaming machine learning algorithms. The streaming active learning strategy assumes a continuous stream of data, where for each instance, a decision must be made either to request its labeling or not. On the other hand, the streaming algorithm can issue a prediction on each new data instance and learn incrementally from those that are labeled by the oracle.

The main contributions of this research are (i) a comparative study between the five streaming machine learning algorithms for automated defect detection and (ii) two scenarios: no active learning (NO AL) and active learning (AL). We developed the machine learning models with images provided by the *Philips Consumer Lifestyle BV* corporation. The dataset comprises shaver images classified into three categories, based on the defects observed regarding the logo printing: good, double print, and interrupted print (see Fig. 1).

We evaluate the models using the area under the receiver operating characteristic curve (AUC ROC, see [5]), a widely adopted classification metric. The metric is invariant to *a priori* class probabilities, which makes it an excellent choice to measure classifiers performance on imbalanced datasets.

This paper is organized as follows. Section 2 describes related works, Section 3 describes the *Philips Consumer Lifestyle BV* use case, Section 4 provides a brief description of the methodology and experiments we conducted, while Section 5 outlines the results obtained. Finally, Section 6 concludes and describes future work.

2 RELATED WORK

Automated visual inspection refers to image processing techniques used to perform quality control ([2]). Automated visual inspection techniques can be divided into two groups (supervised and unsupervised methods), depending on whether labeled or unlabeled data is required. Supervised machine learning methods require a dataset of labeled images, from which features are extracted to train the machine learning model later. Current state-of-the-art image processing techniques frequently make use of deep learning models, whether it is for end-to-end learning, or as feature extractors on top of which other models can be developed ([13, 28, 33]), though manually engineered features are used too ([34]).

The adoption of automated visual inspection for defect detection is being fostered in the context of Industry 4.0 ([6]). While it was applied in several use cases in the past, it is considered that the

field is still in its early stages and that artificial intelligence can revolutionize product inspection ([1]). Machine learning for visual inspection has been applied to control the quality of printing ([37]), inspect printed circuit board production ([10]), determine whether bottles met the required specifications ([35]), detect defects during metallic powder bed fusion in additive manufacturing ([14]), manufactured vehicle parts ([27]), aerospace components ([2]), or textile quality control ([21]). While the aforementioned authors report using machine learning algorithms, such as the Support Vector Machines, k-Nearest Neighbours, or Multilayer Perceptrons, in a batch setting, [29] describes a visual inspection use case where evolving fuzzy classifiers are used. In addition, an online (streaming) active learning setting is used to reduce manual labeling work.

Active learning (AL) is a sub-field of machine learning that assumes that many unlabeled data exists while the cost of labeling those data instances is expensive. Thus, it explores different techniques to find and select the most informative data instances, which are then presented to an *oracle* (e.g., a human annotator), who can label them. Online active learning assumes a stream of unlabeled instances exists, where one data instance is drawn at a time, and a decision must be made whether to label or discard such piece of data ([36]). One such criterion is uncertainty sampling, where the sample is selected based on the measured uncertainty according to a particular metric or machine-learning model ([26]). While active learning was applied in the manufacturing domain, not much research has been performed in this direction ([30]).

Streaming machine learning is a sub-field of machine learning that studies how a learner can provide adequate forecasts while learning from a sequence of data instances, learning one data instance at a time ([18]). Such models can be immediately updated and thus provide high scalability when dealing with large amounts of data arriving at a high velocity ([15]). In this work we compare five streaming classification algorithms: Hoeffding Tree (HT) ([20]), Hoeffding Adaptive Tree (HAT) ([4]), Stochastic Gradient Tree (SGT) ([16]), streaming Logistic Regression (SLGR), and streaming k-Nearest Neighbors (SKNN).

Combining online active learning and streaming machine learning classifiers allows to reduce downtimes required for new model deployments, updating the classification model one instance a time, and enabling an *inspection by exception*, only requesting the labeling of images for which ones the model is most uncertain about ([8]).

