Spatio-Temporal Anomaly Detection with Graph Networks for Data Quality Monitoring of the Hadron Calorimeter

Mulugeta Weldezgina Asres
Center for Artificial Intelligence Research (CAIR)
University of Agder, Norway
mulugetawa@uia.no

Christian Walter Omlin
Center for Artificial Intelligence Research (CAIR)
University of Agder, Norway
christian.omlin@uia.no

Long Wang
University of Maryland, USA

David Yu
Brown University, USA

Pavel Parygin
University of Rochester, USA

Jay Dittmann
Baylor University, USA

Georgia Karapostoli
University of California, Riverside, USA

Markus Seidel
Riga Technical University, Latvia

Rosamaria Venditti
Bari University and INFN, Italy

Luka Lambrecht
Ghent University, Belgium

Emanuele Usai
University of Alabama, USA

Muhammad Ahmad
Texas A&M University, USA

Javier Fernandez Menendez
Universidad de Oviedo, Spain

Kaori Maeshima
Fermi National Accelerator Laboratory, USA

The CMS-HCAL Collaboration
CERN, Switzerland

November 8, 2023

Abstract

The compact muon solenoid (CMS) experiment is a general-purpose detector for high-energy collision at the large hadron collider (LHC) at CERN. It employs an online data quality monitoring (DQM) system to promptly spot and diagnose particle data acquisition problems to avoid data quality loss. In this study, we present semi-supervised spatio-temporal anomaly detection (AD) monitoring for the physics particle reading channels of the hadronic calorimeter (HCAL) of the CMS using three-dimensional digi-occupancy map data of the DQM. We propose the GraphSTAD system, which employs convolutional and graph neural networks to learn local spatial characteristics induced by particles traversing the detector, and global behavior owing to shared backend circuit connections and housing boxes of the channels, respectively. Recurrent neural networks capture the temporal evolution of the extracted spatial features. We have validated the accuracy of the proposed AD system in capturing diverse channel fault types using the LHC Run-2 collision data sets. The GraphSTAD system has achieved production-level accuracy and is being integrated into the CMS core production system–for real-time monitoring of the HCAL. We have also provided a quantitative performance comparison with alternative benchmark models to demonstrate the promising leverage of the presented system.

Keywords Anomaly Detection · Monitoring · Spatio-Temporal · Deep Learning · Graph Networks · DQM · CMS · LHC

Acronyms

AD	Anomaly Detection
CMS	Compact Muon Solenoid
CNN	Convolutional Neural Network
DL	Deep Learning
DQM	Data Quality Monitoring
FC	Fully-Connected Neural Network
GNN	Graph Neural Network
GraphSTAD	Graph Based ST AD model
HCAL	Hadron Calorimeter
HE	HCAL Endcap detector
HEP	High Energy Physics
LHC	Large Hadron Collider
LS	Lumisection
MAE	Mean Absolute Error
MSE	Mean Square Error
QIE	Charge Integrating and Encoding
RBX	Readout Box
RNN	Recurrent Neural Network
SiPM	Silicon Photo Multipliers
ST	Spatio-Temporal
VAE	Variational Autoencoder
γ	Digi-occupancy
$i\eta$	$ieta$ axis coordinate of the spatial CMS-HCAL channels
$i\phi$	$iphi$ axis coordinate of the spatial CMS-HCAL channels
$depth$	$depth$ axis coordinate of the spatial CMS-HCAL channels

1 Introduction

The large hadron collider (LHC) is the largest particle collider ever built globally. It is designed to conduct experiments in physics and increase our understanding of the universe–with the expectation that new findings will lead to practical applications. The LHC was started in 2008 and consists of a 27 km ring tunnel located 100 meters underground at the France-Switzerland border near Geneva Evans and Bryant [2008]. The LHC is a two-ring superconducting hadron accelerator and collider capable of accelerating and colliding beams of protons and heavy ions with the unprecedented luminosity of 10^{34} cm^{-2}s^{-1} and 10^{27} cm^{-2}s^{-1}, respectively, at a velocity close to the speed of light–3×10^{8} ms^{-1} Evans and Bryant [2008], Heuer [2012]. The LHC consists of several experiments on its sites, and its ring holds several detectors for these experiments. The four major detectors of the LHC are *a toroidal LHC apparatus* (ATLAS), *compact muon solenoid* (CMS), *large hadron collider beauty* (LHCb), and *a large ion collider experiment* (ALICE). Each detector studies particle collisions from a different perspective with different technologies. The ATLAS at *point 1* (P1) and the CMS at *point 5* (P5) are the two high-luminosity general-purpose detectors at the LHC, and they are located in diametrically opposite sections.

The CMS experiment employs the data quality monitoring (DQM) system to guarantee high-quality physics data through online monitoring that provides live feedback during data acquisition, and offline monitoring that certifies the data quality after offline processing Azzolini et al. [2019]. The online DQM identifies emerging problems using reference distribution and predefined tests to detect known failure modes using summary histograms, such as a digi-occupancy map of the calorimeters Tuura et al. [2010], De Guio and Collaboration [2014]. A digi-occupancy map contains the histogram record of particle hits of the data-taking channels of the calorimeters. The CMS calorimeters could have several flaws, such as issues with the frontend particle sensing scintillators, digitization and communication systems, backend hardware, and algorithms, which are usually reflected in the digi-occupancy map. The growing complexity of the detector and the variety of physics experimentation make data-driven anomaly detection (AD) systems essential tools for the CMS to identify and localize detector anomalies automatically. Recent efforts in the CMS have proposed deep learning (DL) for AD applications for the DQM Azzolin et al. [2019], Azzolini et al. [2019], Pol et al. [2019a,b]. The CMS detector consists of a tracker to reconstruct particle paths accurately, two calorimeters, the electromagnetic (ECAL) and the hadronic (HCAL) to detect electrons, photons, and hadrons, respectively, and of several muon detectors. The synergy in AD has thus far achieved promising results on spatial 2D histogram maps of the DQM for the ECAL Azzolin et al. [2019] and the muon detectors Pol et al. [2019b]. Previous studies only considered extreme anomalies, such as dead–no reading–and hot–high noise–particle sensing channels. Detecting degrading channels–essential for quality

