Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2026) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

The 17th International Conference on Ambient Systems, Networks and Technologies (ANT)
April 14-16, 2026, Istanbul, Türkiye

Tailored ensemble anomaly detection for Internet disruptions

Mike Vandersanden[a,b,*], Jelle Beerts[b], Hanna Kreitem[d], Amreesh Phokeer[d], Wim Lamotte[a,b], Peter Quax[a,b,c]

[a]*Digital Future Lab, Wetenschapspark 2, 3590 Diepenbeek, Belgium*
[b]*Hasselt University, Martelarenlaan 42, 3500 Hasselt, Belgium*
[c]*Flanders Make, Oude Diestersebaan 133, 3920 Lommel, Belgium*
[d]*Internet Society, 1551 Emancipation Highway 1506, Fredericksburg, VA. 22401, U.S.A.*

Abstract

The foundational role of the Internet necessitates robust methods for detecting and understanding disruptions. While machine learning offers a promising avenue for anomaly detection in heterogeneous, high-dimensional Internet measurement datasets, it faces significant challenges, including handling diverse disruption patterns, ensuring consistent performance across use cases, a scarcity of labeled ground truth, and difficulty explaining the black-box models. This paper proposes and thoroughly evaluates an *ensemble anomaly detection approach* for Internet disruption detection that addresses these limitations. This approach leverages multiple *univariate unsupervised base detectors*, each tailored to specific data sources, and combines their outputs through simple, intuitive aggregation mechanisms. An evaluation framework extensively assesses the proposed approach against alternative machine learning methods, demonstrating that the proposed ensemble detector achieves *competitive* and *robust* performance against other multivariate approaches, while crucially offering *inherent explainability* and significantly reducing the burden of parameter tuning. These findings highlight the practical efficacy of explainable, robust ensemble methods for Internet infrastructure monitoring.

Keywords:
Ensemble learning; Machine learning; Anomaly detection; Internet disruptions

1. Introduction

A reliable Internet is foundational to many modern services, including global communication, commerce, and entertainment. Nonetheless, various types of disruptions can compromise this crucial infrastructure, including technical failures [1], natural disasters [2], or deliberate actions [3], such as the recurrent shutdowns in Iran [4]. To strengthen the Internet and ensure future reliability, it is crucial to detect and understand disruptive events, as stated by Aceto et al. [5].

* Corresponding author. Tel.: +32-11-268405.
E-mail address: mike.vandersanden@uhasselt.be

Platforms such as IODA [6] and Internet Health Report [7] gauge service health by probing for availability or performance. Combining diverse measurements enables monitoring multiple aspects of the Internet; distinct disruptions manifest via distinct means. Combining datasets is supported by Raman et al. [8], who indicate that reducing data gaps enhances classification, and highlight the strength of integrating heterogeneous data for Internet health assessment.

Detecting and classifying anomalies within heterogeneous, high-dimensional datasets presents an apparent application for *machine learning*. However, applying traditional machine learning anomaly detection approaches to Internet measurements introduces significant *challenges*. A detector might perform excellently for one use case, yet the performance is not guaranteed to carry over to others, requiring *retraining* or *fine-tuning*. For specific data sources, labeled *ground truth* can be scarce, posing a significant hurdle for (semi-)supervised models. Furthermore, machine learning approaches can operate as *black boxes*, providing detection outputs with no explanation for the detected events.

This paper proposes and thoroughly evaluates an *ensemble anomaly detection* approach for detecting Internet disruptions, addressing these challenges by leveraging *univariate unsupervised base detectors* tailored to specific data sources and combining their outputs through intuitive *aggregation strategies*. This approach achieves competitive and robust performance against complex multivariate anomaly detection methods, while offering *inherent explainability*.

A custom framework evaluates various anomaly detection approaches, highlighting the challenge of identifying a universally well-performing detector. However, it reveals *categories of data sources* with similar characteristics. As specific univariate detectors perform well on certain characteristics, it motivates leveraging tailored base detectors instead of needing to uncover the best-performing base detectors for each data source. Furthermore, the proposed ensemble approach is evaluated against several multivariate anomaly detection approaches. This approach demonstrates *robust* and *competitive performance* across all tested scenarios, without needing extensive parameter tuning.

To gain insights into traditional multivariate approaches, the framework utilizes post-hoc explainability techniques. This work presents similar *intuitive insights* for the proposed ensemble approach through the exposed base detectors, which can be obtained without computationally intensive post-hoc techniques. This enables the direct identification of the specific anomalous features and the contributing base detectors that contribute to the final prediction.