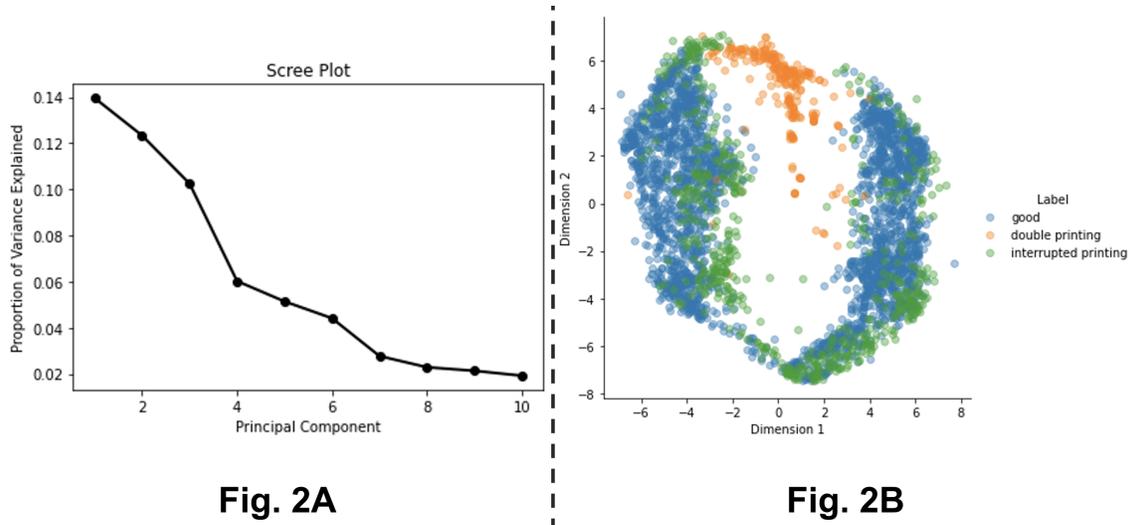


Fig. 2A

Fig. 2B

Figure 2: Fig. 2A shows a scree plot used to assess dimensionality reduction before creating the t-SNE plot. While for the t-SNE plot, we use ten variables, these could be further reduced to eight. Fig. 2B shows a t-SNE plot, with clusters of data points that correspond to the three target labels.

3 USE CASE

In this research, we focus on the visual inspection task performed on shavers produced by *Philips Consumer Lifestyle BV*. Such inspection aims to determine if the *Philips Consumer Lifestyle BV* company logo was properly printed and eventually detect the kind of defective printing. To that end, the company uses different pad-printing setups. A vast amount of products are produced every day, each of which is manually handled and inspected. If a defective print is found, the product is removed from the production line. It is expected that the use of an automated visual quality inspection system would strongly reduce manual work, speeding up the process by more than 40%. A labeled dataset of 3,518 images was provided for the task at hand to train and test the models. Two kinds of defects were labeled (double printing and interrupted printing), in addition to good prints. We treat the problem as a multiclass classification task. The images distribution can be visualized in the t-SNE plot at Fig. 2. A scree plot was used to assess the best amount of variables to consider when performing the dimensionality reduction.

4 METHODOLOGY

We framed the automated defect detection as a multiclass imbalanced classification problem. We make use of a pre-trained ResNet-18 model ([17]) as a feature extractor to obtain image embeddings. In particular, we extract features from the Average Pooling Layer, which translates each image into a feature vector of 512 values. To prevent overfitting, we perform feature selection selecting the *top K* features ranked by mutual information ([22]), considering $K = \sqrt{N}$, where N is the number of data instances in the train set ([19]).

To compare the performance of different streaming machine learning algorithms and assess the benefits of a streaming active learning approach, we divided the dataset using a stratified ten-fold cross-validation ([23, 40]), using one fold for testing, and divided the remaining data into the training set, and a stream of data for

the active learning setting, simulating unlabeled samples that could be labeled on request. We always queried the unlabeled instances until exhausting the stream. When doing so, we asked for labels of those unlabeled instances with a classification uncertainty above 0.55. Then, we measured the models' performance over the test set for each new instance obtained from the unlabeled data stream. To that end, we computed the AUC ROC metric with the "one-vs-rest" heuristic method. The "one-vs-rest" heuristic splits the multiclass dataset into multiple binary classification problems, computing the AUC ROC metric for each class and summarizing the final value as the weighted average considering the number of true instances for each class. We repeated the procedure mentioned above ten times to estimate better the performance of the models and learning scenarios (no active learning, and with active learning).

