EuroVis Workshop on Visual Analytics (2022)
M. Angelini and J. Bernard (Editors)

Interactive Visual Explanation of Incremental Data Labeling

R. Beckmann[1,2], C. Blaga[1,2], M. El-Assady[2], M. Zeppelzauer[3] and J. Bernard[1,4]

[1]University of Zürich, Switzerland [2]Swiss Federal Institute of Technology Zürich, Switzerland
[3]St. Pölten University of Applied Sciences, Austria [4]Digital Society Initiative, Zürich, Switzerland

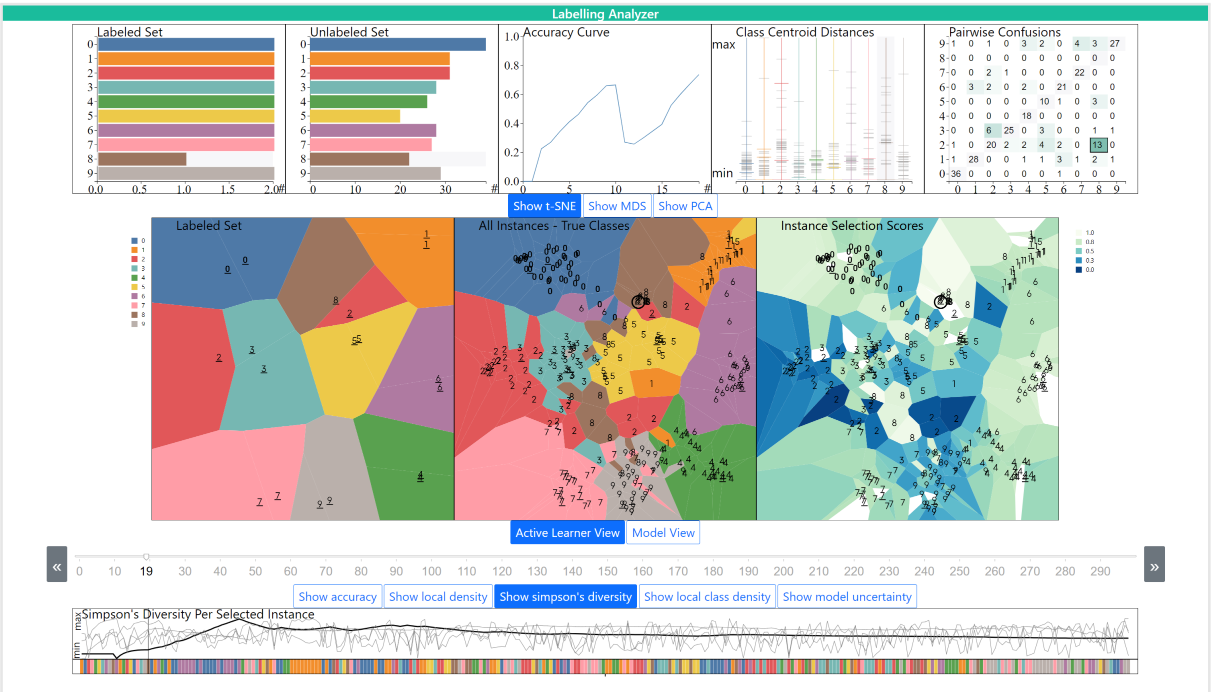

Figure 1: *Visual observation of an instance selection strategy in a data labeling process. Different perspectives on the data set reveal item and model characteristics (here: iteration 19). The visualization of metrics along the process provides navigation support for further inspection.*

Abstract

We present a visual analytics approach for the in-depth analysis and explanation of incremental machine learning processes that are based on data labeling. Our approach offers multiple perspectives to explain the process, i.e., data characteristics, label distribution, class characteristics, and classifier characteristics. Additionally, we introduce metrics from which we derive novel aggregated analytic views that enable the analysis of the process over time. We demonstrate the capabilities of our approach in a case study and thereby demonstrate how our approach improves the transparency of the iterative learning process.

