

Article

Data-Mining-Based Detector Calibrations for High-Energy Physics

Melinda Orosz ^{1,2,*}  and Balázs Ujvári ^{2,3,*} ¹ Doctoral School of Physics, University of Debrecen, 4032 Debrecen, Hungary² HUN-REN Institute for Nuclear Research, 4026 Debrecen, Hungary³ Department of Data Science and Visualization, Faculty of Informatics, University of Debrecen, 4032 Debrecen, Hungary

* Correspondence: orosz.melinda@inf.unideb.hu or orosz.melinda@atomki.hu (M.O.); ujvari.balazs@inf.unideb.hu (B.U.)

† These authors contributed equally to this work.

Abstract

The generation of Dead Hot Maps (DHMs) is an important part of PHENIX data analysis, as it helps reduce detector-related biases. During this step, detector units with abnormal behavior are identified, such as dead, hot, or extremely hot channels. Removing these channels from the analysis prevents detector effects from changing the measured collision distributions. This helps ensure that the final data sample represents the underlying collision physics more reliably. In this work, the standard DHM construction procedure is compared with anomaly detection methods that are commonly used in data analysis and data mining. These methods can offer simpler and faster alternatives to the conventional approach, with lower computational demand and less need for manual work. Five techniques were tested: Z-score analysis, Local Outlier Factor (LOF), One-Class Support Vector Machine (OCSVM), Isolation Forest (IF), and Kernel Density Estimation (KDE). The main goal was to examine which method gives a detector mask that is closest to the established statistical reference and is also physically reasonable. The results suggest that these anomaly detection approaches may be useful for detector monitoring and calibration tasks. Some of the tested methods are easy to apply, computationally efficient, and do not require special hardware. Because of this, they could be included in future automated detector quality control workflows. Their simple implementation also makes them suitable for use in different software environments. In this way, these methods may support faster and more consistent identification of anomalous detector towers.

Keywords: big data analytics; High-Energy Physics (HEP); detector calibration; Dead Hot Map (DHM); statistical analysis tools; data mining; anomaly detection



Academic Editor: Chunping Li

Received: 6 July 2026

Revised: 10 August 2026

Accepted: 16 August 2026

Published: 7 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The reliability of high-energy physics measurements depends on the precise calibration of detectors, which is also crucial for quark–gluon plasma (QGP) research conducted in the PHENIX experiment [1,2]. Mapping the properties of QGP relies heavily on data from the electromagnetic calorimeter (EMCal), particularly through the rigorously controlled measurement of photons and neutral pions (π^0). However, the calorimeter’s extremely fine segmentation and complex architecture present unique challenges to researchers, requiring the use of specialized calibration procedures to obtain reliable data [3].

The PHENIX electromagnetic calorimeter has eight sectors, and each sector has many towers (PHENIX at RHIC, Figure 1). These towers may not work properly, which could lead to errors in the analysis [4].

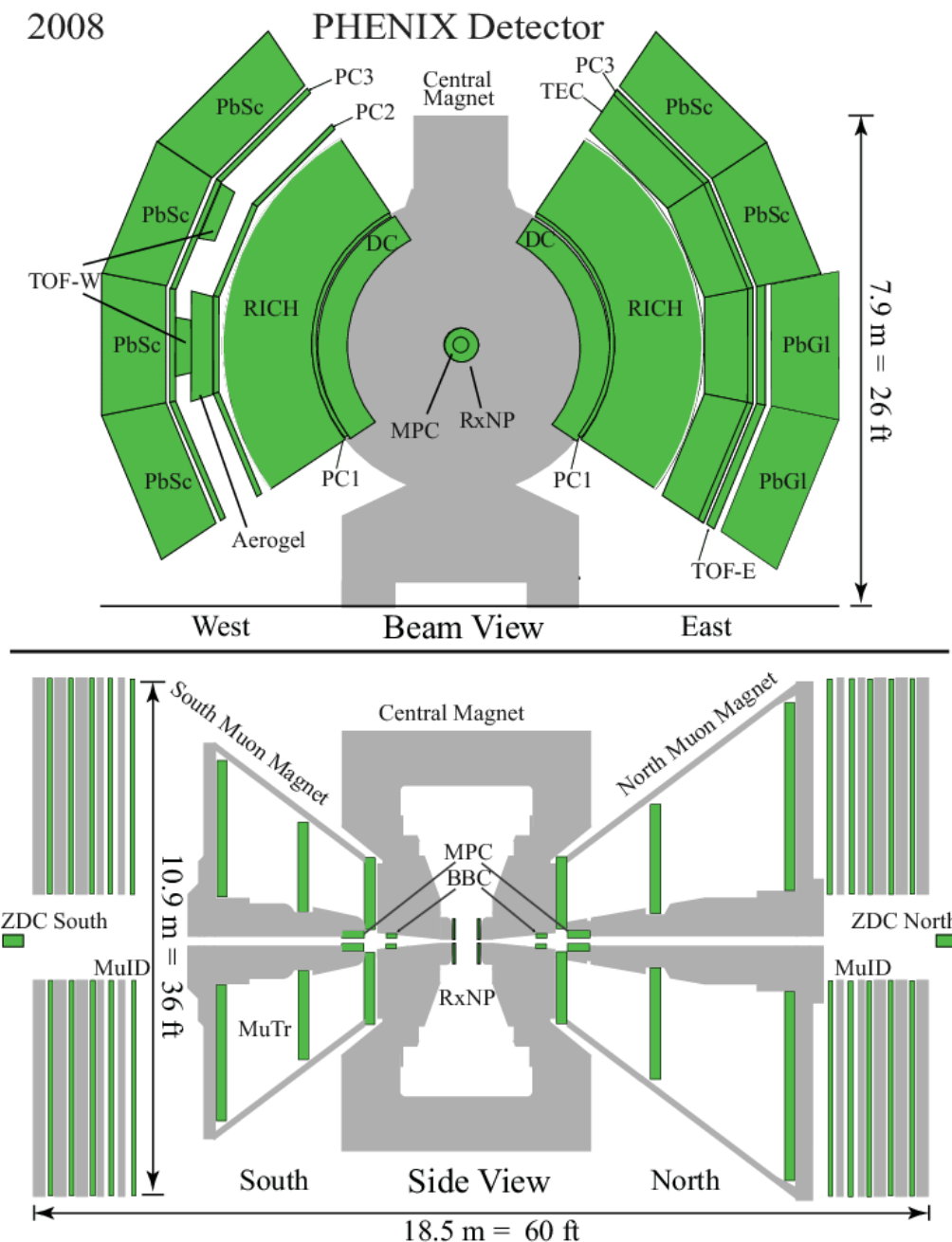


Figure 1. The PHENIX spectrometer setup viewed both from the beam direction and the side [5].

The purpose of the DHM (Dead Hot Map) is to filter out faulty calorimeter units (towers): to ensure the accuracy of the analysis, both “dead” towers—which do not emit a signal—and “hot” towers—which produce a signal even without actual energy absorption—must be removed. This task can essentially be interpreted as the detection of anomalies within the data. Their identification is based on statistics: we assume that, on average, over a large number of events, the energy distribution should be uniform across the towers, so significant deviations indicate a malfunction. However, determining statistical significance is complicated by the rapidly decaying energy spectrum and the inhomogeneous nature of the faults. Researchers are gradually refining the DHM settings to achieve a balance between preserving data quality and minimizing unnecessary exclusions. By examining the stability of the spectrum’s shape through different configurations, they have ensured that only truly faulty towers are excluded, preserving sensor sensitivity and reducing data loss.

One practical limitation of the DHM method is that it involves an extremely complex calibration and quality assurance process that considers multiple parameters. The identification of “dead” and “hot” towers is not based on a single general condition. Rather, it is based on several interdependent threshold values and statistical assumptions. Examples include effects dependent on the sensor area and validation steps. As a result, optimizing DHM settings requires significant time, computational resources, and expert supervision; it is not a trivial task. The calibration and validation processes are time-consuming and resource-intensive, and can be costly.

2. Related Work

Data mining and anomaly detection based on artificial intelligence have previously been studied in similar detector monitoring contexts. An autoencoder-based anomaly detection system was developed for the CMS electromagnetic calorimeter to achieve Data Quality Monitoring in real time. The method follows a semi-supervised approach. It tries to learn the detector’s normal response from high-quality data and then identifies any deviations from that normal behavior. The system uses both spatial variations and time-dependent evolution of anomalies. Its performance was validated on CMS collision data from 2018 and 2022, and first deployment results from Run 3 demonstrated that the system could detect anomalies missed by conventional monitoring tools [6].

Related studies have also been carried out for the CMS hadron calorimeter. Spatio-temporal anomaly detection models, including graph-based and autoencoder-based architectures, have been investigated using calorimeter occupancy maps [7]. These studies explicitly considered abnormal channel behaviors such as dead, degrading, noisy-hot, and fully-hot channels. Transfer-learning-based approaches were also proposed to improve robustness and reduce training requirements in large detector systems with many readout channels [8]. These works demonstrate that AI-based anomaly detection can be effective for identifying localized and time-dependent calorimeter faults.

Although these studies show the potential of AI-based anomaly detection for calorimeter monitoring, they are mostly designed for general online DQM or detector-status monitoring tasks in CMS-like detector environments. They do not directly address the PHENIX Run-16 Au + Au DHM construction problem, nor do they provide a systematic comparison between several lightweight anomaly detection algorithms and the conventional PHENIX DHM output on the same dataset. Therefore, the present work fills this gap by applying anomaly detection directly to the DHM analysis workflow.

Deep-learning models can learn complex spatial and temporal detector patterns, but they generally require large training datasets, substantial memory, and additional preprocessing and post-processing steps. In the referenced CMS autoencoder implementation, several consecutive time windows had to be combined to reduce false-positive detections, which increased the detection latency. Such computational and deployment requirements are particularly important for complex calorimeter systems [9] and motivate the investigation of lightweight anomaly detection methods that can operate with fewer resources and simpler processing workflows.

3. Our Contribution

Many of the existing approaches rely on relatively complex deep-learning architectures and are designed mainly for online DQM or general detector-status monitoring. In contrast, the present work focuses on lightweight and computationally less demanding anomaly detection methods that can be applied directly to the DHM construction workflow. This is important because a practical DHM alternative should not introduce excessive computational cost or require large training infrastructures. Instead, it should remain simple enough

to be integrated into existing calibration pipelines while still producing a physically reliable detector mask.

Our contribution is to compare and apply an anomaly detection approach for the DHM analysis on PHENIX Run-16 Au + Au that is faster and more efficient, while still providing results comparable to those of the traditional method. To achieve this, the conventional DHM procedure was implemented and used as a reference for comparison. In addition, five different anomaly detection algorithms were implemented and evaluated: Z-score analysis, Local Outlier Factor (LOF), One-Class Support Vector Machine (OCSVM), Isolation Forest (IF), and Kernel Density Estimation (KDE). By applying all methods to the same initial dataset, their ability to identify faulty calorimeter units could be directly compared, allowing us to assess whether AI-based and statistical anomaly detection techniques can serve as reliable alternatives to the traditional DHM workflow.