5 RESULTS AND ANALYSIS

We compared two scenarios by running five streaming machine learning classification algorithms across them. The scenarios of interest were: streaming without active learning (NO AL) and streaming active learning labeling only data instances with an uncertainty higher than 0.55 (AL). We summarize the results obtained for both scenarios in Table. 1. We found that the best mean discrimination was achieved without using active learning (NO AL): active learning degraded the models' discrimination performance. In both scenarios, we observed that the models' mean performance improved over time.

We present the models' performance across both scenarios in Table 2. The results show that the SKNN model achieved the best discrimination performance in both cases. The second-best results were reported for the SLGR. The worst performance was achieved by the HF and HAT models, which had a higher false-positive rate than the true-positive rate in the AL scenario.

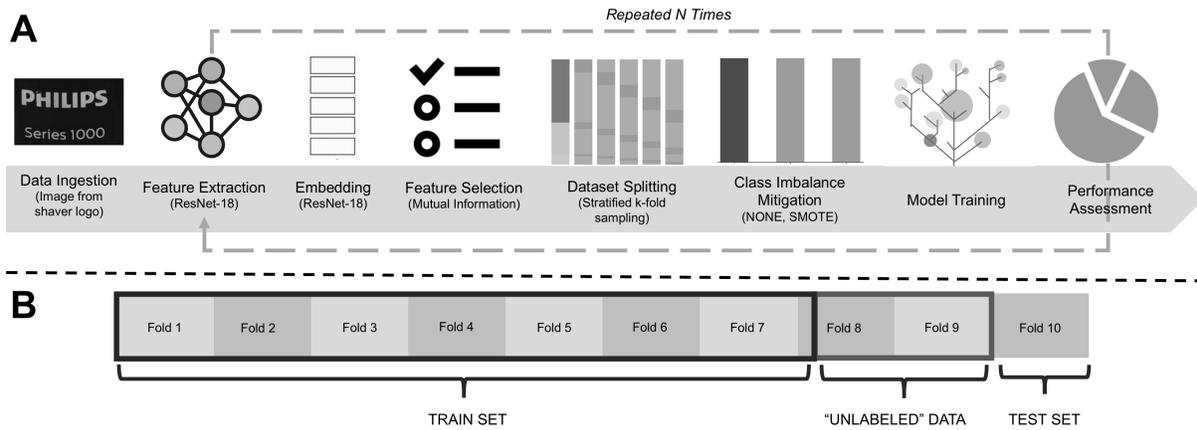


Figure 3: Fig. 3A shows the methodology we followed to develop and assess the models. In Fig. 3B we depict how data was partitioned into train, test, and a stream of instances used to simulate unlabeled data and their labeling in the active learning scenario.

SCENARIO	AUC ROC _{MEAN+CI}	AUC ROC _{MEAN+CI Q1}	AUC ROC _{MEAN+CI Q4}
NO AL	0,6350±0,0007	0,6085±0,0013	0,6532±0,0015
AL	<i>0,6025±0,0007</i>	<i>0,5903±0,0012</i>	<i>0,6103±0,0015</i>

Table 1: The table describes average AUC ROC values, and confidence intervals (CI) at a confidence level of 95%, obtained across ten times repeated ten-fold cross-validation for two scenarios. Best results are bolded, second-best results are highlighted in italics. Q1 and Q4 denote the first and last quartiles, respectively.

MODEL	NO AL	AL
SLGR	<i>0,7655±0,0010</i>	<i>0,6920±0,0012</i>
SKNN	0,8266±0,0005	0,8186±0,0004
SGD	0,5194±0,0005	0,5144±0,0004
HT	0,5164±0,0005	0,4972±0,0004
HAT	0,5470±0,0010	0,4901±0,0005

Table 2: The table presents mean AUC ROC values and confidence intervals (CI) at a confidence level of 95%, obtained for five streaming machine learning algorithms across two scenarios when performing a ten times repeated ten-fold cross-validation. Best results are bolded, second-best results are highlighted in italics.