1. Introduction

Interactive machine learning (IML) is a complex analytical process that is difficult to understand and to make transparent [ACKK14]. A key component in IML that directly steers the incremental learning process is instance selection strategies (ISS). An ISS tries to select *the* instance(s) from an unlabeled pool of candidates that best facilitate(s) the learning of the IML model at the current stage of the learning process. A prominent type of ISS proposed in the ML community is active learning [Set12, FZL12]. Instance selection is an ill-posed problem and thus hard to solve. Existing strategies use heuristics, relying on either model-specific aspects (e.g., model uncertainty [WKBD06]) or data-specific aspects (e.g., density [VPS*]). Recent experiments have shown, however, that these strategies perform far beyond what would theoretically be possible with an "ideal" strategy [BHL*18] in terms of accuracy.

Different expert user groups are involved in research on ISS. First, ML engineers aim at both validating ISS implementations and revising ISS algorithms to improve the performance with respect to some measure of ISS quality [SC08, KCH*17, RXC*21]. Second, VIS researchers are observing users when selecting instances, following the goal to formalize human-based ISS that tend to complement or even outperform model-based ISS [SG10, BHZ*18, CBC*20]. Finally, ML and VIS researchers seek method-

DOI: 10.2312/eurova.20221073

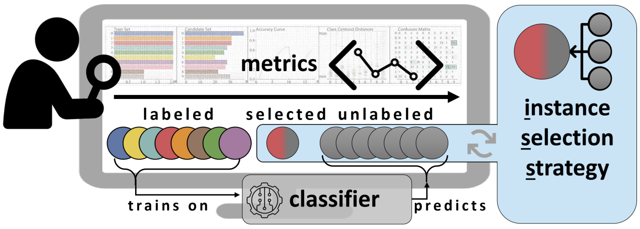

Figure 2: *Observational study of the labeling process, including (labeled and unlabeled) data, underlying classifier, and ISS.*

ologies and techniques to design ISS that will perform significantly better by making better decisions (instance selections) [BHL*18, BHS*21]. While individual aspects of the analysis scenarios may differ in their complexity, all expert groups are confronted with a similar problem: The functional behavior and internal working of the ISS, and/or the underlying classifier, and/or the characteristics of the dataset are unknown, and, thus, cannot be explained. Figure 2 provides an overview of this special analyis context.

We contribute a VA approach for the observation and explanation of ISS for IML along a labeling process. To conduct this novel type of study, our approach provides analysts the means to analyze entire precomputed incremental data labeling processes, performed by any ISS. Multiple linked views show individual labeling iterations from different perspectives, including dimensionality-reduced data representations, label distributions, classifier characteristics, class confusions, class distance relations, and ISS scores for every candidate instance. In addition, visual comparison techniques support the assessment of data characteristics of labeled data versus unlabeled data. To facilitate effective analysis, we visualize data- and model-driven metrics to derive visual cues, which serve as entry points for the exploration of the ISS process. We demonstrate the usefulness of our approach with the MNIST handwritten digits dataset [LBBH98] (without loss of generality) in an observational case study on quasi-optimal ISS assessment.

2. Related Work

The assessment of decisions of ISS relates to the spectrum of explainable AI approaches [AB18, ABC*19]. While such approaches are often categorized into data-driven [CDH*16] or model-driven explanations [ZYMW19], we explicitly incorporate both data and model characteristics to observe ISS behavior. One type of study is focusing on the decision of an ISS for a winning instance per iteration, from a possibly large set of unlabeled candidate instances [BHS*21]. This is typically the case for pool-based ISS [Set12] in active learning, or for ISS based on human decisions, observed in user studies [BHL*18, CBC*20]. However, to also account for the entire labeling process, users need a means to explain ISS behavior based on *all* instances selected in the process so far, i.e., the training data with respect to the currently observed iteration. Along these lines, we draw connections to the assessment of model learning processes, including the assessment of class confusions over time [HRS*22], works studying the involvement of users to steer model building during training [TKC17], as well as the identification of stable layers for in-depth investigation [PHV*18]. However, none of these approaches can be used to observe the behavior of ISS in a labeling process. In a pioneering experiment, five experts in data analysis went through a tedious observational study with five datasets, and reflected on their findings about principal behaviors of previously executed quasi-optimal instance labeling [BHL*18], an experimental design that was difficult to replicate so far. There are several VA approaches in the context of instance selection, some of them also providing multiple linked views [KPSK17, BHZ*18, SJS*21, EBJ*22]. However, none of these approaches can be used to observe and study ISS behavior, as opposed to just labeling the data interactively.