The conventional DHM was used as the reference anomaly map because its performance had previously been validated through detector-level and subsequent physics analyses, and it was shown to provide physically consistent results. Establishing whether a complete detector mask is reliable requires extensive evaluation, including checks of the corrected energy distributions and the effects of masking on downstream reconstruction. Such validation had already been performed for the conventional DHM, allowing it to serve as a practical reference for evaluating the anomaly detection methods. However, the traditional DHM construction and validation workflow is computationally demanding, time-consuming, and difficult to parallelize in some processing steps. The objective of the present study was therefore to determine whether a simpler and more computationally efficient anomaly detection method could reproduce the behavior of this previously validated reference mask.

To the best of our knowledge, only limited studies are publicly available on AI-based anomaly detection applied directly to the PHENIX Run-16 Au + Au DHM construction problem. Most existing PHENIX-related studies focus on conventional statistical DHM construction or timing calibration rather than on systematic comparisons with anomaly detection algorithms. This lack of directly comparable studies motivates the present work, since the identification of dead and hot EMCal towers is naturally interpretable as an anomaly detection problem.

Although related deep-learning-based anomaly-detection systems have already been developed for existing detector-monitoring infrastructures, the present work is motivated by a different deployment scenario. Our objective is not only to analyze detector data offline, but also to investigate methods that could eventually be incorporated at a lower level of the data-acquisition and calibration chain. In such an environment, computational resources, memory, latency, and hardware compatibility, particularly with FPGA-based processing, represent important constraints. Complex deep-learning models may provide strong detection performance, but their computational and implementation requirements could place a substantial load on the detector system and make real-time or near-real-time integration more difficult.

For this reason, the investigated methods were selected as lightweight alternatives that require less computational capacity and are potentially more suitable for low-level implementation in future detector systems. Unlike approaches developed primarily as additional monitoring layers for already-existing infrastructures, our long-term aim is to support anomaly detection closer to the detector readout and preprocessing stages. The present study does not demonstrate a complete FPGA implementation. Rather, it evaluates whether comparatively simple statistical and machine-learning methods can reproduce the behavior of the previously validated DHM while remaining compatible with the resource and latency requirements expected in future online detector-calibration workflows.

4. Preliminaries

4.1. Data Mining

Data mining is an interdisciplinary scientific field concerned with the analysis of large datasets for the extraction of implicit, previously unknown, non-trivial, and potentially useful patterns [10,11]. Its objective extends beyond descriptive summarization by identifying hidden regularities, dependencies, clusters, and anomalies that can support decision-making, prediction, and knowledge-based inference. It is commonly regarded as a core operational stage of the broader Knowledge Discovery in Databases (KDD) process, which includes data selection, cleaning, integration, and transformation; pattern extraction; and the evaluation and interpretation of results [12]. Within this framework, data mining specifically denotes the application of analytical and computational methods that transform raw data into structured, interpretable, and practically relevant knowledge [13].

A central goal of data mining is that systematic analysis can reveal knowledge that is not directly observable in raw data. In large and representative datasets, patterns may emerge with greater statistical stability; consequently, the quality, size, noise level, and information content of the data may be as decisive as the complexity of the applied algorithm. Data mining is therefore particularly suited to the analysis of high-dimensional, noisy, incomplete, heterogeneous, and partially unstructured datasets, where semi-automatic or automatic methods are required to discover complex and previously unknown patterns [12].

Although data mining is closely related to machine learning, the two concepts are not identical. Machine learning provides many of the principal methodological tools used in data mining, especially for classification, regression, clustering, and anomaly detection. Data mining, however, represents a broader workflow that also includes preprocessing, transformation, validation, and interpretation, while machine learning primarily supports pattern recognition and predictive modeling within this process [13].

4.1.1. Anomaly Detection

One important sub-task of data mining and machine learning is the identification of abnormal or outlying observations, commonly referred to as anomaly detection. An outlier is a data point, object, or set of examples that deviates significantly from the rest of the dataset [14]. These deviations are important because they can mean rare, critical, or unusual patterns. For example, in financial systems, an unusual transaction can mean malicious activity. In an industrial setting, a malfunctioning sensor reading can mean equipment failure.

Various anomaly detection tasks can be formalized depending on the goal of the process. These may select anomalous observations, rank observations based on their abnormality, or evaluate a new data point [14]. Let

$$D = \{x_1, x_2, \dots, x_m\}$$

denote a database consisting of records. An anomaly score

$$s : D \rightarrow \mathbb{R}$$

is assigned to each record $x \in D$, where $s(x)$ is the degree to which example deviates from the normal patterns in the dataset. Given a predefined threshold t , the set of anomalies can then be defined as

$$A = \{x \in D \mid s(x) > t\}.$$

In another formulation, no fixed threshold is specified. Instead, the records are ranked according to their anomaly scores, and the n observations with the highest values of $s(x)$ are considered anomalous [14]. A third formulation concerns the evaluation of a new, previously unseen test point with respect to a reference dataset that predominantly contains normal observations. In this case, the task is to determine how well the new data point conforms to the known normal behavior and to what extent it deviates from it.

Anomaly detection is often treated as an unsupervised learning problem, since in practical applications, labeled information indicating which observations are normal and which are abnormal is frequently unavailable [12]. As a consequence, the validation of anomaly detection results is challenging, similarly to clustering tasks. Another important characteristic of the problem is that datasets usually contain substantially more normal observations than anomalous ones. This class imbalance makes it difficult to detect rare but potentially important events or erroneous values [14].

The general methodological principle of anomaly detection is based on constructing a model or profile of normal behavior. This profile may take the form of a statistical summary, a probability distribution, a cluster structure, a rule system, or a machine learning model [13]. Once such a normal profile is established, each observation can be evaluated to determine how it differs from the expected behavior. We classify a tower an anomaly if its value does not fit well into the normal behavior of the model.

4.1.2. Anomaly Detection in the Construction of the Dead Hot Map

Applying anomaly detection in the DHM comes from the nature of the problem. The goal of the DHM is to identify towers whose responses significantly differ from normal sensor behavior. Such differences can appear as low activity, complete inactivity, extremely high activity, or noisy signals. Since these anomalous towers can mislead the photon and π^0 reconstruction, and increase systematic uncertainties, their reliable identification is essential for accurate EMCal calibration. In this context, anomaly detection methods provide data-driven tools for identifying normal tower behavior.

4.2. Data Preparation

The data went through several filtering phases to ensure that only high-quality runs were included. These filtering procedures can be divided into run-level and event-level filters.

4.2.1. Run-Level Selection Criteria

A run is considered suitable for analysis only if the detector is operating stably and provides sufficient statistical data.

- **Statistical Threshold:** We exclude runs containing fewer than 1 million recorded events to ensure that each tower has sufficient data for a reliable statistical analysis.
- **Operational Stability:** Runs affected by hardware malfunctions, beam instabilities, or missing detector sectors are excluded based on the experiment's logbooks.
- **Long-Term Consistency:** Only “good long runs” are selected, where the detector conditions remained stable throughout the entire data acquisition period.

4.2.2. Event-Level Quality Filtering

In addition to ensuring global stability, we also performed meticulous checks to prevent local or temporary detector errors from distorting the DHM:

- **Zero-Hit Towers:** A run is removed if there are more than 100 “zero-hit” towers in any sector. A high concentration of inactive units in a given sector typically indicates a major electronic failure or high-voltage (HV) issues.
- **Cluster-Per-Event Distribution:** To ensure data integrity, runs are deleted if fewer than 10 clusters are recorded on average per event in any sector.
 - This threshold ensures that the analysis is not biased by inefficient or partially non-functional detector elements.
 - The stability is monitored across all eight separate sectors to account for the different responses of the PbSc and PbGl calorimeter types.

4.2.3. Pseudorapidity Correction

Due to the detector’s geometry, our Hit Map, and thus the distribution it generates, will be minimally distorted. As a result, the towers at the edges cover a smaller solid angle, so the number of hits will be lower, creating an angle dependence. We call this angle dependence the pseudorapidity correction, or η -correction for short. This normalization process works as follows: we sum the hits along the z-axis of the Hit Map, which gives us a distribution, and we use this distribution to normalize back to 1. This makes the error detection algorithm more accurate in the subsequent steps (Figure 2).

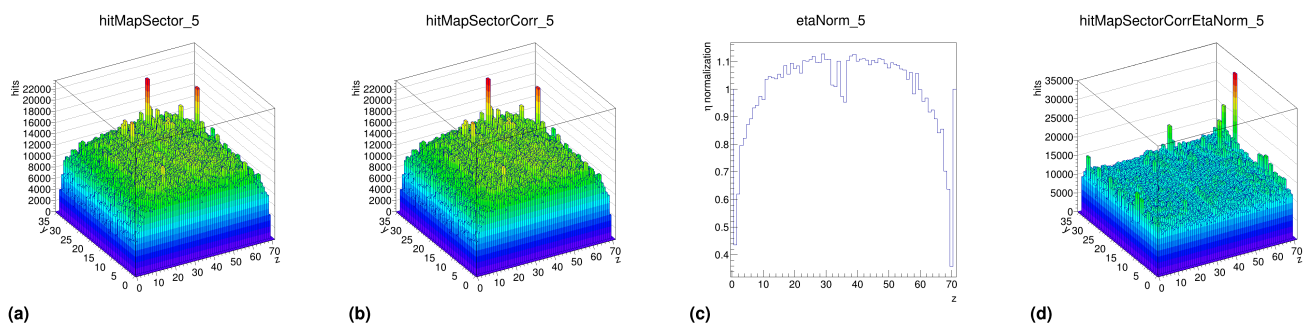


Figure 2. The raw hit distribution (a), removing the extreme large hits (b), the corrected raw hit distribution. The η normalization as the z-projection of the (c) and the η normalized corrected raw hit distribution (d) in energy: 0.4–0.5 GeV. The height of each tower corresponds to the number of hits recorded in the respective towers, which are color-coded for enhanced clarity.

4.3. Algorithms

Original DHM Method

The DHM is a fundamental calibration framework used in high-energy physics experiments to ensure data integrity. For Runs 14 and 16 of the PHENIX experiment Au + Au collisions at $\sqrt{s_{NN}} = 200$ GeV, the DHM algorithm served as the primary tool for filtering out malfunctioning EMCal towers, while ensuring the long-term preservation of the dataset. Following a similar methodology based on previous protocols, this work also implements a multi-step approach. The fundamental mechanisms and analytical steps of the procedure are detailed in the following sections.

Run14 and Run16 have a similar problem in the beginning of the data analysis. Over- or under-working towers need to be excluded.

The Raw Hit Map in Figure 3 provides a direct visualization of the detector’s initial response.