Section 2 posits the proposed approach against related works. Section 3 elaborates on various methods for anomaly detection. Section 4 presents the methodology and evaluation framework. Section 5 evaluates the proposed method on Internet measurements. Finally, Section 6 will conclude this work and suggest future work.

2. Related work

Raman et al. describe the analysis pipeline for Censored Planet, an Internet measurement platform [8]. They emphasize the importance of integrating metadata, as it facilitates understanding of the measurements and, for instance, helps eliminate false positives. These data points are condensed into fingerprints that help identify Internet states as measurements are mapped to them. The importance of integrating multiple sources is substantiated by Sermpezis et al., who reveal biases in Internet measurement platforms [9]. A biased data source may offer faulty insights. Thus, relying on multiple measurement sources may help provide a better understanding of the data. For example, one of the biases is measurement granularity, and different data sources may conduct measurements with different granularity.

There are various methods to detect anomalies within Internet measurements. IODA partially relies on empirically tested heuristics for its anomaly detection [10]. By using domain knowledge, these heuristics can detect a large number of disruptions in real-time. However, these threshold-based heuristics may be brittle to specific disruptions or behavior. Additionally, specific data sources, such as the seasonal Google data source, do not map well to threshold detectors.

Other works achieve a higher performance by exploiting data characteristics. DarkSim, by Gao et al., is an analysis framework for Internet background radiation [11]. Through similarity analysis, this framework defines anomaly pattern templates and detects various anomaly types, including one-off, repeating, and concurrent anomalies. Guillot et al. designed Chocolatine, a machine learning technique that predicts the behavior of a seasonal Internet measurement data source based on ARIMA [12]. Errors between predictions and measurements can indicate a disruption. By understanding the seasonal data sources, this approach can obtain a high true positive rate and few false positives. Richter et al. developed a disruption detection technique that relies on CDN data [13]. This technique identifies anomalous periods by defining usual behavior. On a global scale, this technique identifies several patterns. For example, when disruptions occur or which parts of the Internet are affected. These works highlight the strength of tailored detection models. The proposed ensemble approach leverages a similar concept by finding tailored detectors for a data source.

Vanerio et al. evaluate ensemble learning approaches in a network security context [14]. Their Super Learner overcomes the difficulties of heterogeneous data and performs asymptotically as well as a weighted combination of the best base learners. This provides a robust approach to detecting distinct anomalies with a single model. Internet measurement data displays similarities to heterogeneous network security data, which exhibits varying characteristics.

Explaining the machine learning results of Internet measurements, which, to a large extent, are time series data, can be complex, as highlighted by Rojat et al. [15]. Furthermore, many methods cannot present confidence in the model, mostly indicating which region of the input data receives attention. These methods do provide the user with some explanations or trust in the model. However, the computational complexity poses a significant barrier to getting these insights. The intrinsic explainability of the proposed ensemble approach addresses this limitation.

3. Anomaly detection approaches

This work evaluates three types of anomaly detection approaches: univariate, multivariate, and the proposed ensemble approach. The evaluated approaches are *unsupervised*, relying solely on measurement data and established domain knowledge, as accurate ground truth may not be available for approaches that depend on it. Anomaly Detection Toolkit (ADTK) provides univariate and multivariate detectors [16]. Secondary models for specific multivariate detectors come from scikit-learn [17]. The proposed ensemble approach utilizes ADTK for the base detectors.

3.1. Univariate detectors

Univariate approaches operate on individual data sources, with their effectiveness depending on the characteristics of the data and the anomalies. A data source can remain relatively constant over extended periods, with different sources showing various amounts of noise and variance. Anomalies prevail when a measurement deviates from the baseline. Other data sources display repeating patterns. These seasonal data sources have predictable periodic fluctuations. Anomalies are characterized by deviations from this expected cyclic behavior. This heterogeneity mandates fine-tuning approaches to specific data sources. The IODA anomaly detection system uses empirically tested thresholds [10]. Depending on the data source, the thresholds may be set as high as 99% or as low as 25%. Besides heuristic approaches, IODA employs advanced detectors for specific data sources, such as the Chocolatine detector [12].