3. Research Questions

The analysis context at hand is special, as the dataset is always separated into disjoint sets of already labeled and unlabeled instances. Also, users either focus on a *single* winning instance from the pool of unlabeled instances or *all* instances that have already been selected to form the training data. Figure 2 sheds light on these special analysis challenges and their data and model-centered ingredients. The selected instance per iteration transitions from the unlabeled to the labeled set. The entire labeling process means an iterative swap from 100% unlabeled to 100% labeled data in the dataset. To examine the potential of visualization techniques for the explanation of decisions of possibly not well-understood ISS, we focus on the detailed analysis of individual iterations and the corresponding ISS decision. This can be broken down into four research questions. Can the instance selection of an ISS be explained by observing:

- R1: the given data characteristics?
- R2: the categorical label distribution of the (un)labeled data?
- R3: the data characteristics of the individual classes?
- R4: the predictions of the underlying classifier?

4. Visual Workspace for ISS Behavior Assessment

Assessment Metrics We define a set of metrics to further examine context information of the ISS and to better understand ISS decisions. This approach is inspired from the long list of endeavors making complex data and model characteristics measurable by expressing key aspects in numbers. Examples include Scagnostics measures [WAG05], the per-instance uncertainty of classifiers [RAL*17], or the separability of classes [SNLH09] and clusters [STMT12]. Here, metrics produce one value for each iteration in the labeling process, with respect to the selected instance, and present one possible metric for each research question.

- R1: Local density: Implemented with a k-Nearest Neighbor algorithm [Kra13]. It captures the local spatial density around the selected instance. High values imply high density.
- R2: Simpson's diversity measure: The Simpson's diversity measure [Sim49] was chosen to assess the label diversity in the labeled dataset. High values imply low diversity (balance).
- R3: Local class density: Same as R1, but only instances of the same class are considered as nearest neighbors. This metric measures how densely the neighborhood is populated with members from the same class. High values imply high density.
- R4: Model uncertainty: Denotes the probabilistic confidence that the underlying model assigns to the selected instance [YL16]. High values imply high uncertainty, i.e., low confidence.

General Visualization and Interaction Designs Two driving design principles were a) using standard chart types where possible, and b) re-using visual structures from inspiring and/or related approaches when appropriate. In addition, visual encodings are provided in a way that recurrence across views is leveraged whenever possible. One example is a categorical color coding, used when class information is depicted (six views). The class of the selected instance is highlighted in all class-granularity views, mostly by encoding the background area (three views). The solution for representing instances in projection plots is using a generic

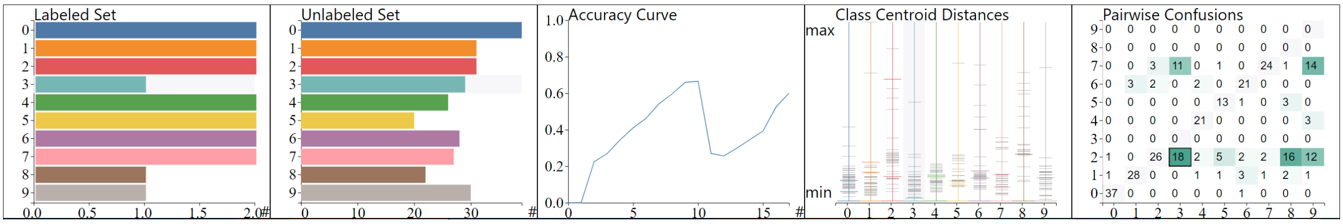

Figure 3: *Data and classifier metrics referring to iteration 17.*

point mark, or, if appropriate, a small icon/shape representing the semantics of the classes (here: digits 0-9, three views). Labeled instances are additionally underlined to be better distinguishable from unlabeled instances (three views). The selected instance in the current iteration has a circular outline (two views). When users hover over components of the class confusion view, the corresponding instances are highlighted across views (two views). Finally, the metrics along the time axis provide a visual navigation support to directly identify interesting iterations.