deterioration monitoring and early intervention–is often overlooked. For instance, the improperly tuned bias voltage on the HCAL physics particle sensing channels caused non-uniformity in the hit map of the DQM, but the channels were neither dead nor hot Viazlo and Collaboration [2022]. Calorimeter channels may degrade with subtle abnormality before reaching extreme channel fault status. Capturing such subtle anomalies–e.g., slow system degradation–makes temporal AD models appealing for early anomaly prediction before ultimate system failure. Time-aware models extract temporal context to enhance AD performance. A few efforts have thus far been focused on temporal models despite the acknowledged potential in the future automation technology challenges at the LHC Azzolin et al. [2019], Wielgosz et al. [2018]. Our study focuses on DQM automation through time-aware AD modeling using digi-occupancy histogram maps of the HCAL. The digi-occupancy data of the HCAL is 3D, and it poses multidimensional challenges due to its depth-wise calorimeter segmentation; it is relatively unexplored with ML endeavors. The particle hit map data of the HCAL are highly dependent on the collision luminosity–a measure of how many collisions are happening in a particle accelerator–and the number of particles traversing the calorimeter. The effort on data normalization that enhances the learning generalization of machine learning models is still limited.

We address the above gaps while investigating the performance enhancement of temporal AD DL models for the HCAL DQM. We propose to detect anomalies of the HCAL particle sensing channels through a semi-supervised AD system–GraphSTAD–from spatial digi-occupancy maps of the DQM. Anomalies can be unpredictable and come in different patterns of severity, shape, and size–often limiting the availability of labeled anomaly data covering all possible faults. We employ a semi-supervised approach for the AD system; the concept for the AD is that an autoencoder (AE)–trained to reconstruct healthy digi-occupancy maps–would adequately reconstruct the healthy maps, whereas it yields high reconstruction error for maps with anomalies.

Since abnormal events can have spatial appearance and temporal context, we combine both the spatial and temporal features–spatio-temporal (ST)–for AD Xu et al. [2017], Chang et al. [2022], Luo et al. [2019], Hasan et al. [2016], Wu et al. [2020], Hsu [2017], Ullah et al. [2021], Hu et al. [2019], Banković et al. [2012], Zhang et al. [2020]. Moreover, modeling the digi-occupancy map of the HCAL is challenging–the spatial nature may exhibit irregularity; although adjacent channels with Euclidean distance are exposed to collision article hits around their region, the channels may belong to different backend circuits–resulting in non-Euclidean spatial behavior on the digi measurements. The GraphSTAD system captures the behavior of channels from regional collision particle hits, and electrical and environmental characteristics due to a shared backend circuit of the channels to effectively detect the degradation of faulty channels. The AD system attains these utilities using a deep AE model that learns local spatial behavior, physical connectivity-induced shared behavior, and temporal behavior through convolutional neural networks (CNN), graph neural networks (GNN), and recurrent neural networks (RNN), respectively.

We have evaluated our proposed AD approach in detecting spatial faults and temporal discords on digi-occupancy maps of the HCAL. We simulated different realistic types of anomalies–*dead channels* without registered hits, *hot channels* dominated by electronic noise resulting in a much higher hit count than expected, and *degrading channels* with deteriorated particle detection efficiency resulting in lower hit counts than expected–to analyze the effectiveness of the AD model. The results have demonstrated promising performance in detecting and localizing the anomalies. We have validated the accuracy in detecting real anomalies and discussed a comparison to the existing DQM system.

In this paper, we briefly describe the DQM and HCAL systems in Section 2, and our data sets in Section 3. Section 4 explains the methodology of the proposed GraphSTAD model, and Section 5 presents the performance evaluation and result discussion. Finally, we summarize the contribution of our study in Section 6.

2 Background

This section describes the CMS DQM and the HCAL systems.

2.1 The Data Quality Monitoring of the CMS Experiment

The particle collision data of the LHC is aggregated into runs, where each run contains thousands of lumisections. A lumisection (LS) corresponds to approximately 23 seconds of data taking and comprises hundreds or thousands of collision events containing particle hit records. The DQM of the CMS provides feedback on detector performance and data reconstruction; It generates a list of certified data for physics analyses–the "Golden JSON" Azzolini et al. [2019]. The DQM employs online and offline monitoring mechanisms: 1) the online monitoring is real-time DQM during data acquisition, and 2) the offline monitoring–after 48 hours since the collisions were recorded–provides the final fine-grained data quality analysis for data certification. The online DQM populates a set of histogram-based maps on a selection of events and provides summary plots with alarms that DQM experts inspect to spot problems. The digi-occupancy–one of the histogram maps generated by the online DQM–contains particle hit histogram records of the

particle readout channel sensor of the calorimeters. A digi–also called hit–is a reconstructed and calibrated collision physics signal of the calorimeter. Several faults in the calorimeter affecting the frontend particle sensing scintillators, digitalization, communications, the backend hardware, and the algorithms–could appear in the digi-occupancy map. Previous efforts by Azzolin et al. [2019], Azzolini et al. [2019], Pol et al. [2019a,b] demonstrate the promising AD efficacy of using digi-occupancy maps for calorimeter channel monitoring using machine learning. However, end-to-end DL with temporal models are relatively unexplored Azzolin et al. [2019], Pol et al. [2019b].

The purpose of leveraging the DQM through machine learning is to address particular challenges: 1) latency of human intervention, and thresholds require sufficient statistics; 2) the volume of data a human can process in a finite time is limited; 3) rule-based approaches do not scale and assume limited potential failure scenarios; 4) dynamic running conditions change reference samples; 5) the effort to train human shifters who monitor DQM dashboards, and maintain instructions is expensive. Developing machine learning models for the DQM comes with some impediments despite the potential promises; data normalization to handle variation in experimental settings, the granularity of the failures to spot, and limited availability of the ground truth labels are among the challenges Pol et al. [2019b].