ADTK offers a range of univariate anomaly detection models, which can be categorized based on the type of anomaly being targeted. *Outliers* are anomalous measurements, deviating from regular measurements or exceeding certain parameters. Anomalies can be time-dependent; a *spike* or *level shift* is characterized by deviating values from neighboring measurements. Other anomalies are characterized by a *pattern change* or a break in the *seasonal trend*.

3.2. Multivariate detectors

Measurements can be considered related in the case of Internet data, as services rely on the same infrastructure. However, data sources can also be considered individually, as measurements are taken independently from different services, and depending on the cause, disruptions may affect services differently. Multivariate approaches can be used to understand the complex relationships between different data features. As these multivariate models capture complex relationships among features, they typically function as black boxes, providing a single, aggregated anomaly prediction without directly indicating which feature or relationship triggered the detection. Consequently, these approaches require an additional, often computationally intensive, step to provide explainability and clarify the relationships.

Multivariate approaches can identify different types of *outliers* in the data. Measurements can be clustered based on their features. A clustering detector will report the *smallest cluster* as anomalous. Another option is to use regression and rely on the *regressive error* to identify deviating measurements. Similarly, Principal Component Analysis (PCA) can indicate anomalous measurements in a high-dimensional space through the *vector reconstruction error*.

3.3. Proposed ensemble detectors

The proposed ensemble approaches can be classified as multivariate; they consider all features at once. Figure 1 illustrates the proposed approach. Base detectors consider each feature individually, and an aggregation strategy combines these base results into a final prediction. However, instead of a black box, the inner workings remain transparent.

Several types of aggregation strategies exist, including voting, score aggregation, and stacking. This work evaluates four voting aggregation strategies, as they best fit the boolean base detector results. The most straightforward are *majority* and *unanimous voting*, where a percentage of base detectors have to agree on an anomaly for the final prediction. This threshold depends on the desired behavior. For example, it is possible to require a strict or weak majority, specifying whether more than half or at least half of the detectors must agree. The *time window voting* strategy seeks to leverage the temporal relationships between data sources. Different data sources may experience the effects of a disruption at different times. This can result in a disruption period being cut short as the voting threshold may not be reached in time for all data sources. This strategy searches within a specified window to determine whether the data sources are showing anomalous behavior. The *weighted voting* strategy utilizes domain knowledge. Specific data sources, and thus base detectors, are given more weight in determining the final anomaly indication. For example, known trustworthy data sources will be relied upon more, without discarding any other sources. This is similar to how certain univariate detectors use domain knowledge, such as the seasonality of data, to achieve a higher performance.

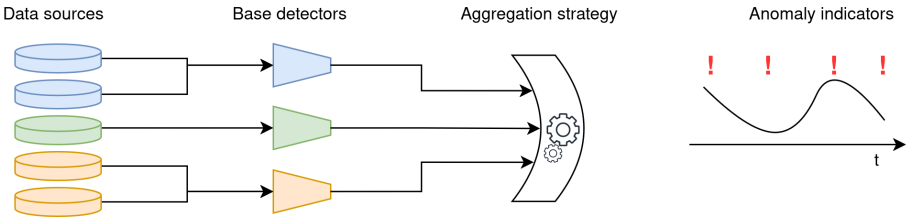


Fig. 1. The proposed ensemble anomaly detection approach. Data sources are fed into detectors, whose results are unified into anomaly indicators.

4. Experimental design

A custom *evaluation framework* assesses the different anomaly detection approaches for several use cases with various data sources. The project repositories contain the framework code, relevant datasets, and experiment results [18][19]. This work considers *four diverse cases*, consisting of measurements from several countries that have a history of using Internet disruptions during exam periods [20]. The diversity of these cases is crucial for evaluating model robustness, as they include periods with missing data sources and a case that appears largely disruption-free.

The framework runs all approaches for a use case and generates a comprehensive report, which includes individual results, visualizations, and summaries. ADTK provides evaluation metrics: the *F1 score*, *precision* (P), *recall* (R). Additionally, an F_β calculation is implemented to calculate *F2 scores*. Whereas F1 is the harmonic mean of recall and precision, F2 weighs recall more than precision. Runs are repeated to prevent biased one-off results, and the mean and standard deviation are shown. During the evaluation, two sets of base detectors will be used for the proposed ensemble approach. The first set utilizes tailored detectors, matching a data source to a univariate detector that is expected to perform well in several cases. A second set of base detectors uses the best-performing univariate detectors. To address the explainability deficit of multivariate models, the framework incorporates SHAP to provide insights. The report examines the explainability of multivariate models using SHAP and the proposed approach through its custom explainability method.