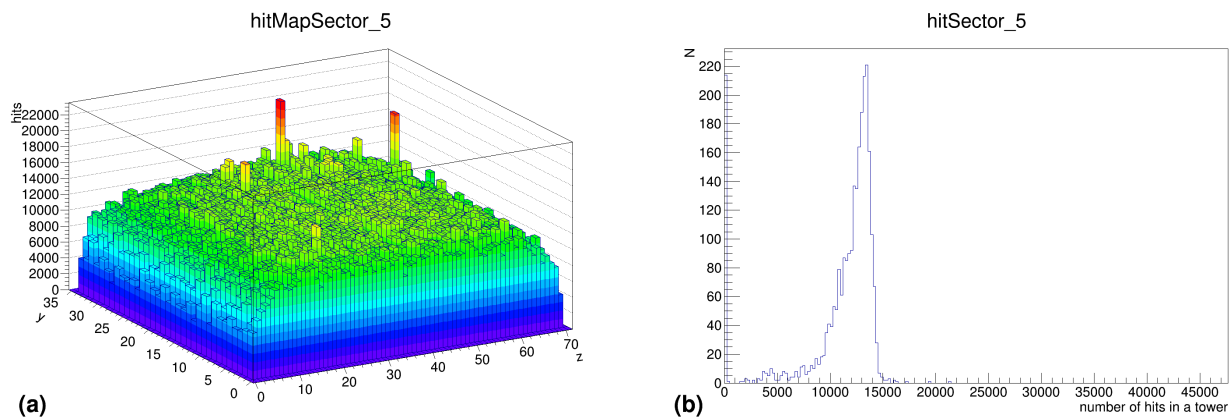


Figure 3. Raw Hit map (a) and Hit Distribution (b) for sector 5 in energy: 0.4–0.5 GeV. The height of each tower corresponds to the number of hits recorded in the respective towers, which are color-coded for enhanced clarity.

4.4. Statistical Analysis of the Number of Hits

The measurements cover a very wide energy range (0–30 GeV), which cannot be treated uniformly due to physical constraints. As energy increases, the hit count tends to decrease; therefore, the distributions derived from the Hit Map cannot be described by the same statistical model across all ranges. For this reason, we employed a three-method framework to identify faulty towers:

1. **Gaussian Distribution (low-energy bins):** The Gaussian distribution is already a well-established method for describing the distribution of low-energy event counts. Typically, the data here form a bell-shaped, symmetric curve. In the low-energy range (0–1 GeV), the Gaussian distribution provided a perfect fit to the hit count distribution (Figure 4a). We used the mean value of the distribution to identify dead towers, selecting $Mean \times 10\%$ (marked by the blue line), while the σ value of the distribution provided the thresholds for hot and extra hot towers. We identified hot towers at values above 6σ (marked by the red line), and extra hot towers at values above 100σ .
2. **Poisson distribution (medium-energy bins):** At higher energies, the Gaussian distribution does not necessarily describe the shape of the hit rate distribution well, so we had to switch to the Poisson distribution here (1–5 GeV) (Figure 4b). At these energies, the occurrence of clusters decreases, and we treat the hits as independent events. We use statistical fluctuations to establish the threshold value for hot towers. We identified hot towers at values above 6σ (marked by the red line).
3. **Binomial distribution (high-energy bins):** At even higher energies, the hits appear completely randomly in the detector, thus exhibiting a “heavy-tailed” pattern, which we were able to represent using a binomial distribution (Figure 4c). For the cutoff of the hot towers, we used the intersection point of a linear fit on the logarithmic scale of the Hit Distribution (marked by the red line).

We specifically optimized the chosen framework for the observed energy intervals. The essence of the method is that each of the three distribution types used (Gaussian, Poisson, and binomial) models a specific, physically well-distinguishable aspect of the hit statistics within the entire 30 GeV range.

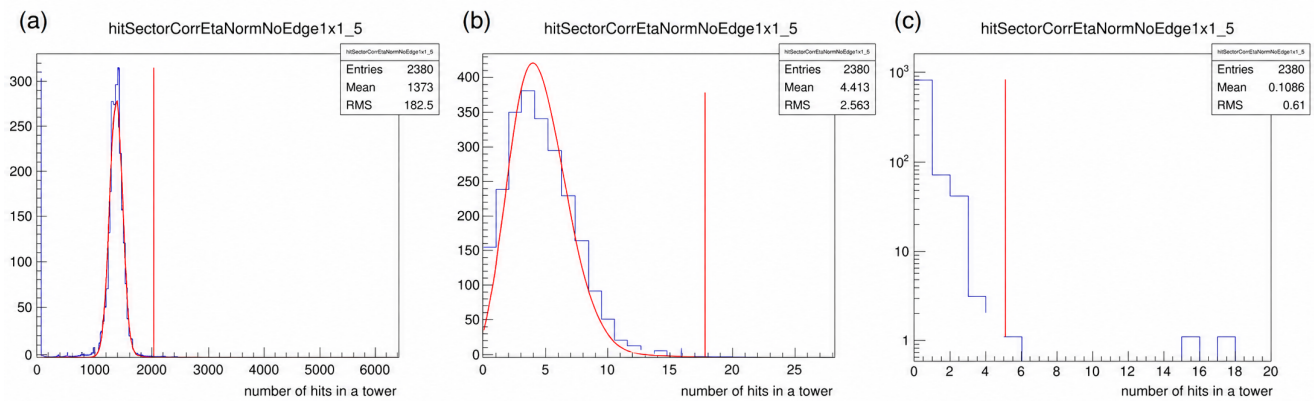


Figure 4. The three distribution type in different energy ranges. (a) 0.4–0.5 GeV, (b) 2.0–3.0 GeV, (c) 5.0–6.0 GeV.

4.5. Step 4: Create Dead Hot Map

After identifying the dead, hot, and extra hot towers, we needed to create a global DHM. To do this, we first had to handle the individual runs and energy intervals separately (Figure 5), and then, using appropriate quality control criteria, we merged them both at the run level and at the energy range level.

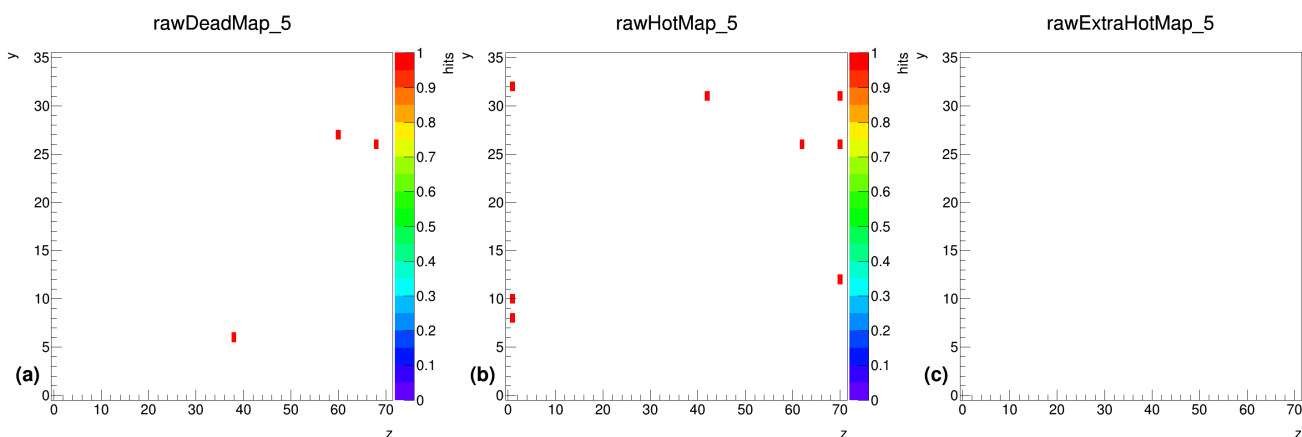


Figure 5. Run-level raw dead (a) hot (b) and extra hot map (c) for the 0.4–0.5 GeV energy range.

After applying the cuts mentioned in Sections 4.2.1 and 4.2.2, we combined the remaining runs, and we called it Selected DHM. This included the results of all good runs. However, as mentioned in Section 4.4, we had to work with different energies for each run. In the final phase, we aimed to obtain a Dead Hot Map covering the range from 0.05 GeV to 30 GeV, for which the data processed in the 39 separate energy bins also had to be combined. After merging the energy bins, we created the so-called Grass-level cut, which helped us show in percentage terms which towers consistently malfunction and which ones malfunction in only 1–2% of cases (Figures 6 and 7). This Grass-level % cut gave us the Global DHM.

The final DHM was the end result of this complex statistical parameterization and filtering process. This allowed us to obtain a final image for each sector individually, while treating the nearly 600 runs and the entire energy spectrum as a single unit.

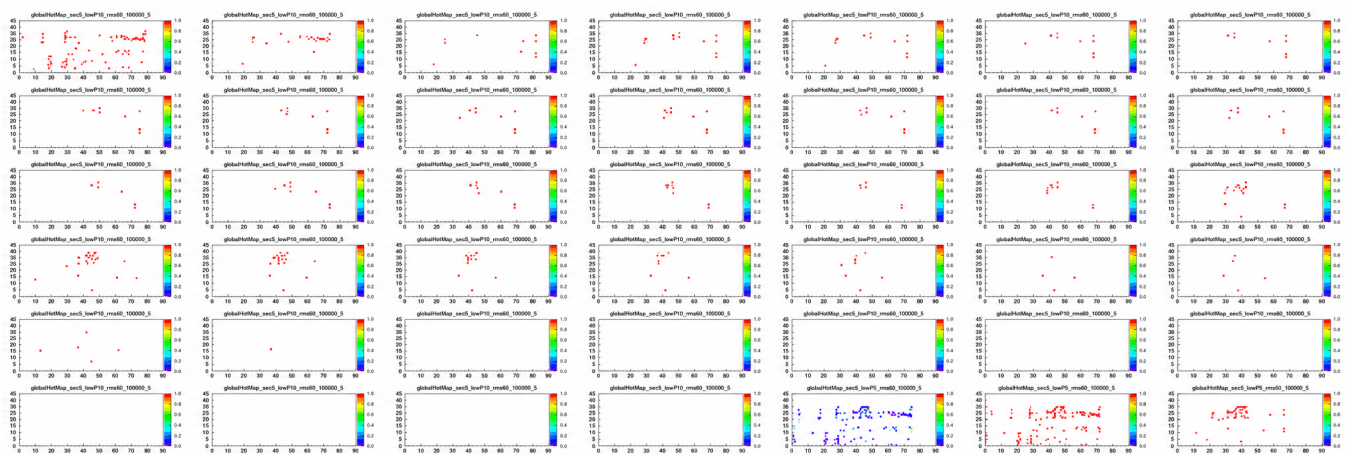


Figure 6. The 38 energy bins exhibit differences, and the last three panels show the following: the 38th panel is all the energy bins merged together, where the colors show how many times the flagged tower was wrong. The 39th panel shows if we applied the Grass-level 0% cut, which means we did not provide any chances for a tower to show a malfunction. The 40th panel shows if we applied the Grass-level 5%.