We extend the efforts in AD with spatio-temporal (ST) modeling of the digi-occupancy maps of the DQM for the HCAL. Several promising AD models have been proposed in the literature for ST data in non-HEP domains Atluri et al. [2018]–such as crowd monitoring using visual streaming data Xu et al. [2017], Chang et al. [2022], Luo et al. [2019], Wu et al. [2020], Hasan et al. [2016], Ullah et al. [2021], Hu et al. [2019], traffic monitoring Hsu [2017], Deng et al. [2022], cyber-security on sensor systems Banković et al. [2012], Tišljarić et al. [2021], Jiang et al. [2022], medical diagnosis Ahmedt-Aristizabal et al. [2021], and environment monitoring Zhang et al. [2020]. A unique quality of ST data that differentiates it from other traditional data studied is the presence of dependencies among measurements induced by the spatial and temporal attributes, where data correlations are more complex to capture by conventional techniques Atluri et al. [2018]. Spatio-temporal anomaly is defined as a data point or cluster of data points that violate the nominal ST correlation structure of the normal points Hsu [2017], Deng et al. [2022], Banković et al. [2012], Tišljarić et al. [2021], Xu et al. [2017], Chang et al. [2022], Luo et al. [2019], Wu et al. [2020], Hasan et al. [2016], Ullah et al. [2021], Ahmedt-Aristizabal et al. [2021], Jiang et al. [2022]. The previous ST AD studies on video data sets Chang et al. [2022], Luo et al. [2019], Wu et al. [2020], Hasan et al. [2016] focus on CNN models for regular spatial feature extraction, and GNNs are gaining popularity for sensor and traffic flow data Hsu [2017], Deng et al. [2022] that exhibit irregular spatial attributes with non-Euclidean distance among nodes. GNNs have recently achieved promising results at the LHC Duarte and Vlimant [2022], Shlomi et al. [2020], and outperformed CNN in learning irregular calorimeter geometry Qasim et al. [2019] and in pileup mitigation Martínez et al. [2019]. The spatial characteristics of the HCAL channels exhibit regular spatial positioning of particle hits in the calorimeter and irregularity in measurement, as adjacent channels may share different backend circuits. Our proposed model integrates both CNN and GNN Bruna et al. [2013], Kipf and Welling [2016] to capture Euclidean and non-Euclidean spatial characteristics, respectively, and RNN for temporal learning for the HCAL channels.

2.2 The Readout Boxes of the HCAL

The HCAL is a specialized calorimeter to capture hadronic particles. The calorimeter is made of multiple subsystems such as HCAL endcap (HE), HCAL barrel (HB), HCAL forward (HF), and HCAL outer (HO) (see Fig. 1).
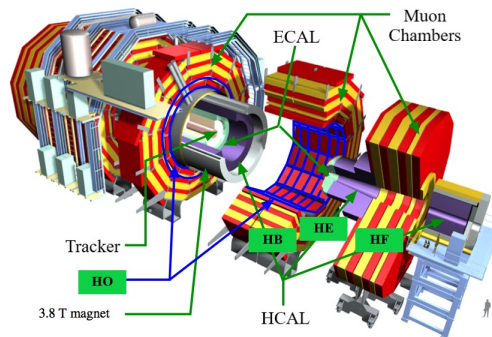

Figure 1: Schematic of the CMS detector at P5 of the LHC and its calorimeters Focardi [2012].

The HCAL frontend electronic systems readout boxes (RBXes) to house the data acquisition electronics. The RBXes provide high voltage, low voltage, backplane communications, and cooling to the data acquisition electronics. The use-case of our study–the HE–is made of 36 RBXes arranged on the plus and minus hemispheres of the CMS. Its frontend particle detection system is built on brass and plastic scintillators, and transmits the produced photon particles

through the wavelength-shifting fibers to the Silicon photomultipliers (SiPMs) (see Fig. 2). Each RBX houses four readout modules (RMs) for signal digitization Strobbe [2017]; each RM has a SiPM control card, 48 SiPMs, and four readout cards–each includes 12 charge integrating and encoding chips (QIE11) and a field programmable gate array (Microsemi Igloo2 FPGA). The QIE integrates charge from each SiPM at 40 MHz, and the FPGA serializes and encodes the data from the QIE. The encoded data is optically transmitted to the backend system via the CERN versatile twin transmitter (VTTx) at 4.8 Gbps. The current HCAL system has 17 detector scintillator layers that are read out in seven groups–hereafter referred to as $depths$; the light from the scintillators in any given group is optically added together by sending it to a single SiPM. More channels allow for a more refined depth segmentation–ideal for precisely calibrating the depth-dependent radiation damage on the HCAL Azzolini et al. [2019].
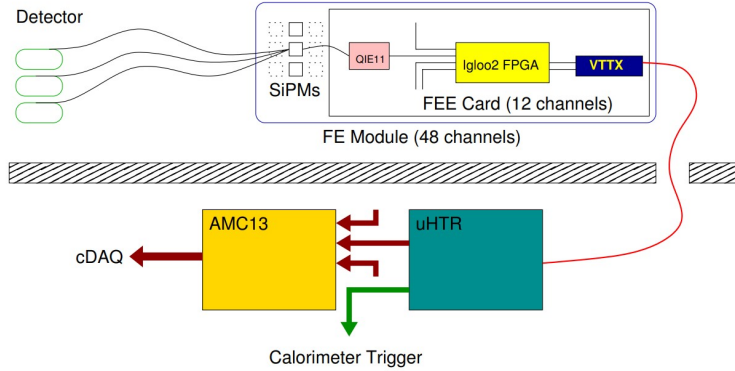
Figure 2: The data acquisition chain of the HE–including the SiPMs, the frontend readout cards, and the optical link connected to the back-end electronics Strobbe [2017]. Each readout card contains twelve QIE11 for charge integration, an Igloo2 FPGA for data serialization and encoding, and a VTTx optical transmitter. A fault in the chain may cause anomalous digi-occupancy reading in the online DQM.