4.1. Data sources

This work integrates data from two publicly available Internet measurement platforms: IODA and Cloudflare Radar. Both platforms cover all use cases and provide disruption notifications as a ground truth. Queries were executed with a uniform one-hour resolution. Since platform-provided ground truth can be incomplete, it was manually inspected and amended using domain knowledge, without relying on any model. This ensures accuracy against all significant Internet disruptions in the data, providing a reliable basis for evaluation. IODA is a *monitoring* service that probes Internet services for availability and aggregates external data into unique sources [10]. *Active Probing* uses the Trinocular technique [21]. *Network Telescope* processes Merit Network Telescope traffic to indicate Internet background radiation [22]. *BGP* converts BGP tables from several collectors into a measure of network visibility. *Google* shows the

normalized service usage reported by the Google Transparency Report [23]. *Cloudflare* is a Content Delivery Network (CDN), its Radar platform exposes global performance and quality statistics [24]. *NetFlows* displays the saturation of the CDN edge, both overall and HTTP-specific saturation [25]. *Internet Quality Index* (IQI) presents several key Internet performance metrics that influence the user experience, such as bandwidth, latency, and DNS response time [26].

4.2. Model configurations

The ADTK and scikit-learn models are mostly kept default, configuring some parameters to align wih the Internet measurement data. For example, the seasonal window covers a day, which is the length of a period, similar to some IODA heuristic detectors. Several multivariate detectors can be tuned by changing the number of components. The framework generates models with varying numbers of components, ranging from one to the maximum number of data sources. This allows these models to capture patterns with low and high dimensionality. Table 1 summarizes the configuration parameters of the detectors.

Table 1. Configuration parameters for univariate and multivariate anomaly detectors.

Category	Model	Parameter	Value / Logic
Outlier	Threshold	Max / Min threshold	1.0 / 0.95
	Quantile	Max / Min quantile	0.95 / 0.05
	InterQuartileRange	c (IQR multiplier)	1.5
	GeneralizedESDTest	Significance (α)	0.05
Shift & Spike	Persist, LevelShift	Time window	20 hours
	VolatilityShift	Time window	5 hours
Seasonality	Seasonal	Frequency	24
	Autoregression	Step size / steps	24 / 1
Multivariate	Outlier	Detection model	LocalOutlierFactor, IsolationForest, EllipticEnvelope, OneClassSVM
	Cluster detctors	Components	[1, *len*(`features`)]
	LocalOutlierFactor	Neighbors	24
	RegressionAD	Targets	Active Probe, Network Telescope, BGP, Google, NetFlows, IQI

5. Evaluation and discussion

This section summarizes and discusses the results of the reports generated by the evaluation framework. Through the analysis of the cases, several findings can be highlighted.

5.1. Matching univariate detectors to data sources

By running all *univariate detectors* on all cases, it becomes apparent that detectors have *similar performance* with *data sources* that have *comparable characteristics*. Investigating the predictions reveals that many anomalies appear through local extrema, as lower values indicate lower performance or utility. Hence, it is evident that specific matches emerge; for example, detectors that can identify seasonal patterns perform well on seasonal data sources.

Based on these results, it is possible to hand-pick a *set of tailored detectors* for data sources with similar characteristics. This selection process is substantiated by domain knowledge, choosing detectors based on the inherent traits and expected noise levels of specific Internet measurements. For the IQI data sources, the *quantile detector* is recommended. The *interquartile range detector* shows good performance for the seasonal data sources. All other data sources are matched to the *persist detector* as it achieves well overall. These recommendations prioritize *robust* and *above-average performance* across all cases. These criteria result, for example, in the exclusion of seasonal detectors for seasonal data, as these detectors yield poor results in certain cases, especially with limited input data. Additionally, it is possible to define a set of best-performing detectors to evaluate against.

5.2. Multivariate approaches are well-performing black boxes

Multivariate approaches show promising results for detecting anomalies in high-dimensional data. These approaches can be broadly categorized into plug-and-play and parameter-tuned detectors. IsolationForest and EllipticEnvelope are examples of plug-and-play detectors that perform quite well in all cases. OneClassSVM is similarly plug-and-play, but achieves lower F1 scores due to a higher number of false positives.

