ZEBRA: Explaining rare cases through outlying interpretable concepts

Pedro Madeira[1]
pedro.madeira@fraunhofer.pt

André Carreiro[1]
andre.carreiro@fraunhofer.pt

Alex Gaudio[2]
agaudio@andrew.cmu.edu

Luís Rosado[1]
luis.rosado@fraunhofer.pt

Filipe Soares[1]
filipe.soares@fraunhofer.pt

Asim Smailagic[2]
asim@andrew.cmu.edu

[1]Fraunhofer Portugal AICOS
Porto, Portugal

[2]Carnegie Mellon University
Pittsburgh, PA, USA

Abstract

Anomaly detection methods can detect outliers, but what are the properties of an outlier? In this paper, we propose ZEBRA, a novel framework for generating explanations of an outlier based on the analysis of feature rarity in an interpretable feature space. The contributions of our work include: (a) a modular model-agnostic framework for explanations of outliers; (b) a statistical explanation method based on a rarity score and weighted aggregation functions; (c) multimodal explanations combining visual, textual, and numeric explanations. ZEBRA simplifies the mapping of low-level features to high-level concepts to generate multimodal and human-readable explanations of outliers.

1. Introduction

"When you hear hoofbeats, think of horses, not zebras". This well-known maxim is commonly taught in medical schools to stress how common diagnoses should be preferred over rare ones. However, detecting and understanding outliers or abnormalities remains crucial, especially in high-stakes settings like the medical domain. While anomaly detection methods can identify outliers, understanding their anomalous nature is challenging. ZEBRA aims to explain outliers by analyzing and presenting their outlying and interpretable properties. Anomaly detection methods for "observations which appear to be inconsistent with the remainder of that set of data" [2] have evolved since 1852 [30], with numerous statistical techniques and AI algorithms now available [5, 28]. Nevertheless, recent advances in Explainable AI (XAI) highlight the need for explainable anomaly detection methods [27, 29].

A recent survey by Panjei *et al*. [29] identifies three types of outlier explanations: **(1)** importance level of outliers, also known as ranking; **(2)** causal interactions among outliers; and **(3)** outlying attributes of outliers. Type (1) explanations can be further divided into model-specific [5, 13] and model-agnostic methods [1, 11, 12, 14, 17, 34]. As for type (3) explanations, current tools face three main challenges: **(3.a)** limiting subspace search, **(3.b)** generating readily interpretable output, and **(3.c)** incorporating user's prior knowledge about the attributes. Reviewing existing works, Refout [12] and LookOut [11] address both challenges (3.a) and (3.b) through heuristic-based subspace search; Explainer [14] tackles (3.b) by developing sapling random forests and methods for anomaly clustering; COIN [17] considers user knowledge about contributing attributes to the outlierness, satisfying challenges (3.b) and (3.c). None of these techniques address all three challenges. In this taxonomy, ZEBRA embodies a model-agnostic hybrid method, providing explanations of types (1) via per-sample numerical rarity scores to rank outliers, and (3) with learning the contributions of features to the inlier/outlier classification. ZEBRA uniquely addresses all three challenges by conducting rarity analysis on interpretable features and mapping them into higher-level user-provided concepts.

With respect to evaluating outlier explanations, most techniques address only their truthfulness [29]. For numerical rankings, usual supervised performance metrics can be used after choosing a threshold for the minimum outlier score (e.g., ROC AUC, accuracy). However, regarding outlying attributes, ground-truth information is more challenging, often restricting evaluation to an empirical discussion. To overcome this limitation, researchers can modify real-world datasets by adding noise attributes [17] or reducing the examples in specific categories to become outliers [8], create synthetic datasets [8] or simulate analyst feedback [34]. The present paper uses a mix of techniques to create a toy example to showcase the framework since the real-world use case under study does not have any ground-truth, nor do we have access to domain experts for validation.

The contributions of ZEBRA are three-fold: a) a modular model-agnostic framework for explaining outliers; b) a statistical-based rarity score for individual features, along with score aggregation functions that integrate mappings of low-level features to user-relevant concepts; c) multimodal explanations combining visual, textual, and numeric components, presented through a versatile Rarity Card.