MODEL	MIN	MAX	MEAN	STD DEV	Q1	Q2	Q3
SLGR	13,21%	100,00%	35,11%	32,50%	17,76%	19,89%	21,70%
SKNN	0,71%	<i>68,61%</i>	14,72%	25,46%	1,70%	2,13%	3,08%
SGD	0,00%	100,00%	20,00%	40,00%	0,00%	0,00%	0,00%
HT	0,00%	100,00%	39,42%	44,45%	0,00%	0,00%	86,79%
HAT	17,76%	100,00%	75,12%	25,40%	52,45%	92,04%	96,98%

Table 3: The table quantifies how much labeling effort is saved in a streaming active learning setting, considering models developed for different streaming machine learning algorithms. Only data instances with a classifier uncertainty higher than 0.55 were presented to the user. Measurements are taken for ten-fold cross-validation, repeated ten times. Best results are bolded, second-best results are highlighted in italics. Q1, Q2, and Q3 denote quartiles one to three, respectively.

To understand the contribution of the active learning approach regarding manual labeling, we measured how much effort was saved (effort gain) as the percentage of all non-labeled samples that were not presented to the *oracle* (human annotator) to request a label. We present the results in Table 3. We found that all the algorithms reported some effort gain on average. The HAT reported the highest gains, which skipped a median of 92% of the unlabeled samples, and reported effort gains of almost 97% at the third quartile. We consider the second-best algorithm, in terms of effort gain, was the SLGR, which did not achieve higher mean savings than the HT, but saved some labeling effort in almost every case, which is reflected by the values observed at all quartiles. We consider the best overall performance (when measuring effort gain and discriminative power assessed through the AUC ROC), was obtained by the SLGR algorithm, which had the second best performance in terms of discriminative power, and competitive results in terms of effort gain.

6 CONCLUSION

This paper compared five streaming machine learning algorithms across two scenarios: streaming without active learning (NO AL) and streaming with a streaming active learning setting (AL). Our results show that all algorithms improve their performance over time regardless of active learning is used. We consider the best overall performance was achieved by the SLGR algorithm, which achieved the second-best scores regarding discriminative power while also reporting competitive effort gains in terms of manual labeling effort (reduced 35% on average). While the SKNN algorithm obtained better results in terms of discriminative power, it reported low gains in manual labeling effort (nearly 15% on average). We thus conclude that active learning can provide effective means towards reducing the manual labeling effort. However, special attention must be put to the model to understand the relationship between models' learning, discriminative performance, and effective manual labeling effort gains. Regardless of the active learning setting, experts estimate that the deployment of machine learning defect detection models can speed up visual inspection up to 40% for this particular use case. Future work will study different approaches to tackle class imbalance in a streaming setting and further improve the classification performance. In addition, we will explore methods to provide cues to the human annotator, showing whether an error is expected and where it could be located. We expect such a piece of information will help to speed up the annotation times. Furthermore, we will develop explainable artificial intelligence methods to understand which segments of the image are used by the classifier, assess the quality of their learning, and identify opportunities for further model improvement.

ACKNOWLEDGMENTS

This work was supported by the Slovenian Research Agency and the European Union's Horizon 2020 program project STAR under grant agreement number H2020-956573. The authors acknowledge the valuable input and help of Jelle Keizer and Yvo van Vegten from Philips Consumer Lifestyle BV.