4.1. Observing the Given Data Characteristics (R1)

We provide three views to show the underlying data distribution. Building upon best practices in visualization research, we represent high-dimensional data instances in projection plots, in combination with different dimensionality reduction methods [BLBC12, SA15, HMdCM17], as shown in Figure 1 (center row). The choice of the dimensionality reduction method is a user parameter, to account for shortcomings of individual methods and to always allow different perspectives on the data. Inspired by the visualization of high-quality decision boundary maps [REHT19], our decision was in favor of coloring the background of the (2D) data space [BHS*21], as opposed to coloring individual instances [BHZ*18]. Here, we make use of a Voronoi tesselation of the 2D manifold, leading to non-regular area marks to be colored. The background color refers to the categorical class labels of the instances (Labeled Set view and All Instances view) or to the ISS scores per instance (Instance Selection Scores view).

To account for the two disjoint sets of instances, we always use the left data view (Labeled Set view) to only show labeled instances. Using Figure 1 as an example, analysts can always see how well current labeled instances represent the dataset to be labeled. The view at the center (All Instances view) shows all instances and can serve two purposes: a) Showing the ground truth information for all instances, and coloring all instances (Active Learner view), b) Showing the predictions of all unlabeled instances, and coloring the false predictions only (Model view). In Figure 1, the Active Learner view perspective is shown, whereas Figure 4 shows the Model view (left) and Active Learner view (right) next to each other. In Figure 1, the ground truth labels are shown in the All Instances view. The visual comparison of the Labeled Set view and the All Instances view thus helps to assess how well the instances in the labeled dataset can already capture the basic structure of the whole dataset. This can be visually assessed by comparing the coloring of the different regions of the Voronoi tesselations. The projection plot on the right (ISS view) encodes the scores of the chosen ISS for every instance. According to the classical active learning approach, the instance with the highest score is always selected to be labeled next. In general, the observed ISS can be an active learning algorithm [Set12, FZL12], a human-based ISS [SG10, BHZ*18, CBC*20], or a quasi-optimal ISS [BHL*18] as observed in the case study (Section 5). A continuous colormap assigns bright colors to instances with high ISS scores.

4.2. Observing the Label Distribution (R2)

The label distribution has an influence on label selection, as in many cases a well-balanced training set is preferable [Set12, BHL*18]. To always be aware of the label distribution, we provide two views, both shown in Figure 3 (left). We use a horizontal barchart to visualize the current label distribution. By using the shown example, the given ISS seems to consider label balance quite intensively: most of the classes have been selected twice, with only three classes (3, 8, 9) being selected only once yet. We use the same encoding also for the visualization of the unlabeled instances, as for pre-executed processes the ground truth can be leveraged as well. The resulting barchart helps to assess whether a given ISS considers the balance/imbalance of unlabeled data when selecting instances.

4.3. Observing Data Characteristics of Individual Classes (R3)

We support answering the question if a given ISS exploits characteristics of data instances, according to their distributions for individual classes. This puts to focus of analysis on the granularity of classes [STMT12, BHZ*21]. The two main design targets are a) the spatial arrangement of instances per class and b) the distances of instances to their class center. A second line of thought is whether the focus shall be on the already labeled instances or not. With the Labeled Set view, we support the analysis of the spatial arrangement of instances per class only for labeled instances, whereas the All Instances view addresses the spatial arrangement of all classes.