Following this introductory section, Sec. 2 describes the ZEBRA framework. Sec. 3 and Sec. 4 present two illustrative use cases and their results. Finally, Sec. 5 concludes with the main findings and future research directions.

2. ZEBRA framework

ZEBRA is a modular, pluggable framework and interface for explanations of outliers. Figure 1 illustrates the workflow of the framework and its modular components. Each modular component is subsequently described.

Feature Extraction. ZEBRA leverages an interpretable feature space. We use the Texture-Color-Geometry Feature Extraction (TCGFE) library [32] to summarize each image with 152 interpretable continuous features. Features describe particular aspects of high-level concepts: Texture, Color and Geometry. If available, image masks may be used to focus the feature representation on regions of interest. Note that this module is extensible to other feature extractors in a pluggable design. The only requirement to maintain the interpretability aspect is to ensure that extracted features can be mapped to human-understandable concepts.

Rarity Scores. A numeric rarity score is computed for each feature of a given image. Rarity of a feature value describes the parameterized probability density estimate of the sample, and its computation is described in Sec. 2.1.

Aggregation of Rarity. The aggregation of feature rarity scores relates samples to higher-level concepts and also enables measuring the overall rarity of a given sample or feature. Model-based aggregation enables context-specific explanations of outliers allowing rarity to account for known or assumed bias. Aggregation is described in Sec. 2.2. This step connects rarity with two modular components of the ZEBRA framework:

i **High-level Concept Mapping.** A mapping between (groups of) lower-level features to high-level user-relevant concepts can be provided, resulting in interpretable concept-level rarity scores. Examples of high level concepts include general image-related concepts like Color, Texture and Geometry features, or clinical criteria in healthcare applications.

ii **Explanations for Outlier Detectors.** Model-based aggregation enables ZEBRA explanations to represent the biases of an outlier detection method or classifier. Examples of context-aware weighted aggregation are given in Sec. 2.2.

Multimodal Explanations in a Rarity Card. The explanations from ZEBRA are visually consolidated into an automatically-generated Rarity Card, as shown in Fig. 1. The multimodal explanations include numerical rankings or scores, visual plots, and textual summaries to aid user understanding of the outlying properties of the sample. Explanations are queryable over a dataset of images with a web application and platform developed using the `Streamlit` library. ZEBRA provides interactive visualizations, such as correlation plots (e.g., outlier detector Anomaly Score vs. overall / concept-aggregate Rarity Score), as shown in Fig. 2, and UMAP-based plots [21] with interactive display of sample information.

Additional implementation features of the ZEBRA platform include:

a **Generalization across modalities.** Rarity scores are calculated from features computed by any feature extractor, which allows for extensive applicability to various data modalities (e.g., image, time series, text, or speech).

b **Class-specific analyses.** Sample rarity is preferably restricted to a specific data class (or group), safeguarding the possible existence of contextual outliers [7] and generating appropriate explanations for them.

c **Outlier labelling.** Outlier detection is built into the framework (e.g., applying a threshold to samples' rarity scores). However, the user can also submit ground-truth anomaly labels, or apply an outlier detector (OD) to obtain approximate results. For the latter, our implementation has two options although others can be used: the Isolation Forest [16] and Local Outlier Factor [4].

2.1. Rarity Score

The proposed numerical Rarity Score aims to characterize each descriptive feature of a sample based on its likelihood of occurrence. The steps for calculating the Rarity Score are described in Algorithm 1. The calculation of the probability density function (PDF) of each feature's distribution is achieved by Kernel Density Estimation (KDE), using the Epanechnikov kernel [9] given its computational efficiency (compact support) and robustness to the presence of outliers:

$$K(u) = \begin{cases} \frac{3}{4}(1-u^2), & |u| \leq 1 \\ 0, & |u| > 1 \end{cases}$$

where u is the distance from the center of the kernel, and $K(u)$ is the weight assigned to a data point at that distance. A bandwidth h is used to adjust the kernel to the data by:

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$$

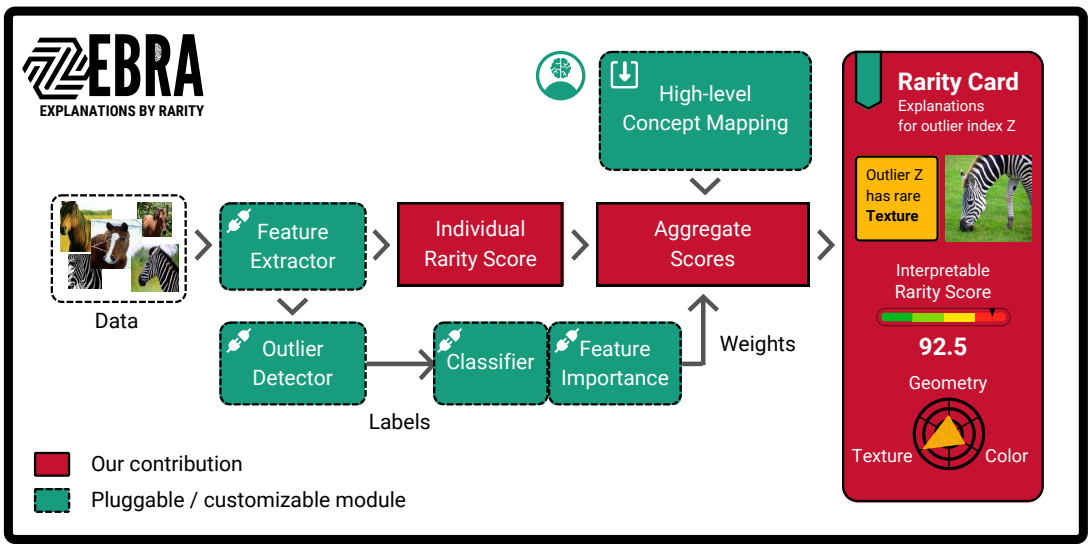

Figure 1. **ZEBRA's modular and pluggable architecture generates multimodal explanations of why an outlier is rare.** Choice of a feature extractor, outlier detector, and high-level concept space enables users to interact with explanations of samples, features, concepts and outlier detectors. The explanations are contextualized on automatically-generated rarity cards displayed to users in an interactive web interface.

Algorithm 1 Rarity Score

Input: A matrix of interpretable features $F \in \mathbb{R}^{n,m}$ for n samples and m features.

Output: Matrix of rarity scores $R \in [0, 100]^{n,m}$, where $R_{i,j}$ is the rarity score for feature $F_{i,j}$.

for each feature column $\mathbf{F}_{:,j}$ **do**
  Estimate kernel density $f(j) \triangleq KDE(\mathbf{F}_{:,j})$
  for each sample, $i \in 1..n$ **do**
    Calculate the density $e_{i,j} \leftarrow f(i|j)$
  end for
  Form the matrix $E \in \mathbb{R}^{n,m}$ from all $e_{i,j}$
  Normalize each row $\mathbf{E}_{i,:} \leftarrow \frac{\mathbf{E}_{:,j}}{||\mathbf{E}_{:,j}||_2}$
  Apply Element-wise Inversion $E_{i,j} \leftarrow \frac{1}{E_{i,j}}$
  Scale each column $\mathbf{R}_{:,j} \leftarrow 100 \times \frac{\mathbf{E}_{:,j} - \min(\mathbf{E}_{:,j})}{\max(\mathbf{E}_{:,j}) - \min(\mathbf{E}_{:,j})}$
end for

and its selection is done via the Freedman-Diaconis method [10], a data-driven rule adaptable to different distributions:

$$h = 2 \times \frac{\text{IQR}(\mathbf{F}_{:,j})}{n^{1/3}}$$

where IQR denotes the interquartile range of the column vector $\mathbf{F}_{:,j}$ of feature j, and n is the number of samples.

2.2. Aggregation of Rarity Scores

Mapping groups of features to a given concept allows the rarity contribution of each feature to be combined and generate human-understandable high-level explanations. To this end, aggregation is performed both at the conceptual level (aggregating scores for feature subsets), and sample-wise (aggregating considered concepts/features into a single score). These aggregate rarity scores can be computed with an inner product $\langle R_{i,:}, \mathbf{w} \rangle$ with feature weights $\mathbf{w} \in \mathbb{R}^k$, $k = c$ if aggregating c conceptually-related features, or $k = m$ if aggregating all m descriptive features of a sample. The feature weights $\mathbf{w}$ can depend on an auxiliary model or not, distinguishing a **model-based** from a **ranked aggregation**.