REFERENCES

- [1] Kareem S Aggour, Vipul K Gupta, Daniel Ruscitto, Leonardo Ajdelsztajn, Xiao Bian, Kristen H Brosnan, Natarajan Chennimalai Kumar, Voramon Dheeradhada, Timothy Hanlon, Naresh Iyer, et al. 2019. Artificial intelligence/machine learning in manufacturing and inspection: A GE perspective. *MRS Bulletin* 44, 7 (2019), 545–558.
- [2] Carlos Beltrán-González, Matteo Bustreo, and Alessio Del Bue. 2020. External and internal quality inspection of aerospace components. In *2020 IEEE 7th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*. IEEE, 351–355.
- [3] Tajeddine Benbarrad, Marouane Salhaoui, Soukaina Bakhat Kenitar, and Mounir Arioua. 2021. Intelligent machine vision model for defective product inspection based on machine learning. *Journal of Sensor and Actuator Networks* 10, 1 (2021), 7.
- [4] Albert Bifet and Ricard Gavaldà. 2009. Adaptive learning from evolving data streams. In *International Symposium on Intelligent Data Analysis*. Springer, 249–260.
- [5] Andrew P. Bradley. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30, 7 (1997), 1145–1159. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)
- [6] JA Carvajal Soto, F Tavakolizadeh, and Dávid Gyulai. 2019. An online machine learning framework for early detection of product failures in an Industry 4.0 context. *International Journal of Computer Integrated Manufacturing* 32, 4–5 (2019), 452–465.
- [7] Lai K Chan, Yer V Hui, Hing P Lo, K Tse Siu, Geoffrey KF Tso, and Ming L Wu. 2003. Consumer satisfaction index: new practice and findings. *European Journal of Marketing* (2003).
- [8] Amal Chouchene, Adriana Carvalho, Tânia M Lima, Fernando Charrua-Santos, Gerardo J Osório, and Walid Barhoumi. 2020. Artificial intelligence for product quality inspection toward smart industries: quality control of vehicle non-conformities. In *2020 9th international conference on industrial technology and management (ICITM)*. IEEE, 127–131.
- [9] Sarah-Jane Cullinane, Janine Bosak, Patrick C Flood, and Evangelia Demerouti. 2013. Job design under lean manufacturing and its impact on employee outcomes. *Organizational Psychology Review* 3, 1 (2013), 41–61.
- [10] Guifang Duan, Hongcui Wang, Zhenyu Liu, and Yen-Wei Chen. 2012. A machine learning-based framework for automatic visual inspection of microdrill bits in PCB production. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 6 (2012), 1679–1689.
- [11] Carlos A Escobar and Ruben Morales-Menendez. 2018. Machine learning techniques for quality control in high conformance manufacturing environment. *Advances in Mechanical Engineering* 10, 2 (2018), 168781401875519.
- [12] Colin Garvey. 2018. A framework for evaluating barriers to the democratization of artificial intelligence. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- [13] Tobias Glasmachers. 2017. Limits of end-to-end learning. In *Asian Conference on Machine Learning*. PMLR, 17–32.
- [14] Christian Gobert, Edward W Reutzel, Jan Petrich, Abdalla R Nassar, and Shashi Phoha. 2018. Application of supervised machine learning for defect detection during metallic powder bed fusion additive manufacturing using high resolution imaging. *Additive Manufacturing* 21 (2018), 517–528.
- [15] Heitor Murilo Gomes, Jesse Read, Albert Bifet, Jean Paul Barddal, and João Gama. 2019. Machine learning for streaming data: state of the art, challenges, and opportunities. *ACM SIGKDD Explorations Newsletter* 21, 2 (2019), 6–22.
- [16] Henry Gouk, Bernhard Pfahringer, and Eibe Frank. 2019. Stochastic gradient trees. In *Asian Conference on Machine Learning*. PMLR, 1094–1109.
- [17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [18] Steven CH Hoi, Doyen Sahoo, Jing Lu, and Peilin Zhao. 2018. Online learning: A comprehensive survey. *arXiv preprint arXiv:1802.02871* (2018).
- [19] Jianping Hua, Zixiang Xiong, James Lowey, Edward Suh, and Edward R Dougherty. 2005. Optimal number of features as a function of sample size for various classification rules. *Bioinformatics* 21, 8 (2005), 1509–1515.
- [20] Geoff Hulten, Laurie Spencer, and Pedro Domingos. 2001. Mining time-changing data streams. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. 97–106.
- [21] JL Jiang and WK Wong. 2018. Fundamentals of common computer vision techniques for textile quality control. In *Applications of Computer Vision in Fashion and Textiles*. Elsevier, 3–15.
- [22] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. 2004. Estimating mutual information. *Physical review E* 69, 6 (2004), 066138.
- [23] Max Kuhn, Kjell Johnson, et al. 2013. *Applied predictive modeling*. Vol. 26. Springer.
- [24] Agnieszka Kujawińska, Katarzyna Vogt, and Adam Hamrol. 2016. The role of human motivation in quality inspection of production processes. In *Advances in Ergonomics of Manufacturing: Managing the Enterprise of the Future*. Springer, 569–579.