To also account for distances of selected instances to the class center, we present a strip plot [BHZ*21] visualization, shown in Figure 3 (Class Centroid Distances view). The view also shows every individual instance as a thin line mark, with gray instances being unlabeled and global class colors for labeled instances. The vertical position shows item distances (global normalization), whilst classes are distributed horizontally. The view allows both the analysis and comparison of density characteristics for every class and the comparison of density characteristics between labeled and unlabeled instances. In the example, the classes 2, 5, 6, and 8 seem to have instances with comparatively large class centroid distances, i.e., the classes have high within-class variations. The largest class centroid distance across all classes can be observed for an instance of class 1, also defining the global maximum distance (y axis).

4.4. Observing the Predictions of the Classifier (R4)

Traditionally, the output of the underlying classifier is maybe the best described relation to ISS behavior, as many ISS are based on active learning leveraging model-based characteristics such as uncertainty [Set12]. For the observation, we deem two aspects important, each of which can be analyzed in a dedicated view. The first aspect is to focus on the predictions per instance (instance granularity), which is why we further extend the All Instances

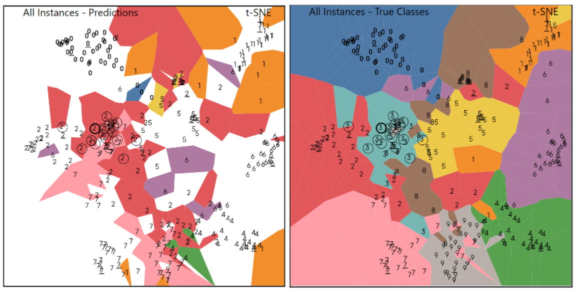

Figure 4: *All Instances View, referring to iteration 17. Here, the two different configurations, i.e. the Model view (left) and the Active Learner view (right), are shown next to each other.*

view with additional optional encodings. Analysts can use the view to observe classifier predictions for every instance using the data space coloring (see Section 4.1). An example can be seen in Figure 4 (left), which shows the predictions for each instance in the unlabeled dataset, highlighting incorrect class predicitons by their color encoding. Analysts can use this as a visual cue for where the model could still be improved. The second aspect of observation shifts the focus towards class characteristics (class granularity). Inspired by confusion matrices [HRS*22], we show all pairwise confusions in the unlabeled data in a corresponding heatmap in Figure 3 (right). The color-coding highlights pairs of classes which are currently particularly difficult to predict correctly and helps analysts to focus on decision boundaries between most problematic pairs of classes, which is a prominent design target for active learning ISS [TVC*11]. In the example, the dominant pairwise class confusion is between class 2 and 3, with 18 occurrences.

5. Case Study

To demonstrate the effectiveness of our approach, we investigate a particularly interesting ISS in a case study. We observe a simulated quasi-optimal greedy strategy that always picks the instance that improves learning performance most using ground truth information [BZL*18] (more details in the supplemental materials.) Even if this ISS is not a strategy that can be used in a practical setting (since it actually cheats), it is of particular interest, because it widely outperforms conventional ISS [BHZ*18] and thus holds unexplored potential. We want to leverage our VA approach to answer the question: *what selection criteria does greedy ISS implicitly apply along the labeling process?* Thereby, we want to shed light on the internal hidden logic and selection criteria to better understand what makes a strategy "optimal". We refer to the supplemental materials for a detailed description of the dataset, the underlying classifier, as well as an encompassing overview of (visual) findings along the process.

The tool at hand guides the user through the exploratory analysis of ISS. Figure 1 shows the state of the tool at iteration 19. This means that at this moment, 19 instances are in the labeled set, and the bar charts reveal that the labeled instances are evenly distributed among all classes, except the class representing no. 8. This number, however, will be the label of the next instance which is selected, as can be seen in Figure 1, e.g., in the Instance Selection Scores view (circle around the instance to be labeled next). This visual cue is given in several components of the visualization. Interestingly, not a single instance in the unlabeled set is predicted to have label 8, as can be seen in the Pairwise Confusions view. This can be explained by the imbalance in the labeled set at this iteration, because the model is biased towards the instances which provide a majority to the train set. Moreover, in the Instance Selection Scores view, it is visible that the instances in the cluster of 8s from which the next instance will be selected all have the highest scores. Such a balancing behaviour was also detected at iteration 9, as well as by Bernard et al. [BHL*18]. Thus, this analysis confirms the assumption that a quasi-optimal greedy ISS favors balanced datasets in the early stages of the labeling process. More interesting findings can be made, e.g., in iteration 17. As it can be seen in the Pairwise Confusions view in Figure 3, the ISS tries to tackle the largest source of class confusions, i.e., 18 instances with true label 3 that are predicted as 2. We use highlighting (circular outline) to analyze the spatial characteristics of these 18 affected instances in the All Instances view, using a visualization where only false positive predictions of the classifier are shown (Figure 4, left). The visualization shows that the ISS not only tackles the largest class confusion, but it also selects an instance in a particularly dense area of the confusion.