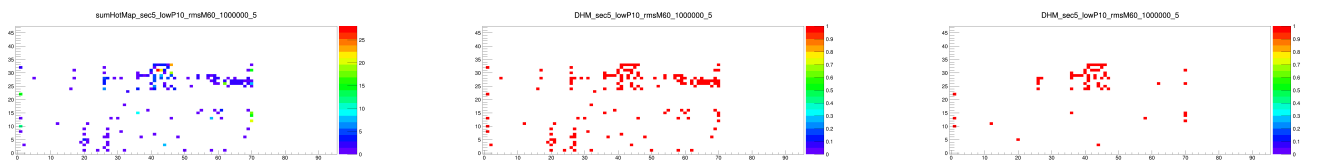


Figure 7. The last three panels zoomed in. The left panel is all the energy bins merged together, where the colors show how many times the tower was flagged wrong. The middle panel shows if we applied the Grass-level 0% cut, which means we did not provide any chances for a tower to show a malfunction. The right panel shows if we applied the Grass-level 5%.

4.5.1. Z-Score

The Z-score is one of the simplest statistical methods for detecting outliers. It measures how much a given observation deviates from the mean of the dataset, expressed in units of the standard deviation. For a one-dimensional observation, x_i , the Z-score is defined as follows:

$$z_i = \frac{x_i - \mu}{\sigma},$$

where μ denotes the mean of the dataset and σ denotes its standard deviation. Observations with large absolute Z-scores are considered potential anomalies [14]. A general decision rule can be:

$$|z_i| > \tau,$$

where τ is a predefined threshold value, which is often $\tau = 2$ or $\tau = 3$.

In the context of DHM construction, the Z-score can be used to identify calorimeter towers whose hit counts deviate significantly from the average tower response. Its main advantage is its low computational complexity and straightforward interpretation. However, the method assumes that the data are approximately normally distributed and is sensitive to extreme values, since outliers may affect both the estimated mean and standard deviation [14].

4.5.2. Local Outlier Factor

The Local Outlier Factor (LOF) is a density-based anomaly detection method. Instead of comparing each data point to the global distribution of the entire dataset, LOF evaluates how isolated a point is with respect to its local neighborhood. This makes the method

suitable for datasets in which normal observations may form regions of different local densities [12,15].

Let $N_k(x)$ denote the set of k -nearest neighbors of a point x . The k -distance of x is the distance to its k -th nearest neighbor. The reachability distance between two points x and o is defined as

$$\text{reachdist}_k(x, o) = \max\{k\text{-distance}(o), d(x, o)\},$$

where $d(x, o)$ is the distance between x and o . The local reachability density of x is then given by

$$\text{lrd}_k(x) = \left(\frac{1}{|N_k(x)|} \sum_{o \in N_k(x)} \text{reachdist}_k(x, o) \right)^{-1}.$$

The LOF score is defined as

$$\text{LOF}_k(x) = \frac{1}{|N_k(x)|} \sum_{o \in N_k(x)} \frac{\text{lrd}_k(o)}{\text{lrd}_k(x)}.$$

If $\text{LOF}_k(x) \approx 1$, the point has a density similar to its neighbors. Values that are significantly larger than one indicate that the point is in a lower-density region and considered as an anomaly [15].

For DHM analysis, LOF is useful because a faulty tower does not necessarily have to be globally extreme. It can be abnormal compared with neighboring or similarly behaving towers. However, the method is sensitive to the choice of the number of neighbors k , and its performance may decrease in high-dimensional spaces where distance measures become less informative [12].

4.5.3. One-Class Support Vector Machine

The One-Class Support Vector Machine (OCSVM) is a model-based anomaly detection method. It aims to learn the boundary of the normal data distribution and identifies observations outside this boundary as anomalies. Unlike supervised SVM classification, OCSVM is trained primarily on data assumed to represent normal behavior [16].

Given training samples $x_1, x_2, \dots, x_n$, the OCSVM solves the following optimization problem:

$$\min_{w, \rho, \xi} \frac{1}{2} \|w\|^2 + \frac{1}{\nu n} \sum_{i=1}^n \xi_i - \rho,$$

subject to

$$w \cdot \phi(x_i) \geq \rho - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, n.$$

Here, $\phi(x_i)$ is a feature-space mapping, w is the normal vector of the separating hyperplane, ρ is the offset, ξ_i are slack variables, and $\nu \in (0, 1]$ controls the expected fraction of outliers and the number of support vectors [17].

The decision function is

$$f(x) = \sum_{i=1}^n \alpha_i K(x_i, x) - \rho.$$

where $K(x_i, x)$ is the kernel function and α_i are the learned coefficients. A point is classified as normal if

$$f(x) \geq 0,$$

and anomalous if

$$f(x) < 0.$$

The predicted class label is obtained by applying the sign function to the decision function:

$$\hat{y} = \text{sign}(f(x)).$$

Thus, $\hat{y} = +1$ indicates a normal observation, while $\hat{y} = -1$ indicates an anomaly.

The OCSVM is advantageous because nonlinear detector-response structures can be handled through kernel functions, for example, the radial basis function kernel [17],

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right).$$

However, the method is sensitive to parameter choices, particularly the kernel type, the kernel width γ , and the parameter ν . In the DHM context, this sensitivity must be considered carefully because small parameter changes may affect which calorimeter towers are classified as faulty.

4.5.4. Isolation Forest

Isolation Forest is a tree-based unsupervised anomaly detection method. Its main assumption is that anomalies are rare and easier to isolate than normal observations. The algorithm constructs an ensemble of random isolation trees by recursively selecting a feature and a random split value. Points that can be isolated with fewer splits are considered more anomalous [18].

Let $h(x)$ denote the path length of a point x , defined as the number of edges required to isolate x in an isolation tree. The average path length over an ensemble of trees is denoted by $E[h(x)]$. The anomaly score is defined as

$$s(x, n) = 2^{-\frac{E[h(x)]}{c(n)}},$$

where n is the sample size and $c(n)$ is the average path length of an unsuccessful search in a binary search tree [18]:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n},$$

where $H(i)$ is the i -th harmonic number, which can be approximated as

$$H(i) \approx \ln(i) + \gamma,$$

with γ denoting the Euler–Mascheroni constant.

The interpretation of the anomaly score is as follows:

$$s(x, n) \rightarrow 1 \quad \Rightarrow \quad x \text{ is likely anomalous,}$$

$$s(x, n) \ll 0.5 \quad \Rightarrow \quad x \text{ is likely normal.}$$

Isolation Forest is computationally efficient because it does not require the computation of a full pairwise distance matrix [14]. It is therefore suitable for processing large datasets from detectors. In the DHM workflow, this property is particularly important because the method may be applied across many runs, detector sectors, and energy intervals. However, its performance may decrease if anomalous towers form structured groups rather than isolated points.

4.5.5. Kernel Density Estimation

Kernel Density Estimation (KDE) is a non-parametric density estimation method. It estimates the probability density function of the data without assuming a predefined

parametric form [19]. Given observations $x_1, x_2, \dots, x_n$, the KDE estimate of the density at a point x is

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right),$$

where $K(\cdot)$ is the kernel function and $h > 0$ is the bandwidth parameter. In the multivariate case, for d -dimensional observations, the estimator can be written as

$$\hat{f}_h(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right).$$

A commonly used kernel is the Gaussian kernel [19],

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right).$$

In anomaly detection, observations located in regions of low estimated density are considered potential anomalies. This can be expressed using a density threshold ϵ [14]:

$$x_i \in A \quad \text{if} \quad \hat{f}_h(x_i) < \epsilon,$$

where A denotes the set of anomalous observations.

For DHM analysis, KDE can be used to estimate the most probable range of normal tower behavior. Towers located in low-density regions of the estimated distribution can then be marked as potential dead or hot channels. The method is flexible because it does not require the hit-count distribution to follow a Gaussian, Poisson, or binomial model. However, KDE is sensitive to the bandwidth parameter h : a too small bandwidth may produce a noisy density estimate, while a too large bandwidth may over-smooth the distribution and hide relevant anomalies [19].

5. Results

5.1. DHM Anomaly Map

The rigorous selection process used in compiling the final DHM (Figure 8) was designed to ensure that only the highest-quality data were included in the analysis. The following criteria were used during the classification and screening process:

- Minimum Events (MEs) = 1,000,000.
- Grass-Level (GL) = 5%.
- Tower classification:
 - Dead towers: Low cluster per event (<10%).
 - Hot towers: Towers exceeding 6σ above mean activity.
 - Extra hot towers: Towers exceeding 100σ above mean activity.

The final DHM (Figure 8) is the result of this rigorous quality control process, which guarantees the stability and reliability of the data used. By removing the faulty towers, we minimized data corruption produced by malfunctioning instruments. On the accompanying color-coded map, red marks units that consistently malfunctioned during measurements; masking these is essential for precise energy measurement. This selection procedure directly improves the quality of the calorimeter data, which increases the accuracy of photon reconstruction and π^0 measurements, thereby enhancing the scientific credibility of the experimental results. For the evaluation of the other algorithms, this was used as the baseline.

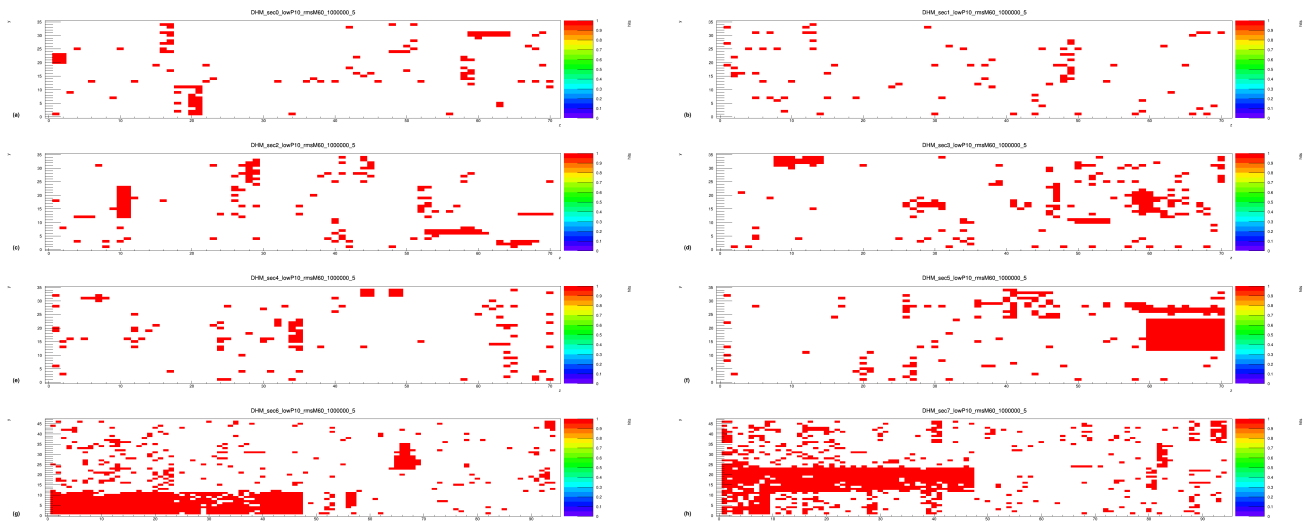


Figure 8. Final DHM with ME = 1,000,000, GL = 5%, LD = 0, LH = 0, LEH = 0 for all sectors in the Au + Au collision. Sector 0 (a), sector 1 (b), sector 2 (c), sector 3 (d), sector 4 (e), sector 5 (f), sector 6 (g), sector 7 (h).