3 Data Set Description

We used digi-occupancy map data of the online DQM system of the CMS experiment to train and validate the proposed AD system. The digi-occupancy map data has 3D spatial dimensions with η ϕ, and $depth$ axes, and contains digi histogram records of the calorimeter readout channels referenced by $i\eta = [-32, 32]$, $i\phi = [1, 72]$ and $depth = [1, 7]$ axes (see Fig. 3). The value of the digi-occupancy map varies with the received luminosity–the recorded by the CMS and hereafter referred to as the luminosity–and the number of events–particles traversing the calorimeter–that may differ across LSs. The maps from a sequence of LSs constitute attribution of ST data with correlated spatial and temporal relations Atluri et al. [2018].

We utilized data collected in 2018 during the LHC Run-2 collision experiment. The data set contains about 20K LSs from 20 healthy runs–collected by the CMS experiment–pre-scrutinized by the CMS certifiers, and registered the "Golden JSON" of the DQM as declared of good quality Rapsevicius et al. [2011]. The maps–one per LS–were populated with the per LS received luminosity up to 0.4 pb^{-1}, and the number of events up to 2250. Our working dataset contains 20K map samples–each with a dimension of $[i\eta = 64 \times i\phi = 72 \times depth = 7]$).

4 Methodology

This section presents the proposed GraphSTAD system for online DQM of the HCAL using digi-occupancy map data.

An anomaly is an odd observation from the bulk of observations–often indicating peculiar underlying incidents Chalapathy and Chawla [2019]. AD methods can broadly be categorized as supervised or unsupervised: 1) supervised approaches require annotated ground-truth anomaly observations, and 2) unsupervised approaches do not require labeled anomaly data and are more generally pragmatic in many real-world application settings, as data annotation is an expensive task. Unsupervised AD models trained with only healthy observations are called semi-supervised AD approaches.

We present an ST reconstruction AE to detect abnormality in the HCAL channels using reconstruction deviation scores on ST digi-occupancy maps from consecutive lumisections (see Fig. 4). The AE combines CNN, GNN, and RNN to capture ST characteristics of digi-occupancy maps. The spatial feature extraction of the CNNs is leveraged with GNNs
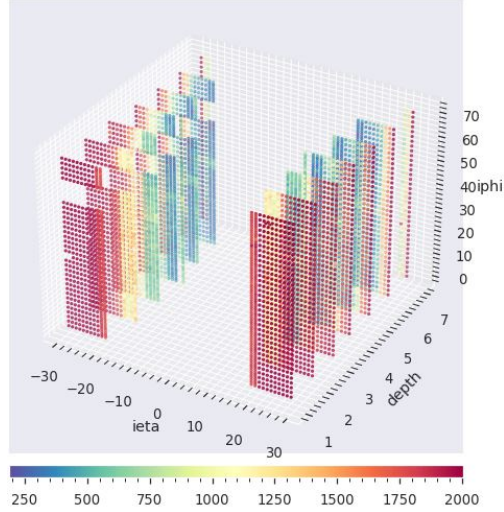
5

Figure 3: Digi-occupancy map (*year=2018, RunId=325170, LS=15*) of the HE. The HE channels are placed in $|i\eta| = [14, 29]$, $i\phi = [1, 72]$, and $depth = [1, 7]$. Each pixel in the map corresponds to the readout of a calorimeter channel after a particle hit. The HCAL covers a considerable volume of CMS and has a fine segmentation along three axes ($i\eta$, $i\phi$, and $depth$). The missing section near the top-left of the map is due to two failed RBX (HEM15 and HEM16) sectors during the 2018 collision runs.

to learn circuit and housing connectivity-induced spatial behavior irregularities among the HCAL sensor channels. There are approximately 7K channels–pixels–on the digi-occupancy map of the HCAL endcap subsystem–housed in 36 RBXes. The channels in a given RBX are susceptible to system faults in the RBX due to the shared backbone circuit and environmental factors, such as temperature and humidity. Behavior variations among RBXes have also been observed due to intrinsic deviations of the custom-built electronic components in the RBXes. Our proposed GraphSTAD employs GNNs–in its spatial feature extraction pipeline–to capture the characteristics of the HCAL channels owing to their shared physical connectivity to a given RBX. GNNs have recently achieved promising results in several applications at the LHC Duarte and Vlimant [2022], Shlomi et al. [2020], and outperformed CNN in learning irregular calorimeter geometry Qasim et al. [2019] and in pileup mitigation Martínez et al. [2019]. The GraphSTAD system exploits both CNN and GNN Bruna et al. [2013], Kipf and Welling [2016] to capture Euclidean and non-Euclidean spatial characteristics of the HCAL channels, respectively.
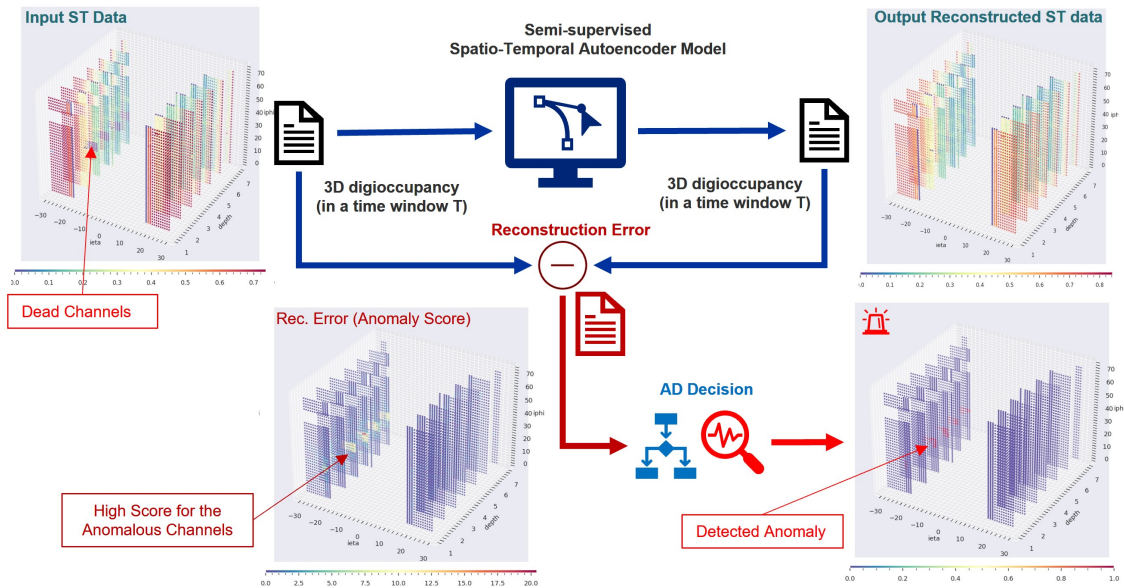

Figure 4: The proposed AE-based reconstruction AD system. The AE reconstructs the input ST digi-occupancy map, and the AD decision is performed on the anomaly scores estimated from the reconstruction errors.