To summarize, our approach is a first step towards enabling a deeper look into the selection criteria of a given ISS. We demonstrate the usefulness of our tool with the validation of many of the findings of previous research on the greedy ISS [BHL*18].

6. Discussion

This VA approach supports the analysis of individual labeling iterations, as well as the entire process. We report on several aspects of discussion items, limitations, and future work ideas. To start with, we used only one dataset in the manuscript, meaning that claims for generalizability and scalability need to be postponed to future work. One remaining challenge is a meaningful iconic representation of classes, beyond handwritten digits. We deem the focus on metrics useful, even if we did not exploit the whole design space for metrics. One future work aspect includes the systematic design of metrics. Further, in our examples and the cases study, we limited scope to one ISS only. Due to space limitations, the study on other interesting cases can not be presented. Finally, we echo the observation of the case study that even our quasi-optimal strategy seemed not always to be "perfect". We point to other quasi-optimal implementations, which however are computationally much more expensive to be pre-computed. Another discussion aspect is the use of dimensionality-reduced data in combination with the Voronoi area coloring. We are aware that this cascade of data representations may communicate errors, but realized through careful study that the benefit seems to outweigh the downsides. Still, this representation needs to be used with care, e.g., by seamlessly switching back and forth between data space coloring and direct instance coloring. Our final point of discussion regards the targeted user group, which is ML and/or VIS experts interested in the systematic study of ISS behavior. To make the approach applicable also for other/larger user groups, collaborative approaches and design study settings may be useful.

7. Conclusion

We have presented a VA approach for the interactive visual observation of ISS to improve the transparency of interactive machine learning. Multiple linked views allow the assessment of both ISS behavior within iterations as well as across the iterations of a longitudinal labeling process. The findings of our case study show the usefulness of our approach, e.g., when observing a quasi-optimal ISS, whose internal working is unknown. This work is one step towards the long-term vision of leveraging lessons learned from user strategies and from quasi-optimal ISS in particular, to design and implement model-based ISS of yet unimaginable good quality.

References

[AB18] ADADI A., BERRADA M.: Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access 6* (2018), 52138–52160. 2

[ABC*19] ARYA V., BELLAMY R. K., CHEN P.-Y., DHURANDHAR A., HIND M., HOFFMAN S. C., HOUDE S., LIAO Q. V., LUSS R., MOJSILOVIĆ A., ET AL.: One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques. *arXiv:1909.03012* (2019). 2

[ACKK14] AMERSHI S., CAKMAK M., KNOX W. B., KULESZA T.: Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine 35*, 4 (2014), 105–120. 1

[BHL*18] BERNARD J., HUTTER M., LEHMANN M., MÜLLER M., ZEPPELZAUER M., SEDLMAIR M.: Learning from the best: Visual analysis of a quasi-optimal data labeling strategy. In *EuroVis Short Papers* (2018), Eurographics Association, p. 95–99. 1, 2, 3, 4

[BHS*21] BERNARD J., HUTTER M., SEDLMAIR M., ZEPPELZAUER M., MUNZNER T.: A taxonomy of property measures to unify active learning and human-centered approaches to data labeling. *ACM Transactions on Interactive Intelligent Systems (TiiS) 11*, 3-4 (2021), 1–42. 2, 3