5.2. Anomaly Detection Algorithms

The parameters of the investigated anomaly detection methods were determined using method-specific parameter-selection procedures. For algorithms requiring explicit hyperparameter selection, several candidate configurations were evaluated at the run level, and the resulting masks were compared with the conventional DHM reference. Parameter selection was based on the combined assessment of Intersection over Union, Dice/F1-score, precision, recall, false-positive rate, and false-negative rate. For KDE, the bandwidth was determined automatically from the statistical properties of the input data using Scott's rule. For the one-dimensional distributions considered in this study, the bandwidth factor is given by $(n^{-1/5})$, where (n) denotes the number of observations. Therefore, the KDE bandwidth adapts to the statistical properties of the analyzed distribution. No individual performance metric was optimized independently, since improving anomaly sensitivity may also increase the unnecessary masking of correctly operating towers. Therefore, the final parameter configurations were selected as the most favorable overall compromise between identifying faulty channels, reducing false detections, and preserving the usable detector area. The physical consistency, stability, and reproducibility of the masks across the investigated runs were also considered before the run-level results were combined into the final sector-level detector masks.

Following the run-level evaluation, the results obtained from different runs were combined at the sector level to construct a final detector mask. This mask was then intended to be used in the energy-dependent analysis to remove malfunctioning or anomalous detector towers from the dataset. The purpose of this procedure was to suppress detector-induced artifacts and to ensure that the subsequent physical analysis was based on a cleaner and more reliable detector response.

Among the investigated methods, Kernel Density Estimation (KDE) showed the closest agreement with the behavior of the original DHM algorithm as can be seen in Figure 9. This indicates that KDE was able to reproduce, at least approximately, the masking logic of the conventional procedure while relying on a data-driven estimate of the underlying detector-response distribution. In contrast, the other methods, namely Z-score analysis, LOF, OCSVM, and IF, were able to indicate certain trends and abnormal regions, but they did not successfully identify the real anomalies or they incorrectly detected anomalies, see the example in Figure 10, where the cluster energy distributions can be seen. Each row

shows a different method for sector 2 and sector 7. The cluster energy distribution was not successfully corrected by all other methods, except the KDE. This caused misidentification in the DHM, which can be seen in Figure 11. Again, each row shows a different method for sector 2 and sector 7. Figure 12 shows the ratio between the cluster energy distributions between the conventional DHM and the anomaly-based methods. Again, each row shows a different method for sector 2 and sector 7. Their weaker performance may also suggest that these algorithms would require larger, more representative datasets and more extensive parameter optimization to achieve comparable stability.

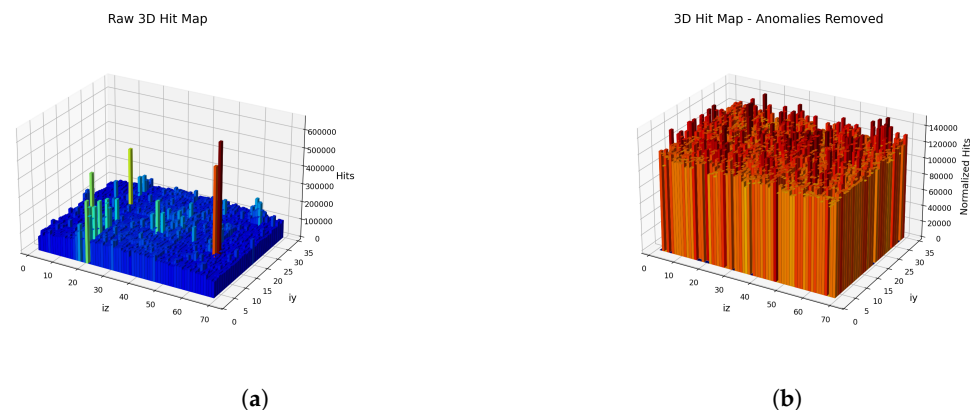


Figure 9. Comparison of the original detector hitmap and the hitmap obtained after removing the towers identified as anomalous by the applied masking procedure. **(a)** Original hitmap before applying the detector mask. **(b)** Hitmap after removing the anomalous detector towers using KDE method. The height of each tower corresponds to the number of hits recorded in the respective towers, which are color-coded for enhanced clarity.

The results were evaluated using multiple metrics and qualitative comparisons, with particular emphasis on physical consistency, reproducibility, and agreement with the established DHM baseline. This is essential because the task is not merely to identify mathematical outliers, but to detect detector towers whose behavior is physically inconsistent with the expected response of the detector system.

Table 1 summarizes the sector-wise mean performance of the five anomaly-detection algorithms relative to the conventional Dead Hot Map reference mask. Bold values indicate the best result within each sector. The metrics are appropriate for evaluating the anomaly-detection-based masks because the dataset is strongly imbalanced because correctly operating towers substantially outnumber faulty towers. Under these conditions, accuracy and specificity alone may be misleading, since an algorithm can obtain high values by classifying nearly every tower as normal while detecting very few actual anomalies. Precision and recall are therefore particularly important. Precision quantifies how many of the towers identified as anomalous are also marked as faulty by the reference Dead Hot Map, thereby characterizing the reliability of the generated alarms. Recall measures how many of the reference faulty towers are successfully detected and therefore reflects the ability of the method to avoid leaving malfunctioning channels in the dataset. Intersection over Union and the Dice/F1-score provide a more comprehensive evaluation of the predicted detector mask because they jointly account for correctly detected anomalies, missed faulty towers, and incorrectly excluded normal towers. These metrics are especially relevant for the present task, where the spatial agreement between the anomaly-based mask and the conventional Dead Hot Map is important. The false-positive rate measures the unnecessary removal of correctly operating detector channels, which reduces the usable detector area, whereas the false-negative rate quantifies the proportion of faulty towers that remain unmasked and may distort subsequent energy reconstruction. Consequently, the combined

evaluation of overlap-based metrics, precision, recall, false-positive rate, and false-negative rate provides a more physically meaningful assessment than accuracy alone, as the objective is not merely to classify most towers correctly, but to produce a reliable detector mask that suppresses detector-induced artifacts while preserving valid measurements.

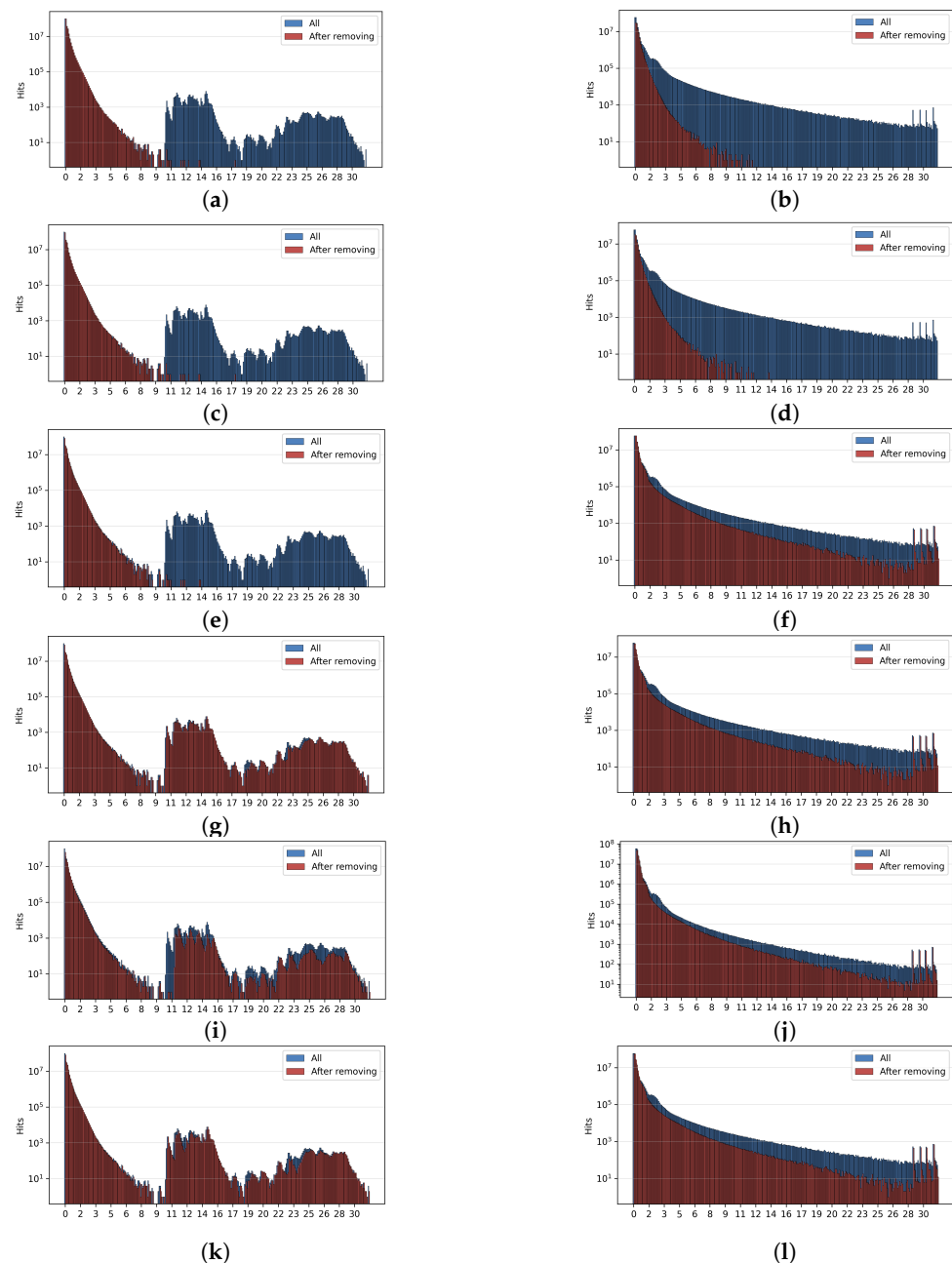


Figure 10. Comparison of the conventional DHM's and the anomaly-detection-based DHM's effect on the cluster energy distributions for sectors 2 and 7. The left column shows sector 2 and the right column shows sector 7. From top to bottom, the rows correspond to the original DHM, KDE, Z-score, OCSVM, LOF, and Isolation Forest. (a) Sector 2 cluster energy distribution, original DHM method. (b) Sector 7 cluster energy distribution, original DHM method. (c) Sector 2 cluster energy distribution, KDE anomaly-based DHM. (d) Sector 7 cluster energy distribution, KDE anomaly-based DHM. (e) Sector 2 cluster energy distribution, Z-score anomaly-based DHM. (f) Sector 7 cluster energy distribution, Z-score anomaly-based DHM. (g) Sector 2 cluster energy distribution, OCSVM anomaly-based DHM. (h) Sector 7 cluster energy distribution, OCSVM anomaly-based DHM. (i) Sector 2 cluster energy distribution, LOF anomaly-based DHM. (j) Sector 7 cluster energy distribution, LOF anomaly-based DHM. (k) Sector 2 cluster energy distribution, IF anomaly-based DHM. (l) Sector 7 cluster energy distribution, IF anomaly-based DHM.