4.1 Data Preprocessing

This section describes the data preprocessing stages of the proposed approach–i.e., digi-occupancy renormalization for particle collision experiment setting variations and graph adjacency matrix generation for the readout channels of the HCAL.

4.1.1 Digi-occupancy Map Renormalization

The digi-occupancy (γ) map data of the HCAL varies with the received luminosity (β) and the number of events (ξ) (see Fig. 5). We devise a renormalization of the γ through a regression model $\mathcal{R}$ to have a consistent quantity interpretation of the γ and build a model that robustly generalizes previously unseen run settings–β and ξ variations. The $\mathcal{R}$ estimates the renormalizing $\bar{\gamma}_s$ at the sth LS using β and ξ as:

$$\bar{\gamma}_s = \mathcal{R}(\xi, \beta) \tag{1}$$

The model $\mathcal{R}$ is trained to minimize the MSE cost function, $\mathbb{E}[(\gamma_s - \bar{\gamma}_s)^2]$, where γ_s is calculated as:

$$\gamma_s = \sum_{\forall i} \gamma(s, i) \tag{2}$$

where the $\gamma(s, i)$ is the digi-occupancy of the ith channel in the map at the sth LS. Finally, the per-channel $\gamma(s, i)$ is renormalized by its corresponding $\bar{\gamma}_s$ as:

$$\hat{\gamma}(s, i) = \frac{K\gamma(s, i)}{\bar{\gamma}_s} \tag{3}$$

where the $\hat{\gamma}$ is the renormalized γ, and the K is a scaling factor to compensate for the difference in the number of channels on the depth axes.

We have employed fully connected (FC) neural networks to build the regression model to effectively capture the non-linear relationships illustrated in Fig. 5:

$$input(\xi, \beta) \to \text{ReLU}(\text{FC}(64)) \to \text{ReLU}(\text{FC}(64)) \to \text{ReLU}(\text{FC}(64)) \to output(\bar{\gamma}_s) \tag{4}$$

Fig. 6 depicts data distribution of the γ_s before and after renormalization with $\mathcal{R}$. The renormalization has successfully handled the discrepancies on the γ_s from several runs–overlaps and centers distributions of $\hat{\gamma}_s$ and minimizes the outliers.

4.1.2 Adjacency Matrix Generation for Graph Network

We deploy an undirected graph network $\mathcal{G}(\mathcal{V}, \Theta)$ to represent the HCAL channels in a graph network based on their connection to a shared RBX system. The graph $\mathcal{G}$ contains nodes $\upsilon \in \mathcal{V}$, with edges $(\upsilon_i, \upsilon_j) \in \Theta$ in a binary adjacency matrix $\mathcal{A} \in \mathbb{R}^{M \times M}$, where M is the number of channel nodes. An edge indicates the channels sharing the same RBX as:

$$A(\upsilon_i, \upsilon_j) = \begin{cases} 1, & \text{if } \Omega(\upsilon_i) = \Omega(\upsilon_j) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $\Omega(\upsilon)$ returns the RBX ID of the channel υ.

There are approximately 7K channels in a graph representation of the digi-occupancy map of the HE, where each RBX network contains roughly 190 nodes. We retrieved the channel to RBX mapping from the HCAL's calorimeter segmentation map.

4.2 Anomaly Detection Modeling

We denote the AE model of the GraphSTAD system as $\mathcal{F}$. The ST data is $X \in \mathbb{R}^{T \times N_{i\eta} \times N_{i\phi} \times N_d \times N_f}$ as a sequence in a time window $t_x \in [t - T, t]$, where $N_{i\eta} \times N_{i\phi} \times N_d$ is the spatial dimension corresponding to the $i\eta$, $i\phi$, and $depth$ axes, respectively, and $N_f = 1$ is the number of input variables–only a digi-occupancy quantity in the spatial data. The $\mathcal{F}_{\theta, \omega} : X \to \bar{X}$–parameterized by θ and ω–attempts to reconstruct the input ST data X and outputs $\bar{X}$. The encoder network of the model $E_\theta : X \to z$ provides low-dimension latent space, $z = E_\theta(X)$, and the decoder $D_\omega : z \to \bar{X}$, reconstructs the ST data from z, $\bar{X} = D_\omega(z)$ as:

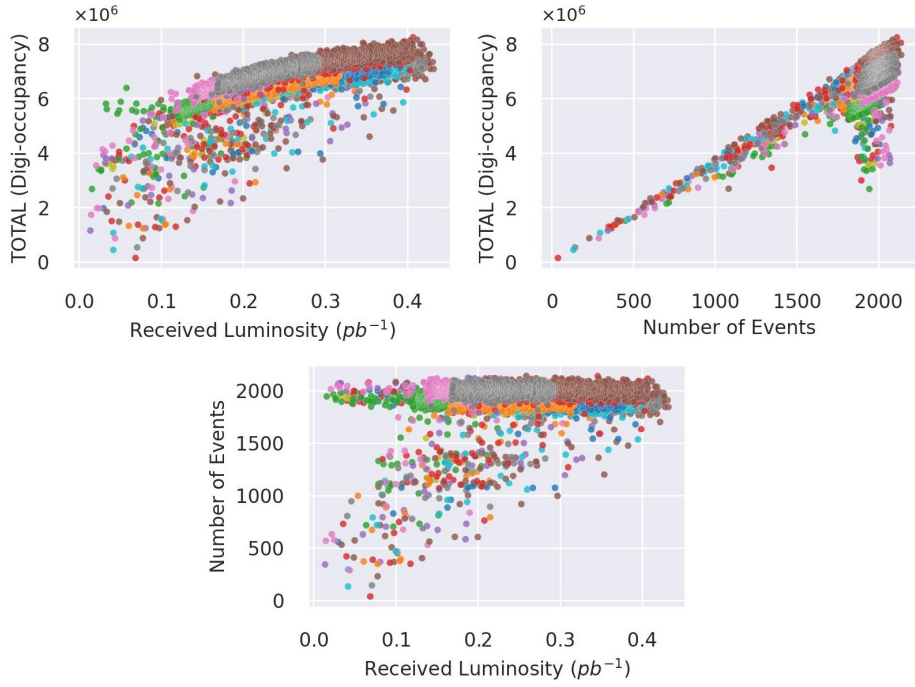$$\bar{X} = \mathcal{F}(X) = D_\omega(E_\theta(X)) \tag{6}$$