[BHZ*18] BERNARD J., HUTTER M., ZEPPELZAUER M., FELLNER D., SEDLMAIR M.: Comparing visual-interactive labeling with active learning: An experimental study. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 24*, 1 (2018), 298–308. 1, 2, 3, 4

[BHZ*21] BERNARD J., HUTTER M., ZEPPELZAUER M., SEDLMAIR M., MUNZNER T.: Proseco: Visual analysis of class separation measures and dataset characteristics. *Computers and Graphics 96* (2021), 48–60. 3

[BLBC12] BROWN E. T., LIU J., BRODLEY C. E., CHANG R.: Disfunction: Learning distance functions interactively. In *IEEE Visual Analytics Science and Technology (VAST)* (2012), IEEE Computer Society, pp. 83–92. doi:10.1109/VAST.2012.6400486. 3

[BZL*18] BERNARD J., ZEPPELZAUER M., LEHMANN M., MÜLLER M., SEDLMAIR M.: Towards user-centered active learning algorithms. *Computer Graphics Forum (CGF) 37* (2018), 121–132. 4

[CBC*20] CHEGINI M., BERNARD J., CUI J., CHEGINI F., SOURIN A., ANDREWS K., SCHRECK T.: Interactive visual labelling versus active learning: an experimental comparison. *Frontiers of Information Technology & Electronic Engineering 21* (2020), 524–535. 1, 2, 3

[CDH*16] CHEN X., DUAN Y., HOUTHOOFT R., SCHULMAN J., SUTSKEVER I., ABBEEL P.: Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in Neural Information Processing Systems* (2016), pp. 2172–2180. 2

[EBJ*22] EIRICH J., BONART J., JÄCKLE D., SEDLMAIR M., SCHMID U., FISCHBACH K., SCHRECK T., BERNARD J.: IRVINE: A design study on analyzing correlation patterns of electrical engines. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 28*, 1 (2022), 11–21. doi:10.1109/TVCG.2021.3114797. 2

[FZL12] FU Y., ZHU X., LI B.: A survey on instance selection for active learning. *Knowledge and Information Systems 35* (2012). 1, 3

[HMdCM17] HUANG L., MATWIN S., DE CARVALHO E. J., MINGHIM R.: Active learning with visualization for text data. In *ACM Workshop on Exploratory Search and Interactive Data Analytics* (2017), ACM, pp. 69–74. doi:10.1145/3038462.3038469. 3

[HRS*22] HINTERREITER A. P., RUCH P., STITZ H., ENNEMOSER M., BERNARD J., STROBELT H., STREIT M.: Confusionflow: A model-agnostic visualization for temporal analysis of classifier confusion. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 28*, 2 (2022), 1222–1236. doi:10.1109/TVCG.2020.3012063. 2, 4

[KCH*17] KOTTKE D., CALMA A., HUSELJIC D., KREMPL G., SICK B.: Challenges of reliable, realistic and comparable active learning evaluation. In *Workshop and Tutorial on Interactive Adaptive Learning* (2017), vol. 1924, CEUR-WS.org, pp. 2–14. 1

[KPSK17] KUCHER K., PARADIS C., SAHLGREN M., KERREN A.: Active learning and visual analytics for stance classification with ALVA. *ACM Transactions on Interactive Intelligent Systems (TiiS) 7*, 3 (2017), 14:1–14:31. doi:10.1145/3132169. 2

[Kra13] KRAMER O.: *K-Nearest Neighbors*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 13–23. 2

[LBBH98] LECUN Y., BOTTOU L., BENGIO Y., HAFFNER P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE 86*, 11 (1998), 2278–2324. doi:10.1109/5.726791. 2

[PHV*18] PEZZOTTI N., HÖLLT T., VAN GEMERT J., LELIEVELDT B. P. F., EISEMANN E., VILANOVA A.: Deepeyes: Progressive visual analytics for designing deep neural networks. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 24*, 1 (2018), 98–108. 2

[RAL*17] REN D., AMERSHI S., LEE B., SUH J., WILLIAMS J. D.: Squares: Supporting interactive performance analysis for multiclass classifiers. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 23*, 1 (2017), 61–70. doi:10.1109/TVCG.2016.2598828. 2