Model-based aggregation. Algorithm 1 outputs rarity scores via properties of the data distributions, representing the statistical rarity of a sample in a space of interpretable features. ZEBRA provides an additional step to refine these scores, allowing a model-based aggregation that enables the reflection of known or assumed bias, such as bias of a given outlier detector, providing model context to an otherwise solely data-driven score. Since an outlier detector $o : \mathcal{X} \to \mathcal{Y}$ where $o \sim \mathcal{O}$ evaluates the sample space $\mathcal{X}$, and not the rarity space $\mathcal{F}$ adopted by ZEBRA, an auxiliary model $a : (\mathcal{F}, \mathcal{O}) \to \mathcal{W}$ is used to obtain importance weights. Methods from interpretable artificial intelligence can suffice for $a(\cdot)$. For instance, it is possible to train an interpretable classifier to mimic the result of an outlier detector, and then extract feature importances from the model. Classifiers of this type include linear models, Logistic Regression [31], and Random Forest [3]. A second

class of methods from the field of post-hoc interpretability compute the importance weights directly. These include permutation feature importance [22] and Shapley values via SHAP [18, 19], and post-hoc attribution methods [25]. The aggregation step retrieves the importance weights vector $\mathcal{W}$ to compute the aggregation, generating both concept-related and overall aggregate sample-wise scores.

Ranked aggregation. Independently from any model, in this case the features of each sample i are ranked according to their rarity score value R_{ij} and a weighting vector, $\mathbf{w_i}$, is generated. For a set of k features, $w_{ij} = k - Rank_{ij} + 1, \forall i \in [1, n]$, where w_{ij} represents the weight of feature j for sample i, n is the total number of samples, and $Rank_{ij}$ denotes the rank of feature j for sample i based on its rarity value. The highest score among $R_{i,:}$ has a $Rank_{ij} = 1$ and the largest weight $w_{ij} = k$, whereas the lowest scoring feature in the k-set has a weight of $w_{ij} = 1$. This type of aggregation is used when one opts not to apply the model-based option, to further evidence the differences in feature rarity for the different samples in a dataset.

2.3. Regularity and Normality of Rarity Scores

Kriegel *et al*. [15] propose that outlier scores should satisfy properties of regularity and normality in order to differentiate outliers from inliers. The individual (feature-level) scores calculated by our method are regular and normalized in the interval [0, 100]. The aggregate scores (by concept or sample) also meet these two requirements, since the transformations applied to the combination of feature-level scores are classified as *regular*, *normal* and *ranking-stable* (according to the taxonomy in [15]).

3. General Image Concept-based explanations: horses and zebras

Taking the aphorism presented in the motivation of this work, the distinction between a horse and a zebra is a good example for a concept-based explanation task. Considering general image-based concepts such as Texture, Color, and Geometry, applicable to a myriad of use cases and domains, it is intuitive to find the rare zebras in a dataset comprised mainly of horses by emphasizing texture and color features over geometry features. Can we use the ZEBRA framework to automatically identify zebras and explain them based on the outlying attributes?

To answer this question, we built a toy example with 290 horses and 10 zebras randomly picked from the Horse2zebra dataset in the UC Berkeley's official directory of CycleGAN Datasets [35]. The TCGFE library was then used to extract texture, color and geometry features from the target images. Subsequent computation of rarity scores and their ranked aggregation resulted in a rarity score for each sample, according to the previously described procedure.