The F2 scores prioritize recall over precision. Since anomalies might not entirely fall within the ground-truth label, a detector overshoot might not be entirely undesirable. For example, to catch the surge after an outage. It can be favorable to catch as many anomalies as possible, even if this results in a number of false positives. The plug-and-play detectors perform similarly or better on the F2 metric.

Parameter-tuned detectors can perform exceptionally well given a specific parameter set; however, these parameters may not be applicable to other cases. Additionally, in the Jordan case, with few disruptions, both univariate and multivariate detectors perform similarly poorly. Table 2 summarizes the results for several multivariate detectors.

Table 2. F1 and F2 scores for several multivariate approaches, for each case. Parameters are shown in square brackets. The standard deviation is shown for models that exhibit variance across runs.

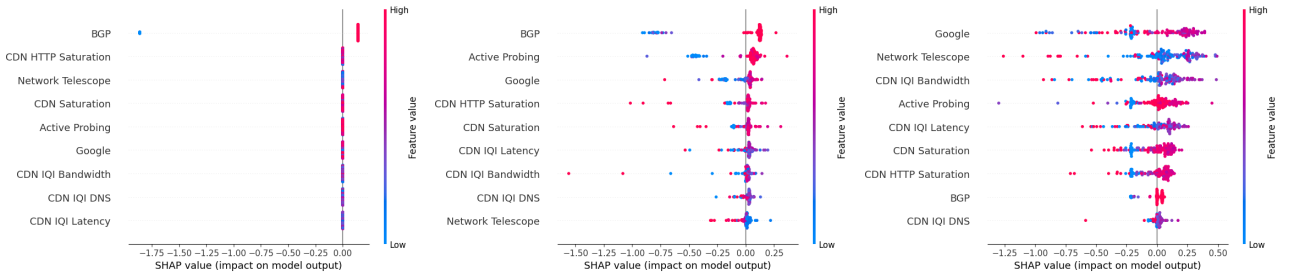
F1	**Elliptic Envelope**	**Isolation Forest**	**One Class SVM**	**Cluster[2]**	**Cluster[6]**	**PCA[3]**	**PCA[6]**	**Regression[BGP]**
Iraq	0.52±.02	0.56±.03	0.15	0.13±.16	0.87±.23	0.17	0.48	0.94
Syria	1.00±.00	0.86±.02	0.48	1.00±.00	0.40±.32	0.10	0.00	0.33
Algeria	0.64±.00	0.55±.05	0.35	0.13±.04	0.02±.09	0.67	0.00	0.00
Jordan	0.21±.00	0.23±.04	0.10	0.09±.00	0.18±.09	0.24	0.44	0.44
F2								
Iraq	0.69±.02	0.76±.02	0.31	0.23±.19	0.85±.21	0.14	0.58	0.91
Syria	1.00±.00	0.94±.01	0.70	1.00±.00	0.42±.25	0.08	0.00	0.56
Algeria	0.55±.00	0.51±.06	0.51	0.23±.08	0.02±.10	0.56	0.00	0.00
Jordan	0.36±.00	0.38±.04	0.21	0.16±.00	0.22±.11	0.29	0.56	0.37

Post-hoc explainability techniques can be applied to better understand detector results and mitigate the black box aspect of these models. Figure 2 presents the SHAP values for several plug-and-play multivariate detectors for the Syria case. The data sources are ranked according to their impact on the anomaly decision, highlighting their influence. Figure 2a shows that the EllipticEnvelope is almost exclusively influenced by the BGP data. This may explain why this detector achieves the highest F1 score and precision in some cases, as it focuses its detection efforts on the BGP data source, which clearly indicates all anomalies in this case. Figure 2b, in contrast, indicates that IsolationForest finds nuanced and subtle relationships between data sources, taking into account when certain measurements are high or low. While this results in a higher recall, the detector has a lower precision, resulting in more false positives. It mostly relies on the BGP and Active Probing data sources. Figure 2c reveals that the OneClassSVM detector also finds relationships between data sources; however, the influence of data sources is very different, resulting in similar recall with more false positives.

These SHAP values help understand the different models by giving insight into which data sources were considered. For example, the OneClassSVM detector places a high consideration on the seasonal and daily Google data, which reveals disruptions, albeit less clearly than the BGP data. This could explain why the detector is less precise, as there might be a larger gray area to consider, and help account for the higher number of false positives.