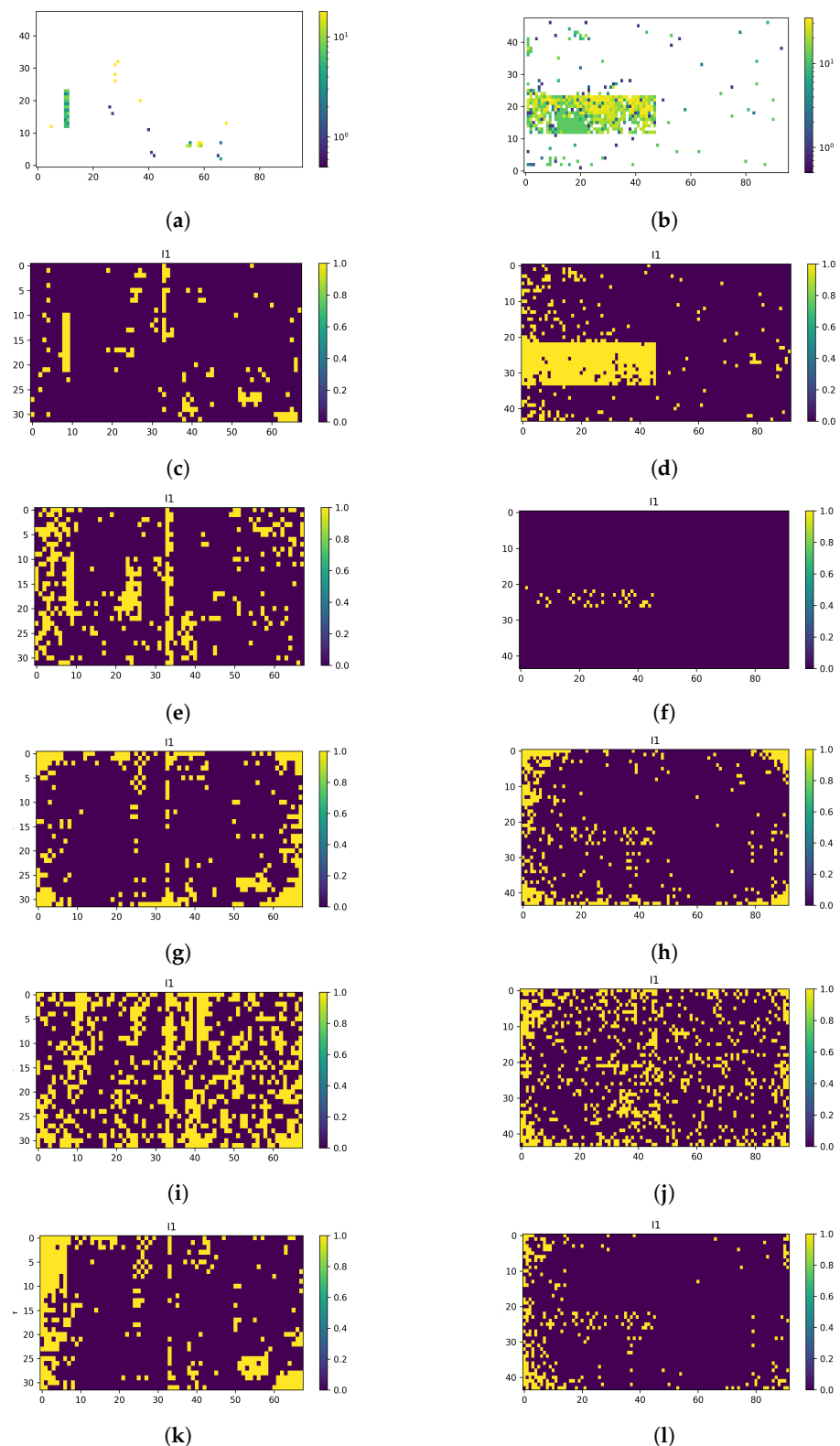


Figure 11. Comparison of the conventional DHM and the anomaly-detection-based DHMs for sectors 2 and 7. The left column shows sector 2 and the right column shows sector 7. From top to bottom, the rows correspond to the original DHM, KDE, Z-score, OCSVM, LOF, and Isolation Forest. (a) Sector 2, original DHM. (b) Sector 7, original DHM. (c) Sector 2, KDE anomaly-based DHM. (d) Sector 7, KDE anomaly-based DHM. (e) Sector 2, Z-score anomaly-based DHM. (f) Sector 7, Z-score anomaly-based DHM. (g) Sector 2, OCSVM anomaly-based DHM. (h) Sector 7, OCSVM anomaly-based DHM. (i) Sector 2, LOF anomaly-based DHM. (j) Sector 7, LOF anomaly-based DHM. (k) Sector 2, IF anomaly-based DHM. (l) Sector 7, IF anomaly-based DHM.

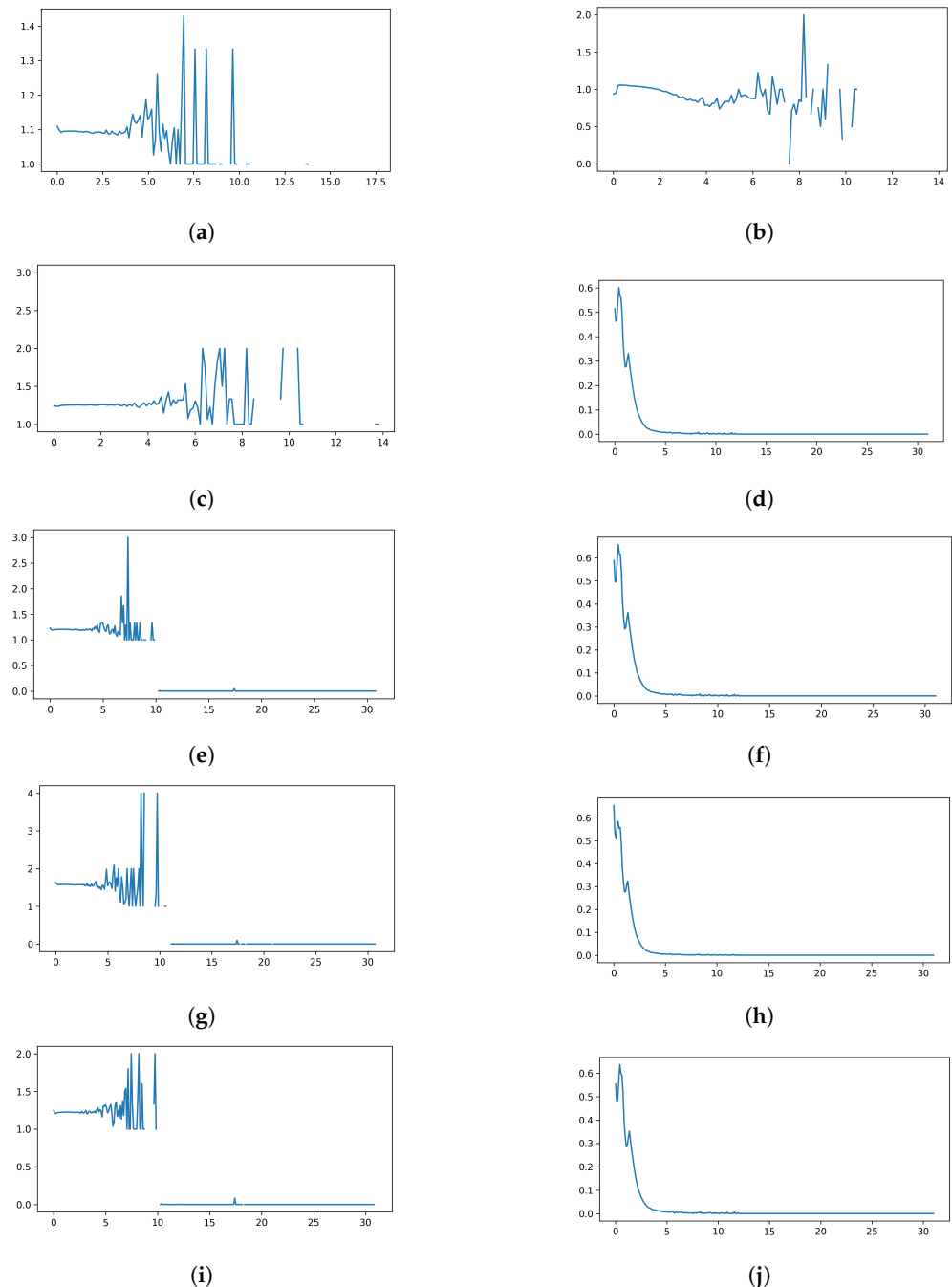


Figure 12. Ratio of the cluster energy distribution between the conventional DHM and the anomaly-detection-based DHMs for sectors 2 and 7. The left column shows sector 2 and the right column shows sector 7. From top to bottom, the rows correspond to the original divided with, KDE, Z-score, OCSVM, LOF, and Isolation Forest. (a) Sector 2, ratio of the cluster energy distribution, between the KDE anomaly-based DHM and the original DHM. (b) Sector 7, ratio of the cluster energy distribution, between the KDE anomaly-based DHM and the original DHM. (c) Sector 2, ratio of the cluster energy distribution, between the Z-score anomaly-based DHM and the original DHM. (d) Sector 7, ratio of the cluster energy distribution, between the Z-score anomaly-based DHM and the original DHM. (e) Sector 2, ratio of the cluster energy distribution, between the OCSVM anomaly-based DHM and the original DHM. (f) Sector 7, ratio of the cluster energy distribution, between the OCSVM anomaly-based DHM and the original DHM. (g) Sector 2, ratio of the cluster energy distribution, between the LOF anomaly-based DHM and the original DHM. (h) Sector 7, ratio of the cluster energy distribution, between the LOF anomaly-based DHM and the original DHM. (i) Sector 2, ratio of the cluster energy distribution, between the IF anomaly-based DHM and the original DHM. (j) Sector 7, ratio of the cluster energy distribution, between the IF anomaly-based DHM and the original DHM.

The lower IoU values observed in sectors 0–5 should be interpreted in the context of the detector structure. Sectors 0–5 belong to the PbSc calorimeter, whereas sectors 6–7 belong to the PbGl calorimeter. These two detector technologies have different response characteristics and hit-count distributions. Therefore, anomalous tower behavior does not necessarily appear in the same statistical form in all sectors. The higher recall and IoU values observed in sectors 6–7 suggest that normal and anomalous tower responses are more clearly separated in these sectors, making the anomalies easier to identify by the investigated methods. In contrast, the low IoU values in sectors 0–5 are mainly associated with low recall rather than with a high false-positive rate. In these sectors, several methods classify most towers as normal, resulting in high specificity and accuracy but limited overlap with the conventional DHM. Thus, the sector dependence should not be interpreted only as weak algorithmic performance, but also as a consequence of the different detector-response characteristics of the PbSc and PbGl sectors.