7

Figure 5: Dependency between digi-occupancy and run settings–the received luminosity and the number of events–in LS granularity. The number of events did not fully follow the drop in the luminosity (right-side plot) and the digi-occupancy (middle plot); it portrays the non-linear behavior of LHC. The different colors correspond to different collision runs.
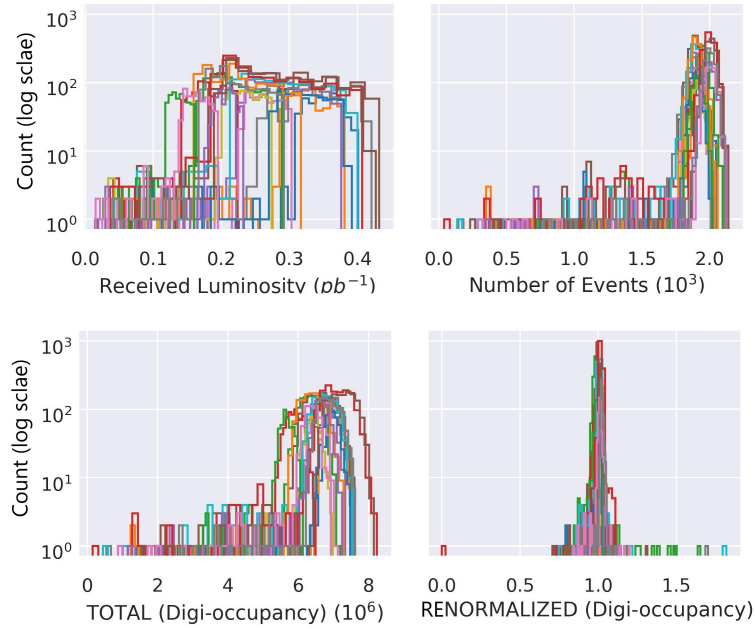

Figure 6: Distribution of total digi-occupancy per LS before and after renormalization. From left to right: a) the received luminosity, b) the number of events, c) the digi-occupancy, and d) the renormalized digi-occupancy obtained with the regression model described in the text. The different colors correspond to different runs.

The channel anomalies can be transients–live short and impact only a single digi-occupancy map–or persist over time–affecting a sequence of maps. The spatial reconstruction error e is calculated to detect a transient anomaly as:

$$e_i = |x_i - \bar{x}_i| \tag{7}$$

where $x_i \in \mathrm{X}$ and $\bar{x}_i \in \bar{\mathrm{X}}$ are the input and reconstructed digi-occupancy of the ith channel. The e_i detects channel abnormality occurrence on isolated maps. We engage an aggregated error in a time window T using mean absolute error (MAE) to capture a time-persistent anomaly as:

$$e_{i,MAE} = \frac{1}{T} \sum_{t'=t-T}^{t} e_i(t') \tag{8}$$

We standardize e_i to regularize the reconstruction accuracy variations among the channels–allowing a single AD decision threshold α to all the channels in the spatial map–as:

$$a_i = \frac{e_i}{\sigma_i} \tag{9}$$

where σ_i is the standard deviation of the e_i (or $e_{i,MAE}$ if the time window is considered) on the training dataset. The anomaly flags are generated after applying α to the anomaly scores a $a_i > \alpha$. The α is a tunable constant that controls the detection sensitivity.

4.3 Autoencoder Model Architecture

Convolutional neural networks have achieved state-of-the-art performance in several AD DL applications with image data Chang et al. [2022], Luo et al. [2019], Hasan et al. [2016], Hsu [2017], Wu et al. [2020]. The shared nature of the kernel filters of the CNNs substantially reduces the number of trainable parameters in the model compared to fully connected neural networks. Directly supplying the learned spatial features into an architecture that can learn temporal data–such as RNN–could become inherently challenging due to the considerable computational demand for high-dimensional data. We employ CNN and GNN with a pooling mechanism to extract relevant features from high dimensional spatial data followed by RNN to capture temporal characteristics of the extracted features (see Fig. 7). We integrate variational layer Kingma and Welling [2013] at the end of the encoder for regularization of AE overfitting by enforcing continuous and normally distributed latent representations Asres et al. [2021], Guo et al. [2018], An and Cho [2015], Chadha et al. [2019].

The CNN of the encoder has L_c networks–each containing $\mathrm{Conv3D}(\cdot, \mathrm{kernel\_size} = [3 \times 3 \times 3])$[1] for regular spatial learning followed by batch normalization (BN)[2] for network weight regularization and faster convergence, ReLU for nonlinear activation, and $\mathrm{MaxPooling3D}$[3] for spatial dimension reduction. The model can be summarized as:

$$y_t^c, \psi_t^c = \mathrm{Pool}(\mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv3D}(x_t^l, N_c^l))|_{l=1,...,L_c} \tag{10}$$

where x_t^l is the input spatial γ map data at time-step t and the N_c^l is the feature size of the lth network. The y_t^c is the extracted feature set of the CNN at t. The $\mathrm{Pool}(\cdot)$ denotes $\mathrm{MaxPooling3D}(\cdot, stride = [2 \times 2 \times 2])$. The ψ_t^c holds the pooling spatial location indices of the $\mathrm{MaxPooling3D}$ layers to be used later for upsampling in the decoder during map reconstruction. The final extracted feature set $\mathrm{Y}_c \in \mathbb{R}^{T \times N_c}$ of the CNN is an aggregation of all y_t^c in the time window T–concatenated on the time dimension–as:

$$\mathrm{Y}_c = [y_1^c, y_2^c, ..., y_T^c] \tag{11}$$

We have used $L_c = 4$ to map the input spatial dimension $[64 \times 72 \times 7]$ into $[4 \times 4 \times 1]$, which yields a reduction factor of 2^{L_c} and expands the feature space of the input from $N_f = 1$ to $N_c = 128$. The $N_c' : [4 \times 4 \times 1 \times 128] = 2048$ spatial features are generated after reshaping.