[REHT19] RODRIGUES F. C. M., ESPADOTO M., HIRATA R., TELEA A. C.: Constructing and visualizing high-quality classifier decision boundary maps. *Information 10*, 9 (2019). 3

[RXC*21] REN P., XIAO Y., CHANG X., HUANG P.-Y., LI Z., GUPTA B. B., CHEN X., WANG X.: A survey of deep active learning. URL: https://doi.org/10.1145/3472291, doi:10.1145/3472291. 1

[SA15] SEDLMAIR M., AUPETIT M.: Data-driven evaluation of visual quality measures. *Computer Graphics Forum (CGF) 34*, 3 (2015), 201–210. doi:10.1111/cgf.12632. 3

[SC08] SETTLES B., CRAVEN M.: An analysis of active learning strategies for sequence labeling tasks. In *Empirical Methods in Natural Language Processing (EMNLP)* (2008), ACL, pp. 1070–1079. 1

[Set12] SETTLES B.: *Active Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2012. doi:10.2200/S00429ED1V01Y201207AIM018. 1, 2, 3

[SG10] SEIFERT C., GRANITZER M.: User-based active learning. In *IEEE Conference on Data Mining Workshops (ICDMW)* (2010), pp. 418–425. doi:10.1109/ICDMW.2010.181. 1, 3

[Sim49] SIMPSON E. H.: Measurement of diversity. *nature 163*, 4148 (1949), 688–688. 2

[SJS*21] SEVASTJANOVA R., JENTNER W., SPERRLE F., KEHLBECK R., BERNARD J., EL-ASSADY M.: Questioncomb: A gamification approach for the visual explanation of linguistic phenomena through interactive labeling. *ACM Transactions on Interactive Intelligent Systems (TiiS) 11*, 3-4 (2021), 1–38. doi:10.1145/3429448. 2

[SNLH09] SIPS M., NEUBERT B., LEWIS J. P., HANRAHAN P.: Selecting good views of high-dimensional data using class consistency. *Computer Graphics Forum (CGF) 28*, 3 (2009), 831–838. 2

[STMT12] SEDLMAIR M., TATU A., MUNZNER T., TORY M.: A taxonomy of visual cluster separation factors. *Computer Graphics Forum (CGF) 31* (2012), 1335–1344. 2, 3

[TKC17] TAM G. K. L., KOTHARI V., CHEN M.: An analysis of machine- and human-analytics in classification. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 23*, 1 (2017), 71–80. 2

[TVC*11] TUIA D., VOLPI M., COPA L., KANEVSKI M., MUNOZ-MARI J.: A survey of active learning algorithms for supervised remote sensing image classification. *IEEE Journal of Selected Topics in Signal Processing 5*, 3 (2011), 606–617. 4

[VPS*] VENDRIG J., PATRAS I., SNOEK C., WORRING M., DEN HARTOG J., RAAIJMAKERS S., VAN REST J., VAN LEEUWEN D. A.: TREC feature extraction by active learning. In *Text REtrieval Conference, TREC*, vol. 500-251, NIST. 1

[WAG05] WILKINSON L., ANAND A., GROSSMAN R. L.: Graph-theoretic scagnostics. In *IEEE Symposium on Information Visualization* (2005), pp. 157–164. doi:10.1109/INFOVIS.2005.14. 2

[WKBD06] WU Y., KOZINTSEV I., BOUGUET J.-Y., DULONG C.: Sampling strategies for active learning in personal photo retrieval. In *IEEE International Conference on Multimedia and Expo* (2006), IEEE, pp. 529–532. doi:10.1109/ICME.2006.262442. 1

[YL16] YANG Y., LOOG M.: Active learning using uncertainty information. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (2016), pp. 2646–2651. doi:10.1109/ICPR.2016.7900034. 2

[ZYMW19] ZHANG Q., YANG Y., MA H., WU Y. N.: Interpreting cnns via decision trees. In *Computer Vision and Pattern Recognition* (2019), pp. 6261–6270. 2