Importantly, the better overall performance of KDE is not determined only by sectors 6 and 7. KDE achieves the highest IoU and Dice/F1-score among the investigated methods in every sector, although the absolute agreement with the conventional DHM remains lower in some PbSc sectors. For example, in sector 3, KDE achieves an IoU of 0.3404, while the IoU values of the other methods remain close to zero. This shows that the relative advantage of KDE is also present in the PbSc sectors, even though the absolute performance is higher in sectors 6 and 7.

As shown in Table 1, KDE demonstrated the most robust overall performance among the evaluated methods. Its superiority is not based on a single metric, but on its consistently stronger balance across the metrics that are most relevant for an imbalanced anomaly-detection problem. In particular, KDE achieved the highest or among the highest Intersection over Union and Dice/F1-score values across the detector sectors, indicating that its predicted masks provided the closest overall agreement with the conventional Dead Hot Map while simultaneously accounting for both missed faulty towers and incorrectly masked normal channels. KDE also generally produced higher precision and recall, demonstrating that its detections were both more reliable and more complete than those of the competing methods.

Although KDE was not optimal for every individual metric, the isolated advantages of the other algorithms were typically obtained at the expense of substantially poorer performance elsewhere. The clearest example is sector 6, where IF achieved a slightly higher recall than KDE (0.6976 versus 0.6582); however, its false-positive rate was more than twice as high (0.1310 versus 0.0548), while its precision, Intersection over Union, and Dice/F1-score were all considerably lower. Thus, the additional anomalies detected by IF were accompanied by a much larger number of correctly operating towers being incorrectly removed. In contrast, KDE preserved a substantially larger fraction of valid detector channels while maintaining a comparably high anomaly-detection rate. Therefore, KDE represents the best overall method because it provides the most favorable and physically meaningful trade-off between anomaly sensitivity, detection reliability, mask agreement, and preservation of the usable detector area.

The superior performance of KDE is consistent with the statistical properties of the PHENIX electromagnetic calorimeter data. Since the hit-count distributions may exhibit Gaussian, Poisson-like, or long-tailed behavior depending on the energy range, they cannot be adequately described by a single parametric model. KDE estimates the probability density function directly from the observed data and identifies observations in low-density regions as potential anomalies. Its non-parametric nature allows it to represent asymmetric, multimodal, and locally distorted distributions, enabling it to reproduce the physically motivated masking behavior of the conventional Dead Hot Map more accurately than the other evaluated algorithms.

Table 1. Average algorithm results by sector. Bold values show the best result in each sector.

Algorithm	IoU	Dice	Precision	Recall/TPR	Specificity	Accuracy	FPR	FNR
Sector 0								
Z-score	0.0006	0.0012	0.0375	0.0006	0.9966	0.9698	0.0034	0.9994
SVM	0.0655	0.1221	0.2235	0.0953	0.9971	0.9710	0.0029	0.9047
LOF	0.0746	0.1380	0.2733	0.1039	0.9976	0.9715	0.0024	0.8961
IF	0.0467	0.0887	0.2000	0.0627	0.9979	0.9714	0.0021	0.9373
KDE	0.1639	0.2760	0.3441	0.2851	0.9951	0.9707	0.0049	0.7149
Sector 1								
Z-score	0.0000	0.0000	0.0000	0.0000	1.0000	0.9674	0.0000	1.0000
SVM	0.0000	0.0000	0.0000	0.0000	1.0000	0.9674	0.0000	1.0000
LOF	0.0000	0.0000	0.0000	0.0000	1.0000	0.9674	0.0000	1.0000
IF	0.0000	0.0000	0.0000	0.0000	1.0000	0.9674	0.0000	1.0000
KDE	0.0255	0.0496	0.1286	0.0410	0.9973	0.9652	0.0027	0.9590
Sector 2								
Z-score	0.0002	0.0004	0.0750	0.0002	0.9992	0.9509	0.0008	0.9998
SVM	0.0344	0.0664	0.2348	0.0428	0.9961	0.9490	0.0039	0.9572
LOF	0.0118	0.0233	0.4000	0.0121	0.9996	0.9515	0.0004	0.9879
IF	0.0522	0.0989	0.2607	0.0673	0.9953	0.9487	0.0047	0.9327
KDE	0.1237	0.2176	0.4963	0.1484	0.9969	0.9516	0.0031	0.8516
Sector 3								
Z-score	0.0007	0.0014	0.0588	0.0007	0.9964	0.9462	0.0036	0.9993
SVM	0.0003	0.0006	0.0500	0.0003	0.9983	0.9479	0.0017	0.9997
LOF	0.0106	0.0209	0.3750	0.0109	0.9994	0.9492	0.0006	0.9891
IF	0.0028	0.0055	0.1200	0.0029	0.9980	0.9478	0.0020	0.9971
KDE	0.3404	0.4953	0.7400	0.4062	0.9965	0.9557	0.0035	0.5938
Sector 4								
Z-score	0.0007	0.0014	0.0786	0.0007	0.9971	0.9537	0.0029	0.9993
SVM	0.0006	0.0011	0.0529	0.0006	0.9964	0.9529	0.0036	0.9994
LOF	0.0000	0.0000	0.0000	0.0000	1.0000	0.9562	0.0000	1.0000
IF	0.0006	0.0011	0.0643	0.0006	0.9971	0.9536	0.0029	0.9994
KDE	0.1764	0.2950	0.4437	0.2600	0.9961	0.9554	0.0039	0.7400
Sector 5								
Z-score	0.0002	0.0004	0.0500	0.0002	0.9991	0.9320	0.0009	0.9998
SVM	0.0340	0.0657	0.3167	0.0372	0.9952	0.9304	0.0048	0.9628
LOF	0.0084	0.0167	0.2875	0.0086	0.9987	0.9321	0.0013	0.9914
IF	0.0260	0.0506	0.2731	0.0282	0.9956	0.9303	0.0044	0.9718
KDE	0.0658	0.1231	0.5114	0.0710	0.9961	0.9330	0.0039	0.9290
Sector 6								
Z-score	0.0860	0.1507	0.1381	0.2237	0.6574	0.5858	0.3426	0.7763
SVM	0.1766	0.2997	0.2791	0.3804	0.8137	0.7311	0.1863	0.6196
LOF	0.0867	0.1589	0.1971	0.1545	0.8686	0.7435	0.1314	0.8455
IF	0.3729	0.5423	0.4808	0.6976	0.8690	0.8182	0.1310	0.3024
KDE	0.4774	0.6396	0.6728	0.6582	0.9452	0.8690	0.0548	0.3418
Sector 7								
Z-score	0.0582	0.1072	0.1125	0.1185	0.7494	0.6351	0.2506	0.8815
SVM	0.2145	0.3532	0.3413	0.3946	0.8382	0.7501	0.1618	0.6054
LOF	0.1639	0.2817	0.3210	0.2667	0.8796	0.7620	0.1204	0.7333
IF	0.3485	0.5164	0.4757	0.6052	0.8621	0.8032	0.1379	0.3948
KDE	0.5965	0.7401	0.7886	0.7314	0.9603	0.8985	0.0397	0.2686

This suggests that, in the DHM construction task, the relevant anomalies are not necessarily defined by a simple threshold, distance measure, or isolation criterion. Rather, they appear as deviations in the empirical density structure of the detector response. KDE is therefore an optimal solution for this problem. It does not depend on strict modeling assumptions. As a result, it tracks the actual distribution of tower activity in a more natural way.

The weaker performance of the Z-score can be attributed to its reliance on the mean and standard deviation. Although this method is effective for approximately normal distributions, its reliability decreases if the data are skewed or have long tails. The LOF method, which detects anomalies based on local density differences, can be sensitive to the choice of neighborhood parameters. If certain detector regions exhibit intrinsically different but physically valid activity levels, LOF can incorrectly classify normal towers as anomalies. The performance of OCSVM is strongly affected by the choice of kernel and by the values of the ν and γ parameters. If these parameters are not optimally tuned, the learned decision boundary may become either too restrictive or too permissive. In addition, OCSVM is less transparent from a physical interpretation perspective and can be sensitive to scaling, noise, and heterogeneity in the data. Consequently, it may identify statistically separated observations that do not correspond well to a physically justified DHM mask. The Isolation Forest method assumes that anomalies can be isolated efficiently through random partitioning. This assumption is appropriate when anomalies are rare and clearly separated from the normal population. However, if defective detector towers form a gradual transition between normal and abnormal behavior, or if the relevant deviation is primarily density-based, Isolation Forest may be less effective. Its random tree-based structure may also introduce additional variability in the resulting masks.

Overall, the results indicate that the most complex algorithm is not necessarily the optimal choice for DHM construction. Based on the present evaluation, KDE appears to be the most promising among the tested data-mining approaches, as it most closely approximates the behavior of the original DHM algorithm while preserving a flexible and empirically grounded description of the detector-response distribution.

5.3. DHM from the Anomaly Algorithms

For the final DHM, an additional filtering step was applied to remove any remaining noise components. Background noise removal was based on morphological background estimation and the intensity masking derived from it. As a first step, the slowly varying component of the heatmap was estimated using grayscale morphological opening. This operation suppresses smaller, locally intense structures while preserving the larger-scale, gradually varying background pattern. We further refined the resulting background estimate with mild Gaussian smoothing, which reduced local irregularities and resulted in a more stable background model.

After subtracting the estimated background, we obtained an intensity map that described the excess intensity of each point relative to the local background level. Based on this representation, we applied a percentile-based threshold determination to identify those areas that can be considered as relevant signal components relative to the background. For the first six sectors, we set the threshold at the 95th percentile; therefore, we retained only the top 5% of the background-corrected intensity values, while the bottom 95% were removed as background. For sectors 6 and 7, we applied a less stringent threshold corresponding to the 70th percentile, which means that we retained the strongest 30% of the values, while removing the lower-intensity 70%. This sector-specific thresholding strategy made it possible to preserve weaker but still relevant intensity patterns in sectors where the signal was less clearly separated from the background.

The binary mask obtained through thresholding was further refined using morphological cleaning operations. Opening removed small, isolated noise components, closing connected nearby regions and reduced minor discontinuities, while hole filling eliminated empty areas within the selected objects. As a result, a spatially more coherent mask with reduced sensitivity to noise was obtained. In the final step of background removal, the cleaned mask determined which parts of the original intensity distribution were retained. The pixels inside the mask retained their original intensity values, while the areas outside the mask, identified as background, were removed. The result was a cleaned final DHM that primarily contained signal areas elevated relative to the local background level and spatially connected. This can be seen in the Figures 13 and 14 for sector 2 in panel (a) and sector 7 in panel (b). The Figure 13 shows the result of the final DHM for the KDE based method, meanwhile the Figure 14 shows the result of the final DHM for the Z-score based method.