5.3. Robust detections through the tailored ensemble approach

For each aggregation strategy, the evaluation framework runs the proposed ensemble approach with the set of tailored base detectors for all cases. A performance ceiling is established for the proposed approach through a reference evaluation using impractical but ideal base detectors, the best-performing detectors for each use case. Not unexpectedly, the best detectors usually outperform the tailored detectors, both in precision and recall. There is no variance between runs, since the base detectors have none. However, the tailored detectors still achieve acceptable performance,



a. b. c.

Fig. 2. SHAP feature importance summaries, for the Syria case, illustrating the feature impact on the model output. (a) Elliptic Envelope (F1=1.00, F2=1.00, P=1.00, R=1.00); (b) Isolation Forest (F1=0.85±.02, F2=0.94±.01, P=0.75±.02, R=1.00); (c) One Class SVM (F1=0.48, F2=0.70, P=0.31, R=1.00).

especially compared to the equally plug-and-play multivariate detectors. Table 3 summarizes the F1 and F2 scores of the aggregation strategies for the proposed tailored ensemble approach.

Table 3. F1 and F2 scores for the proposed tailored ensemble detectors, across all cases. These scores can be compared to those in Table 2.

F1	**Weighted Voting**	**Weak Majority Voting**	**Majority Voting**	**Time Window Voting**	**Unanimous Voting**
Iraq	0.51	0.85	0.85	0.79	0.00
Syria	0.71	0.87	0.87	0.92	0.00
Algeria	0.67	0.29	0.00	0.00	0.00
Jordan	0.17	0.00	0.00	0.00	0.00
F2					
Iraq	0.73	0.87	0.87	0.85	0.00
Syria	0.86	0.83	0.83	0.93	0.00
Algeria	0.79	0.27	0.00	0.00	0.00
Jordan	0.43	0.00	0.00	0.00	0.00

The strategies that require a majority vote with equal weights exhibit inconsistent performance, sometimes performing exceptionally well, as seen in the time window voting for the Syria case. In other cases, a majority is never reached, resulting in very poor results. The weighted voting strategy, however, exploits domain knowledge to achieve more consistent performance. IODA data sources are given more weight as they show anomalies most clearly. In the case of Iraq and Syria, the weighted detector can achieve a higher recall at the cost of lower precision, which is reflected in the F2 score. For the Algeria and Jordan cases, the performance is considerably improved.

These results demonstrate that robust performance can be achieved through an ensemble anomaly detector with tailored base detectors. However, it does require an appropriate aggregation strategy. Taking into account domain knowledge on the reliability and clarity of the data sources is what gives the weighted voting strategy an edge in overall performance.

5.4. Base detector explainability

The ensemble approach exposes the individual base detectors, facilitating detector explainability. Figure 3 presents a visualization inspired by SHAP, delivering similar insights directly drawn from base detector predictions. An important distinction to make is that this visualization only shows predicted anomalies. Marks indicate the measurement value and the contribution of a feature to the prediction, which is why the importance is non-negative. Whereas SHAP shows positive and negative impact for any measurement. The data sources are ranked by their influence.

Figure 3a and 3b highlight that, for the Syria case, the BGP data source is most influential with majority and weighted voting. Additionally, since most high-importance marks have a blue tint, it indicates that a lower value triggers an anomaly. This aligns with the mental model of disruptions for those data sources, where a disruption

results in lower utility or performance. Likewise, low-importance marks for those data sources have a high value. A low-importance mark indicates that a prediction was made, but this data source did not contribute to the acceptance.

With the proposed ensemble approach, it is possible to effortlessly examine the predictions. Additionally, it can expose the potentially complex relationships that exist. This provides an explanation of the predictions, which can instill confidence in the approach. These insights are operationally similar to those available through the more computationally intensive SHAP.