- [25] Nani Kurniati, Ruey-Huei Yeh, and Jong-Jang Lin. 2015. Quality inspection and maintenance: the framework of interaction. *Procedia manufacturing* 4 (2015), 244–251.
- [26] David D Lewis and Jason Catlett. 1994. Heterogeneous uncertainty sampling for supervised learning. In *Machine learning proceedings 1994*. Elsevier, 148–156.
- [27] Wang Liqun, Wu Jiansheng, and Wu Dingjin. 2020. Research on vehicle parts defect detection based on deep learning. In *Journal of Physics: Conference Series*, Vol. 1437. IOP Publishing, 012004.
- [28] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3431–3440.
- [29] Edwin Lughofer. 2017. On-line active learning: A new paradigm to improve practical useability of data stream modeling methods. *Information Sciences* 415 (2017), 356–376.
- [30] Lingbin Meng, Brandon McWilliams, William Jarosinski, Hye-Yeong Park, Yeon-Gil Jung, Jehyun Lee, and Jing Zhang. 2020. Machine learning in additive manufacturing: A review. *Jom* 72, 6 (2020), 2363–2377.
- [31] Paolo Napoletano, Flavio Piccoli, and Raimondo Schettini. 2021. Semi-supervised Anomaly Detection for Visual Quality Inspection. *Expert Systems with Applications* (2021), 115275.
- [32] Josue Obregon, Jihoon Hong, and Jae-Yoon Jung. 2021. Rule-based explanations based on ensemble machine learning for detecting sink mark defects in the injection moulding process. *Journal of Manufacturing Systems* 60 (2021), 392–405.
- [33] Samira Pouyanfar, Saad Sadiq, Yilin Yan, Haiman Tian, Yudong Tao, Maria Presa Reyes, Mei-Ling Shyu, Shu-Ching Chen, and Sundaraja S Iyengar. 2018. A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)* 51, 5 (2018), 1–36.
- [34] S Ravikumar, KI Ramachandran, and V Sugumar. 2011. Machine learning approach for automated visual inspection of machine components. *Expert systems with applications* 38, 4 (2011), 3260–3266.
- [35] Santosh Kumar Sahoo, M Mahesh Sharma, and BB Choudhury. 2019. A Dynamic Bottle Inspection Structure. In *Computational Intelligence in Data Mining*. Springer, 873–884.
- [36] Burr Settles. 2009. Active learning literature survey. (2009).
- [37] Javier Villalba-Diez, Daniel Schmidt, Roman Gevers, Joaquín Ordieres-Meré, Martin Buchwitz, and Wanja Wellbrock. 2019. Deep learning for industrial computer vision quality control in the printing industry 4.0. *Sensors* 19, 18 (2019), 3987.
- [38] Thorsten Wuest, Christopher Irgens, and Klaus-Dieter Thoben. 2014. An approach to monitoring quality in manufacturing using supervised machine learning on product state data. *Journal of Intelligent Manufacturing* 25, 5 (2014), 1167–1180.
- [39] Jing Yang, Shaobo Li, Zheng Wang, Hao Dong, Jun Wang, and Shihao Tang. 2020. Using deep learning to detect defects in manufacturing: a comprehensive survey and current challenges. *Materials* 13, 24 (2020), 5755.
- [40] Xinchuan Zeng and Tony R Martinez. 2000. Distribution-balanced stratified cross-validation for accuracy estimation. *Journal of Experimental & Theoretical Artificial Intelligence* 12, 1 (2000), 1–12.