The GNN of the encoder has L_g networks of a graph convolutional network (GCN)[4] with ReLU activation, and a final global attention pooling[4]. The networks are summarized as:

$$y_t^g = \mathrm{Pool}(\mathrm{ReLU}(\mathrm{GCN}(x_t^l, N_g^l)|_{l=1,...,L_g}))$$
$$\mathrm{Y}_g = [y_1^g, y_2^g, ..., y_T^g] \tag{12}$$

---

[1]https://pytorch.org/docs/stable/generated/torch.nn.Conv3d.html
[2]https://pytorch.org/docs/stable/generated/torch.nn.BatchNorm3d.html
[3]https://pytorch.org/docs/stable/generated/torch.nn.MaxPool3d.html
[4]https://docs.dgl.ai/en/0.2.x/tutorials/models/1_gnn/1_gcn.html
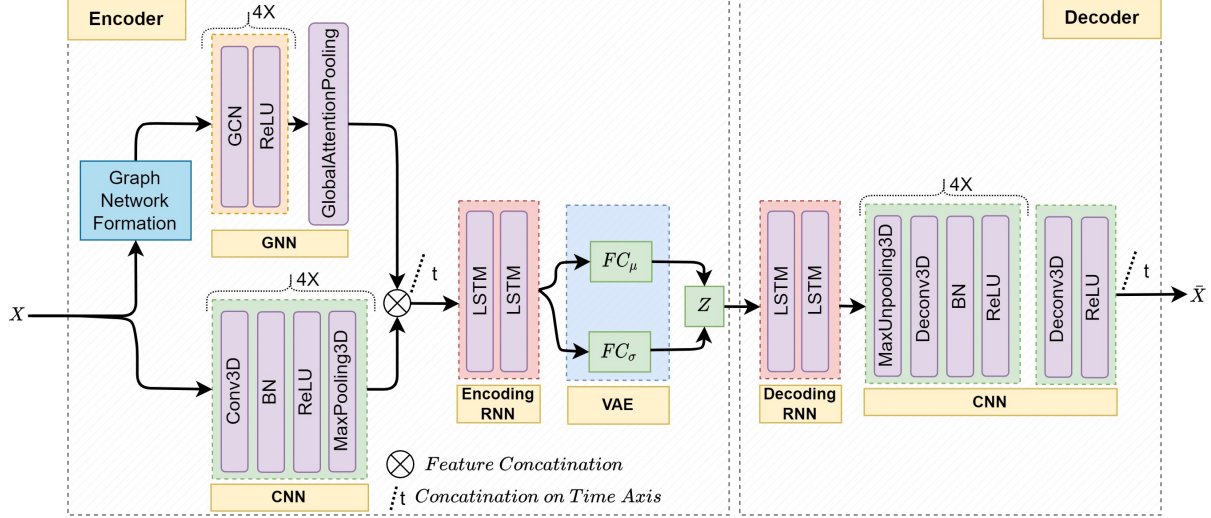
Conv3D: 3D convolutional neural network; GCN: graph convolutional neural networks; Deconv3D: 3D deconvolutional neural networks; BN: batch normalization; LSTM: long short time memory recurrent networks; FC: fully connected neural networks.

Figure 7: The architecture of the proposed AE for the GraphSTAD system. The GNN and CNN are spatial feature extraction on each time step, and the RNN network captures the temporal behavior of the extracted features. The encoder–feature extraction–incorporates the GNN for backend physical connectivity among the spatial channels, CNN for regional spatial proximity of the channels, and RNN for temporal behavior extraction. The decoder–reconstruction–contains RNN and deconvolutional neural networks to reconstruct the spatio-temporal input data from the low dimensional latent features.

where the GCN layers have a feature size of N_g^l, and the $\text{Pool}(\cdot)$ signifies the GlobalAttentionPooling$(\cdot)$ at the end of the GNN. The GlobalAttentionPooling aggregates the graph node features with an attention mechanism to obtain the final feature set of the GNN $Y_g \in \mathbb{R}^{T \times N_g}$. Similar to the CNN, we set $L_g = 4$ and $N_g = 128$ to generate the Y_g.

The encoded ST feature set $\zeta \in \mathbb{R}^{1 \times N_z}$ is obtained by learning the temporal context on the extracted spatial features $Y = [Y_c, Y_g]$ with two layers of long short-time memory (LSTM)[5] as:

$$\zeta = \text{LSTM}(Y, N_r^l)|_{l=1,2} \tag{13}$$

where N_r^l is the feature size of the lth LSTM layer. The last layer ($N_r^2 = N_z = 32$) generates the low-dimensional latent representation of the encoder. The VAE layer of the encoder generates the normally distributed representation latent features $\mathbf{z}$ as:

$$\mathbf{z} = \mu_z + \sigma_z \odot \epsilon \tag{14}$$

where $\odot$ signifies an element-wise product with standard normal distribution sampling $\epsilon \sim \mathcal{N}(0,1)$ An and Cho [2015]. The μ_z and the σ_z of the VAE are implemented with FC[6] layers taking the ζ as input.