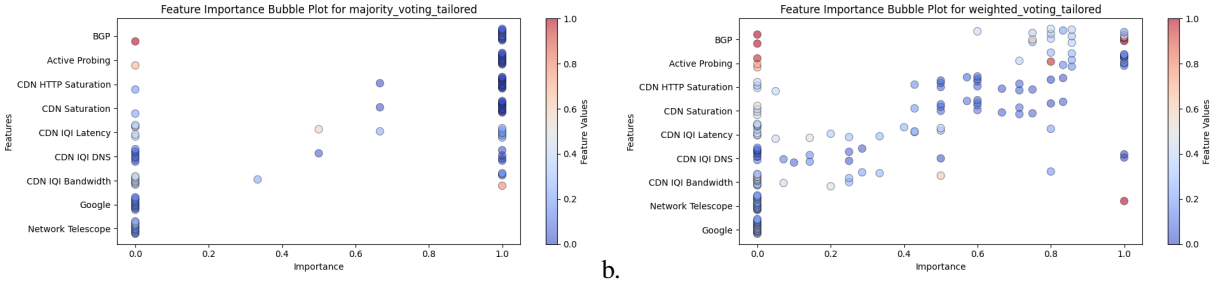


Fig. 3. SHAP-inspired feature importance for the proposed ensemble approach, for the Syria case, illustrating the impact of features and the effect of different voting strategies. (a) Majority Voting (F1=0.87, F2=0.83, P=0.96, R=0.80); (b) Weighted Voting (F1=0.71, F2=0.86, P=0.56, R=1.00).

6. Conclusion

This paper evaluates an ensemble anomaly detection approach that uses tailored base detectors for Internet measurements, demonstrating robustness, high performance, and inherent explainability. The evaluation framework revealed that no univariate detector performs universally well; however, data sources with similar characteristics respond predictably to specific detectors, motivating a tailored ensemble approach. Additionally, the proposed approach was empirically compared against a variety of traditional multivariate approaches. While this method may not outperform every individual multivariate detector in every single case, it consistently demonstrates robust and competitive performance. The proposed approach is similarly plug-and-play compared to several other multivariate methods, as it replaces fine-tuning with the application of domain knowledge. Furthermore, the tailored ensemble models achieve performance on par with impractical variants built from the best-performing univariate detector for each data source, validating our approach.

Beyond its robust performance, the proposed method has inherent explainability. Traditional multivariate models require computationally intensive post-hoc techniques, such as SHAP, for interpretation. In contrast, the ensemble approach provides immediate and intuitive insights into its predictions by revealing the contributing base detectors and their corresponding features. This transparency is crucial for building trust and enabling rapid root-cause analysis.

Future work can take several paths. First, advanced and adaptive aggregation strategies can dynamically adjust to the characteristics of data sources. Second, applying multiple base detectors to a single data source could potentially capture a wider range of anomalous behaviors, with a weighted aggregation strategy that leverages these multiple perspectives. Finally, the inherent explainability of this approach presents a unique opportunity to distill the distinctive fingerprints of various disruptions, which could be used to detect and classify the root causes of future disruptions.

Acknowledgments

Mike Vandersanden (BOF22OWB17) is a Ph.D. candidate at Hasselt University, supported by the Special Research Fund (BOF). This project was made possible through the Internet Society Pulse Research Fellowship.

Declaration of generative AI and AI-assisted technologies

The authors acknowledge the use of Grammarly and Gemini for minor editing, structuring, and proofreading of the paper. The content has been reviewed and edited as needed, and the authors take full responsibility.

References

[1] E. Aben, A Deep Dive Into the Baltic Sea Cable Cuts (Dec. 2024).
URL https://labs.ripe.net/author/emileaben/a-deep-dive-into-the-baltic-sea-cable-cuts/

[2] R. Padmanabhan, A. Schulman, D. Levin, N. Spring, Residential links under the weather, in: Proceedings of the ACM Special Interest Group on Data Communication, SIGCOMM '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 145–158. doi:10.1145/3341302.3342084.

[3] A. Chatzivasileiou, A. Kornilakis, K. Lionta, G. Nomikos, X. Dimitropoulos, G. Smaragdakis, How Russia's Invasion of Ukraine Impacted the Internet Peering of the Conflicted Countries, in: 2024 8th Network Traffic Measurement and Analysis Conference (TMA), 2024, pp. 1–10. doi:10.23919/TMA62044.2024.10559142.

[4] Cloudflare Radar, What we know about Iran's Internet shutdown (Jan. 2026).
URL https://blog.cloudflare.com/iran-protests-internet-shutdown/

[5] G. Aceto, A. Botta, P. Marchetta, V. Persico, A. Pescapé, A comprehensive survey on internet outages, Journal of Network and Computer Applications 113 (2018) 36–63. doi:10.1016/j.jnca.2018.03.026.