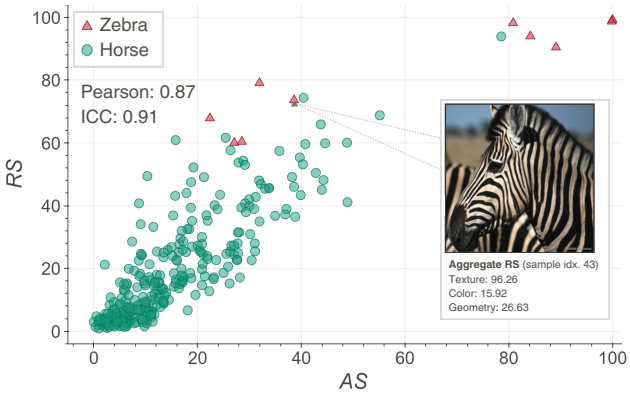

Figure 2. **Horses vs. Zebras:** Correlation plot between sample-wise ranked aggregated Rarity Scores (RS) and Anomaly Scores (AS) returned by the outlier detector. Pearson and Intra-class correlation (ICC) coefficients are presented. In this type of interactive plot, sample cards are shown by hovering over the data points, displaying the sample's image and concept-related aggregated rarity scores.

Figure 2 shows the comparison between our sample-wise ranked aggregated rarity scores (RS) and the anomaly scores (AS) returned by an Isolation Forest OD. We observe a good correlation between RS and AS, with a Pearson correlation of 0.87 and Intra-class Correlation (ICC) of 0.91. Additionally, we can observe that all zebras can be found with a threshold of RS $>$ 60, whereas for AS that threshold should be AS $>$ 22, which would result in a very low precision. Considering a threshold of 60 for both scores we obtain an outlier recall of 50% for AS, and 100% for RS. F1-scores are 62.5% and 74% for AS and RS, respectively. These thresholds may be chosen through visualization using the interactive plot, or objectively based on performance curves for a training or validation set (e.g., Precision-Recall Curve).

4. Clinical criteria-based explanations: Cervical cytology

As another example of application, we illustrate how ZEBRA can be applied in a medical use case, namely cervical cytology. According to the World Health Organization, cervical cancer is the fourth most common cancer in women, and curable if diagnosed early [26]. Squamous cell lesions represent up to 80% of cervical cancers and can be classified into five levels of abnormality [6, 20]: Atypical squamous cells of undetermined significance (ASC-US); Atypical squamous cells, cannot exclude a high-grade squamous intraepithelial lesion (ASC-H); Low-grade squamous intraepithelial lesions (LSIL); High-grade squamous intraepithelial lesions (HSIL); and squamous cell carcinoma (SCC). Additionally, the following nuclei criteria are rele-

vant to assess squamous cell nuclei abnormality: enlarged nucleus (Large), heterogeneous chromatin color (HCC) and texture (HCT), hypo- and hyperchromatism (Chroma), and presence of nucleolus (Pres. Nuc.).

Given the current advances in developing AI-powered mobile-based solutions to support cervical cancer screening in medically underserved areas [23, 24, 33], we used the Nuclei-based Cervical Lesions Dataset [23], which consists of 31698 and 1395 normal and abnormal squamous cell nuclei annotations, respectively. The TCGFE library was then used to extract texture, color, and geometry features from those annotated regions. Moreover, as an example of how domain-related high-level concepts can improve the interpretability of the rarity explanations, an empirical mapping between the TCGFE features and the aforementioned nuclei criteria was explored.

Figure 3 presents the Rarity Cards generated for the ASC-US class, covering a comparative inlier and rare cases for general image concepts (texture, color, and geometry) - top row - and the mapped clinical criteria - bottom row. Considerable differences in texture, geometry and color can be seen in Figures 3b and 3c, respectively. Despite having no expertise in cytological imaging for cervical cancer, it is possible for the reader to observe in Fig. 3f the presence of nucleolus (smaller, round structures).

Based on this information, it is possible to perform input validation of new samples, by checking if they fall out of scope of the training data (see Fig. 3c with no nucleous and extreme color rarity). Moreover, one could also evaluate the rare examples in specific classes, and strive for more targeted collection of similar examples to increase representativity and avoid negative bias.

5. Conclusion

We introduce ZEBRA, a modular model-agnostic framework to identify and explain rare cases using human-interpretable concepts displayed in a multimodal Rarity Card. We additionally propose a statistical explanation method based on a rarity score and weighted aggregation functions, and we contribute with Rarity Cards combining textual, visual, and numerical explanations. This approach uniquely addresses the three challenges of limiting subspace search, generating interpretable outputs, and including prior knowledge. Preliminary results in two different use cases suggest that our rarity score can be used for outlier detection and interpretable explanations, showing good correlation with existing outlier detectors. Without ground-truth labels or expert validation in the cytology example, we rely on visual analysis of the Rarity Cards, where considerable differences in the high-level concepts are observed.