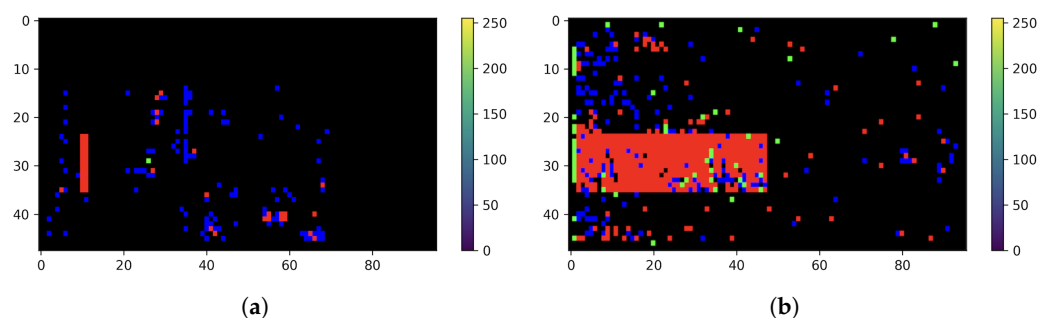


Figure 13. Comparison between the original DHM mask and the KDE-based mask for Sector 2 (a) and 7 (b). Red means towers identified by both methods, blue means towers detected only by the anomaly detection method, and green means towers marked only by the original DHM algorithm.

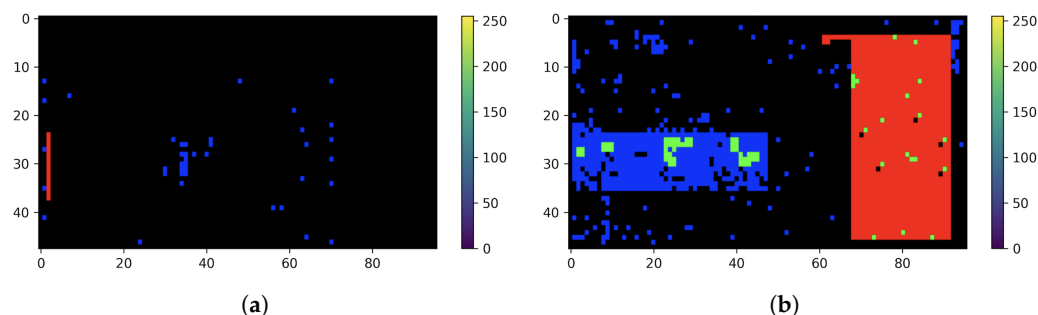


Figure 14. Comparison between the original DHM mask and the Z-score-based mask for Sector 2 (a) and 7 (b). Red means towers identified by both methods, blue means towers detected only by the anomaly detection method, and green means towers marked only by the original DHM algorithm.

First, the anomaly maps were constructed separately at the individual run level. After this run-level evaluation, the results obtained for each run were combined in order to produce an aggregated mask as the final DHM.

Further studies are needed to determine whether this level of accuracy is sufficient, particularly with regard to its impact on subsequent physical analyses. Nevertheless, the results suggest that, at the simulation level, the method functions adequately when applied to energy filtering. In many cases, it was able to identify trends similar to and consistent with those observed using the traditional, well-established DHM-based approach. This suggests that the method may be suitable as a complementary or preliminary screening tool in the detector mask fabrication workflow, although its definitive applicability requires further quantitative validation.

Each investigated anomaly detection method has specific limitations. The Z-score is most reliable for approximately Gaussian data and may perform poorly for skewed or

long-tailed distributions. LOF is sensitive to the neighborhood size and may interpret physically valid regional differences as anomalies. OCSVM depends strongly on feature scaling and the selected kernel parameters, while IF may be less effective when anomalies form structured groups or gradual transitions. The main limitation of KDE is bandwidth selection: insufficient smoothing increases noise sensitivity, whereas excessive smoothing may hide relevant anomalies.

6. Conclusions and Future Work

The use of accurate calibration procedures is essential for improving the reliability of physical analyses in high-energy physics experiments. In the PHENIX experiment, the calibration of the electromagnetic calorimeter (EMCal) with DHM-based methods helps achieve more precise event reconstruction and reduces systematic uncertainties. The DHM procedure identifies detector towers with abnormal responses, including dead, hot, and extra-hot towers. By excluding unreliable detector regions from the analysis, the method improves the quality of the recorded data.

These calibration improvements strengthen the consistency of the dataset used in PHENIX analyses. They also help reduce systematic errors and increase the reliability of the measured observables. In the future, similar methods could also be applied in other calorimetric experiments, where adaptive DHM classification may be useful for maintaining detector performance over longer operating periods.

A substantial practical difference was observed in the computational requirements of the two workflows. The conventional DHM production was executed using approximately 200 CPU cores on a high-performance computing system and required several weeks, whereas the investigated anomaly detection methods were executed on a standard laptop, with the longest complete workflow requiring no more than approximately one week. These observations indicate that the proposed methods can considerably reduce both the required computing infrastructure and the overall processing time, supporting their potential use in resource-constrained detector-monitoring and calibration environments.

Future work could focus on the development of a “live Dead Hot Map.” This approach would make it possible to identify anomalous detector towers during data collection or at an early reconstruction stage. Instead of relying only on detector masks produced after offline analysis, the detector mask could be updated continuously as detector conditions change. In this case, saving failed runs may no longer be necessary, which could reduce the required time, storage capacity, and energy consumption in the long term.

Further studies could also investigate hybrid approaches that combine KDE with simple threshold-based or distance-based criteria. Such methods may improve physical interpretability, robustness, and computational efficiency. They could provide a practical balance between fully data-driven anomaly detection methods and the traditional DHM procedure.

Although the present framework focuses on the evaluation of existing anomaly detection methods, several extensions could further improve its adaptability and physical interpretability. In particular, adaptive threshold optimization, physics-informed constraints, and detector-specific feature learning represent promising directions for future research. Such developments could incorporate detector geometry, temporal evolution, channel correlations, and calibration information into the anomaly detection process.

Author Contributions: Conceptualization, B.U. and M.O.; methodology, B.U. and M.O.; software, M.O.; validation, M.O.; formal analysis, M.O.; investigation, M.O.; resources, B.U.; data curation, M.O.; writing—original draft preparation, B.U. and M.O.; writing—review and editing, B.U. and M.O.; visualization, M.O.; supervision, B.U., project administration, B.U.; funding acquisition, B.U. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Research, Development and Innovation Office (NKFIH) OTKA 143540.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author or can be seen at <https://github.com/giraszitamas/DHM> (accessed on 1 August 2026).

Acknowledgments: We gratefully acknowledge the support of the HUN-REN AI4Science Program, which provides critical resources and infrastructure in the form of inspiration, education, consultation, and technology to the research community. Its mission to address researchers' needs in AI and to enable world-class results in scientific inquiry has been instrumental in advancing our work.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HEP	High-Energy Physics
DHM	Dead Hot Map
EMCal	Electromagnetic Calorimeter
PHENIX	Pioneering High Energy
QGP	Quark-Gluon Plasma
RHIC	Relativistic Heavy Ion Collider
CMS	Compact Muon Solenoid
DQM	Data Quality Monitoring
KDD	Knowledge Discovery in Database
LOF	Local Outlier Factor
OCSVM	One-Class Support Vector Machine
IF	Isolation Forest
KDE	Kernel Density Estimation

References

- Shuryak, E.V. Quantum Chromodynamics and the Theory of Superdense Matter. *Phys. Rep.* **1980**, *61*, 71–158. [[CrossRef](#)]
- Gyulassy, M.; McLerran, L. New forms of QCD matter discovered at RHIC. *Nucl. Phys. A* **2005**, *750*, 30–63. [[CrossRef](#)]
- Shiltsev, V.; Zimmermann, F. Modern and future colliders. *Rev. Mod. Phys.* **2021**, *93*, 015006. [[CrossRef](#)]
- Aphecetche, L.; Awes, T.C.; Banning, J.; Bathe, S.; Bazilevsky, A.; Belikov, S.; Belyaev, S.T.; Blume, C.; Bobrek, M.; Bucher, D.; et al. PHENIX calorimeter. *Nucl. Instrum. Meth. A* **2003**, *499*, 521–536. [[CrossRef](#)]
- Adcox, K.; Adler, S.S.; Aizama, M.; Ajitanand, N.N.; Akiba, Y.; Akikawa, H.; Alexander, J.; Al-Jamel, A.; Allen, M.; Alley, G.; et al. PHENIX detector overview. *Nucl. Instrum. Meth. A* **2003**, *499*, 469–479. [[CrossRef](#)]
- The CMS Collaboration. Autoencoder-Based Anomaly Detection System for Online Data Quality Monitoring of the CMS Electromagnetic Calorimeter. *Comput. Softw. Big Sci.* **2024**, *8*, 1. [[CrossRef](#)]
- Asres, M.W.; Omlin, C.W.; Wang, L.; Yu, D.; Parygin, P.; Dittmann, J.; Karapostoli, G.; Seidel, M.; Venditti, R.; Lambrecht, L.; et al. Spatio-Temporal Anomaly Detection with Graph Networks for Data Quality Monitoring of the Hadron Calorimeter. *Sensors* **2023**, *23*, 9679. [[CrossRef](#)]
- Asres, M.W.; Omlin, C.W.; Wang, L.; Yu, D.; Parygin, P.; Dittmann, J.; The CMS-HCAL Collaboration. Data Quality Monitoring for the Hadron Calorimeters Using Transfer Learning for Anomaly Detection. *Sensors* **2025**, *25*, 3475. [[CrossRef](#)]
- Acharya, S.; Adamová, D.; Adler, A.; Rinella, G.A.; Agnello, M.; Agrawal, N.; Ahammed, Z.; Ahmad, S.; Ahn, S.U.; Ahuja, I.; et al. Performance of the ALICE Electromagnetic Calorimeter. *J. Instrum.* **2023**, *18*, P08007. [[CrossRef](#)]
- Neik, T.X.; Dolatabadian, A.; Danilevich, M.F.; Upadhyaya, S.R.; Zhang, F.; Batley, J.; Edwards, D. Plant disease epidemiology in the age of artificial intelligence and machine learning. *Agric. Commun.* **2025**, *3*, 100089. [[CrossRef](#)]
- Li, G. Big data and computing models. *Big Data Res.* **2024**, *10*, 9–16. [[CrossRef](#)]
- Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*, 3rd ed.; Morgan Kaufmann: Burlington, MA, USA, 2011.
- Tan, P.-N.; Steinbach, M.; Karpatne, A.; Kumar, V. *Introduction to Data Mining*, 2nd ed.; Pearson Education: Singapore, 2018.
- Aggarwal, C.C. *Outlier Analysis*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2017.

15. Breunig, M.M.; Kriegel, H.-P.; Ng, R.T.; Sander, J. LOF: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*; Association for Computing Machinery: New York, NY, USA, 2000; pp. 93–104.
16. Schölkopf, B.; Platt, J.C.; Shawe-Taylor, J.; Smola, A.J.; Williamson, R.C. Estimating the support of a high-dimensional distribution. *Neural Comput.* **2001**, *13*, 1443–1471. [[CrossRef](#)]
17. Schölkopf, B.; Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2002.
18. Liu, F.T.; Ting, K.M.; Zhou, Z.-H. Isolation forest. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*; IEEE: Piscataway, NJ, USA, 2008; pp. 413–422.
19. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; CRC Press: Boca Raton, FL, USA, 1986.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.