[6] CAIDA, Internet Outage Detection and Analysis (IODA), section: projects (Aug. 2016).
URL https://www.caida.org/projects/ioda/

[7] Internet Health Report, Internet Health Report | Monitoring networks health (2025).
URL https://ihr.live/

[8] R. S. Raman, A. Virkud, S. Laplante, V. Fortuna, R. Ensafi, Advancing the Art of Censorship Data Analysis, Free and Open Communications on the Internet (2023).
URL https://petsymposium.org/foci/2023/foci-2023-0003.php

[9] P. Sermpezis, L. Prehn, S. Kostoglou, M. Flores, A. Vakali, E. Aben, Bias in Internet Measurement Platforms, in: 2023 7th Network Traffic Measurement and Analysis Conference (TMA), 2023, pp. 1–10. doi:10.23919/TMA58422.2023.10198985.

[10] IODA, IODA - Help (2025).
URL https://ioda.inetintel.cc.gatech.edu/help

[11] M. Gao, R. Mok, E. Carisimo, E. Li, S. Kulkarni, k. claffy, DarkSim: A similarity-based time-series analytic framework for darknet traffic, in: Proceedings of the 2024 ACM on Internet Measurement Conference, IMC '24, Association for Computing Machinery, New York, NY, USA, 2024, pp. 241–258. doi:10.1145/3646547.3688426.

[12] A. Guillot, R. Fontugne, P. Winter, P. Merindol, A. King, A. Dainotti, C. Pelsser, Chocolatine: Outage Detection for Internet Background Radiation, in: 2019 Network Traffic Measurement and Analysis Conference (TMA), 2019, pp. 1–8. doi:10.23919/TMA.2019.8784607.

[13] P. Richter, R. Padmanabhan, N. Spring, A. Berger, D. Clark, Advancing the Art of Internet Edge Outage Detection, in: Proceedings of the Internet Measurement Conference 2018, IMC '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 350–363. doi:10.1145/3278532.3278563.

[14] J. Vanerio, P. Casas, Ensemble-learning Approaches for Network Security and Anomaly Detection, in: Proceedings of the Workshop on Big Data Analytics and Machine Learning for Data Communication Networks, Big-DAMA '17, Association for Computing Machinery, New York, NY, USA, 2017, pp. 1–6. doi:10.1145/3098593.3098594.

[15] T. Rojat, R. Puget, D. Filliat, J. D. Ser, R. Gelin, N. Díaz-Rodríguez, Explainable Artificial Intelligence (XAI) on TimeSeries Data: A Survey, arXiv:2104.00950 [cs] (Apr. 2021). doi:10.48550/arXiv.2104.00950.

[16] Arundo Analytics, Inc., arundo/adtk: A Python toolkit for rule-based, unsupervised anomaly detection in time series (2020).
URL https://github.com/arundo/adtk

[17] scikit-learn, scikit-learn: machine learning in Python — scikit-learn 1.7.1 documentation (2025).
URL https://scikit-learn.org/stable/index.html

[18] M. Vandersanden, J. Beerts, Explainable anomaly detection for internet measurements (2026). doi:10.5281/zenodo.18376516.

[19] M. Vandersanden, Results: Tailored ensemble anomaly detection for internet disruptions (2026). doi:10.5281/zenodo.18376560.

[20] H. Kreitem, Stop Exam-related Internet Shutdowns in 2024 (Apr. 2024).
URL https://pulse.internetsociety.org/blog/stop-exam-related-internet-shutdowns-in-2024

[21] L. Quan, J. Heidemann, Y. Pradkin, Trinocular: understanding internet reliability through adaptive probing, SIGCOMM Comput. Commun. Rev. 43 (4) (2013) 255–266. doi:10.1145/2534169.2486017.

[22] K. Benson, A. Dainotti, k. claffy, A. C. Snoeren, M. Kallitsis, Leveraging Internet Background Radiation for Opportunistic Network Analysis, in: Proceedings of the 2015 Internet Measurement Conference, IMC '15, Association for Computing Machinery, New York, NY, USA, 2015, pp. 423–436. doi:10.1145/2815675.2815702.

[23] Google, Google Transparency Report (2025).
URL https://transparencyreport.google.com/

[24] Cloudflare Radar, Worldwide Overview | Cloudflare Radar (Jan. 2025).
URL https://radar.cloudflare.com/

[25] Cloudflare, NetFlows · Cloudflare Radar docs (Feb. 2025).
URL https://developers.cloudflare.com/radar/investigate/netflows/

[26] Cloudflare, Cloudflare API | Radar > Quality > IQI (2025).
URL https://developers.cloudflare.com/api/go/resources/radar/subresources/quality/subresources/iqi/