Future work will consider estimated joint distributions of multiple features, and we will extend our framework to other data modalities, such as time series or text, and feature types (e.g., binary and categorical). Additionally, we plan to investigate concept learning, where interpretable concepts can be automatically extracted from the data, leveraging exemplar samples or self-supervision techniques, (e.g., using medical images accompanied by clinical notes).

ZEBRA explains outliers by their rarity in a domain-relevant concept space, empowering users with actionable insights to understand and address outliers across different applications.

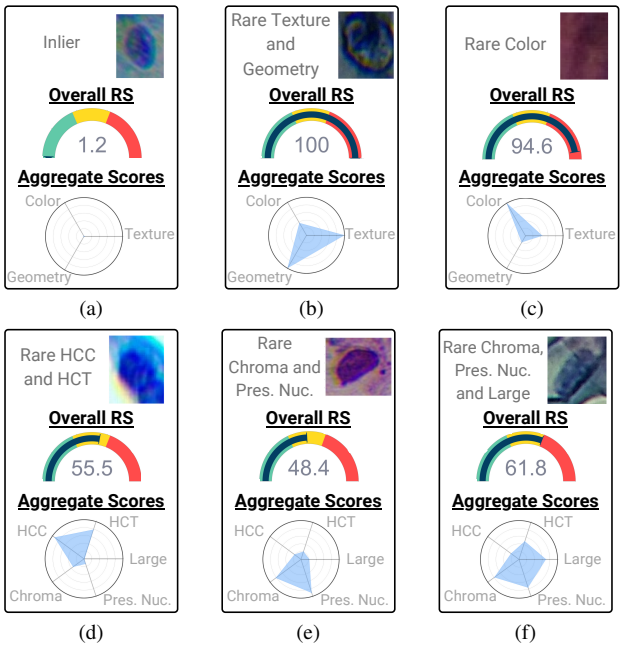

Figure 3. **Examples of inlier sample and Rarity Cards for cytological outliers with multimodal explanations:** text summary of rare concepts, visual meter of overall rarity score for the sample, and plot of rarity scores aggregated by: a-c) general image concepts (color, texture, geometry); d-f) mapped clinical criteria.

6. Acknowledgements

This work was developed under the scope of the Carnegie Mellon University (CMU) Portugal Visiting Students Program, within the project Transparent Artificial Medical Intelligence (TAMI) - co-funded by Portugal 2020 framed under the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), Fundação para a Ciência e Tecnologia (FCT), Carnegie Mellon University, and European Regional Development Fund under Grant 45905 -, and project "Center for Responsible AI project number C645008882-00000055", supported by European funds through Plano de Recuperação e Resiliência.

References

[1] Kasun Amarasinghe, Kevin Kenney, and Milos Manic. Toward explainable deep neural network based anomaly detection. In *2018 11th International Conference on Human System Interaction (HSI)*, pages 311–317, 2018. 1

[2] Vic Barnett, Toby Lewis, et al. *Outliers in statistical data*, volume 3. Wiley New York, 1994. 1

[3] Leo Breiman. Random forests. *Machine Learning*, 45:5–32, 10 2001. 3

[4] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. Lof: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, SIGMOD '00, page 93–104, New York, NY, USA, 2000. Association for Computing Machinery. 2

[5] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3), jul 2009. 1

[6] Teresa Conceição, Cristiana Braga, Luís Rosado, and Maria João M. Vasconcelos. A review of computational methods for cervical cells segmentation and abnormality classification. *International Journal of Molecular Sciences*, 20:5114, 10 2019. 4

[7] Divya D. and Dr Sasidhar Babu. Methods to detect different types of outliers. pages 23–28, 03 2016. 2

[8] Xuan Hong Dang, Barbora Micenková, Ira Assent, and Raymond T. Ng. Local outlier detection with interpretation. In Hendrik Blockeel, Kristian Kersting, Siegfried Nijssen, and Filip Železný, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 304–320, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. 1

[9] V. A. Epanechnikov. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, 14(1):153–158, 1969. 2

[10] David A. Freedman and Persi Diaconis. On the histogram as a density estimator:l2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57:453–476, 1981. 3

[11] Nikhil Gupta, Dhivya Eswaran, Neil Shah, Leman Akoglu, and Christos Faloutsos. Beyond outlier detection: Lookout for pictorial explanation. In *ECML/PKDD*, 2018. 1

[12] Fabian Keller, Emmanuel Müller, Andreas Wixler, and Klemens Böhm. Flexible and adaptive subspace search for outlier analysis. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, CIKM '13, page 1381–1390, New York, NY, USA, 2013. Association for Computing Machinery. 1

[13] Edwin M. Knorr and Raymond T. Ng. Finding intensional knowledge of distance-based outliers. In *Very Large Data Bases Conference*, 1999. 1

[14] Martin Kopp, Tomáv*s*. Pevný, and Martin Holeňa. Interpreting and clustering outliers with sapling random forests. 2014. 1

[15] Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek. Interpreting and unifying outlier scores. In *SDM*, 2011. 4

[16] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422, 2008. 2

[17] Ninghao Liu, Donghwa Shin, and Xia Hu. Contextual outlier interpretation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 2461–2467. International Joint Conferences on Artificial Intelligence Organization, 7 2018. 1

[18] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence*, 2(1):2522–5839, 2020. 4

[19] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017. 4

[20] Christian Marth, Fabio Landoni, Sven Mahner, Mary McCormack, Antonio Gonzalez-Martin, and Nicoletta Colombo. Cervical cancer: Esmo clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 28:iv72–iv83, 7 2017. 4

[21] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2018. 2

[22] Christoph Molnar. *Interpretable Machine Learning*. 2019. https://christophm.github.io/interpretable-ml-book/. 4

[23] Vladyslav Mosiichuk, Ana Sampaio, Paula Viana, Tiago Oliveira, and Luís Rosado. Improving mobile-based cervical cytology screening: A deep learning nuclei-based approach for lesions detection. (submitted). 5

[24] Vladyslav Mosiichuk, Paula Viana, Tiago Oliveira, and Luís Rosado. Automated adequacy assessment of cervical cytology samples using deep learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13256 LNCS:156–170, 2022. 5

[25] Ian E. Nielsen, Dimah Dera, Ghulam Rasool, Ravi P. Ramachandran, and Nidhal Carla Bouaynaya. Robust explainability: A tutorial on gradient-based attribution methods for deep neural networks. *IEEE Signal Processing Magazine*, 39(4):73–84, 2022. 4

[26] World Health Organization. Cervical cancer. 4

[27] Guansong Pang and Charu Aggarwal. Toward explainable deep anomaly detection. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, page 4056–4057, New York, NY, USA, 2021. Association for Computing Machinery. 1

[28] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep learning for anomaly detection: A review. *ACM Comput. Surv.*, 54(2), mar 2021. 1

[29] Egawati Panjei, · Le Gruenwald, · Eleazar Leal, · Christopher Nguyen, and Shejuti Silvia. A survey on outlier explanations. *The VLDB Journal*, 31:977–1008, 2022. 1

[30] Benjamin Peirce. Criterion for the rejection of doubtful observations. *The Astronomical Journal*, 2:161–163, July 1852. 1

[31] Chao-Ying Joanne Peng, Kuk Lida Lee, and Gary M. Ingersoll. An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1):3–14, 2002. 3

[32] Luís Rosado, João Gonçalves, João Costa, David Ribeiro, and Filipe Soares. Supervised learning for out-of-stock detection in panoramas of retail shelves. In *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*, pages 406–411, 2016. 2

[33] Ana Filipa Sampaio, Luis Rosado, and Maria Joao M. Vasconcelos. Towards the mobile detection of cervical lesions: A region-based approach for the analysis of microscopic images. *IEEE Access*, 9:152188–152205, 2021. 5

[34] Md Amran Siddiqui, Alan Fern, Thomas G. Dietterich, Ryan Wright, Alec Theriault, and David W. Archer. Feedback-guided anomaly discovery via online optimization. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, page 2200–2209, New York, NY, USA, 2018. Association for Computing Machinery. 1

[35] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017. 4