The decoder network of the AE is made of RNN and CNN to reconstruct the target ST data from the latent features. The decoding embarks with temporal feature reconstruction using LSTM network as:

$$\bar{\zeta} = \text{LSTM}(\mathbf{z}, N_r^l)|_{l=1,2} \tag{15}$$

where $\bar{\zeta}$ is the reconstructed temporal feature set from the latent space $\mathbf{z}$. Spatial reconstruction follows for each time-step t through a multi-layer deconvolutional neural network. Each network starts with $\text{MaxUnpooling3D}(\cdot, stride = [2 \times 2 \times 2], \psi_l^c)$[7] to upsample the spatial data using localization indices ψ_l^c from the lth MaxPooling3D of the encoder followed by a deconvolutional layer ($\text{Deconv3D}(\cdot, \text{kernel\_size} = [3 \times 3 \times 3])$) Zeiler et al. [2010], BN and ReLU. Eventually, $\text{Deconv3D}(\cdot, \text{kernel\_size} = [1 \times 1 \times 1])$ is incorporated for final output stabilization. The decoder network is summarized as:

$$\bar{x}_t = \text{ReLU}(\text{BN}(\text{Deconv3D}(\text{Unpool}(\bar{\zeta}_t, \psi_t^c), N_c^l)))|_{l=1,...,L_c}$$
$$\bar{x}_t = \text{ReLU}(\text{Deconv3D}(\bar{x}_t, N_f)) \tag{16}$$

---

[5]https://pytorch.org/docs/stable/generated/torch.nn.LSTM.html
[6]https://pytorch.org/docs/stable/generated/torch.nn.Linear.html
[7]https://pytorch.org/docs/stable/generated/torch.nn.MaxUnpool3d.html

where the $\bar{x}_t$ is the reconstructed spatial data, and the $\mathrm{Unpool}(\cdot)$ denotes $\mathrm{MaxUnpool3D}(\cdot)$. The final reconstructed ST data $\bar{\mathrm{X}} \in \mathbb{R}^{T \times N_{i\eta} \times N_{i\phi} \times N_d \times N_f}$ is obtained as:

$$\bar{\mathrm{X}} = [\bar{x}_1, \bar{x}_2, ..., \bar{x}_T] \tag{17}$$

4.4 Model Training

We trained the AE on healthy digi-occupancy maps of LHC collision runs (described in Section 3). The modeling task becomes a multivariate learning problem since the target data contains readings from multiple calorimeter channels in the spatial digi-occupancy map. Appropriate scaling of the spatial data is thus necessary for effective model training; we further normalized the spatial data per channel into a range of $[0, 1]$. We have also observed that the γ distribution of the channels at the first depth of the spatial map is different from the channels at the higher depths (see Fig. 3); distribution imbalance on target channel data affects model training efficacy when well-known statistical algorithms, such as MSE, are employed as loss functions. MSE loss minimizes the cost of the entire space, and it may converge to a non-optimal local minimum in the presence of imbalanced data distribution; this phenomenon is known as the class imbalance challenge in machine learning classification problems. A popular remedy is to employ a weighting mechanism–assigning weights to the different targets. We applied a weighted MSE loss function to scale the loss from the different distributions–the $depth \in 1$ and $depth \in 2, \ldots, 7$–as:

$$\mathcal{L}' = \sum_j \frac{\varsigma_j}{M_j} \sum_{i \in C_j} (x_i - \bar{x}_i)^2 \tag{18}$$

where x_i is the $\hat{\gamma}$ of the ith channel in the jth group set C_j, M_j is the number of channels in C_j, and ς_j is the weight factor of the MSE loss of the jth group. We holistically set $\varsigma_1 = 0.4$ and $\varsigma_2 = 1$ after experimenting with several different ς values.

The VAE regularizes the training MSE loss using the KL divergence loss D_{KL} to achieve the normally distributed latent space as:

$$\mathcal{L} = \underset{W \in \mathbb{R}}{\mathrm{argmin}} \left\{ \mathcal{L}' - \lambda D_{KL} \left[\mathcal{N}(\mu_z, \sigma_z) \| \mathcal{N}(0, I) \right] + \rho \|W\|_2^2 \right\} \tag{19}$$

where $\mathcal{N}$ is a normal distribution with zero mean and unit variance, and $\|.\|$ is the *Frobenius norm* of L_2 regularization for the trainable model parameters W. The $\lambda = 0.003$ and $\rho = 10^{-7}$ are tunable regularization hyperparameters. We finally used *Adam* optimizer with super-convergence *one-cyclic* learning rate scheduling Smith and Topin [2019] for training.

5 Results and Discussion

In this section, we discuss the AD performance of the proposed GraphSTAD on simulated and real anomalies.

The ML studies for the CMS DQM mostly inject simulated anomalies into good data to validate the effectiveness of the developed models Azzolin et al. [2019] since a small fraction of the data is affected by real anomalies. We trained the AE model on 10K digi-occupancy maps–from LS sequence number $[1, 500]$–and evaluated on LSs $[500, 1500]$ injected with synthetic anomalies simulating real dead, hot, and degrading channels. We trained the AE on four GPUs with early-stopping using 20% of the training dataset to estimate the validation loss during each training epoch.

Fig. 8 demonstrates the capability of the proposed ST AE in reconstructing normal digi-occupancy maps from a sequence of lumisections. The AE has accomplished a promising reconstruction ability on the ST digi-occupancy data. High reconstruction accuracy on the healthy data is essential to reduce false-positive flags when a semi-supervised AE is employed for AD application. We further discuss the reconstruction error distribution comparison on the healthy and abnormal channels in the AD performance in Section 5.1.2.

We will discuss below the AD performance of the proposed system and comparisons with benchmark models, and present the detection of real faulty channels to demonstrate the accuracy of the proposed approach.

5.1 Anomaly Detection Performance

We generated synthetic anomalies simulating real dead, hot, and degrading channels and injected them into healthy digi-occupancy maps; the anomaly generation algorithm involves three steps: 1) selection of a random set of LSs from the test set, 2) random selection of spatial locations φ for each LS, where $\varphi \in \{i\eta \times i\phi \times depth\}$ on the HE axes (see Fig. 3), and 3) injection of anomalies such as *dead* ($\gamma_{anomaly} = 0$), *hot* ($\gamma_{anomaly} >> \gamma_{expected}$), and *degrading* channels ($0 \leq \gamma_{anomaly} < \gamma_{expected}$) into digi-occupancy maps of the LSs. We have kept the same spatial locations of the generated anomalies for consistency when evaluating the AD performance of the different anomaly types.
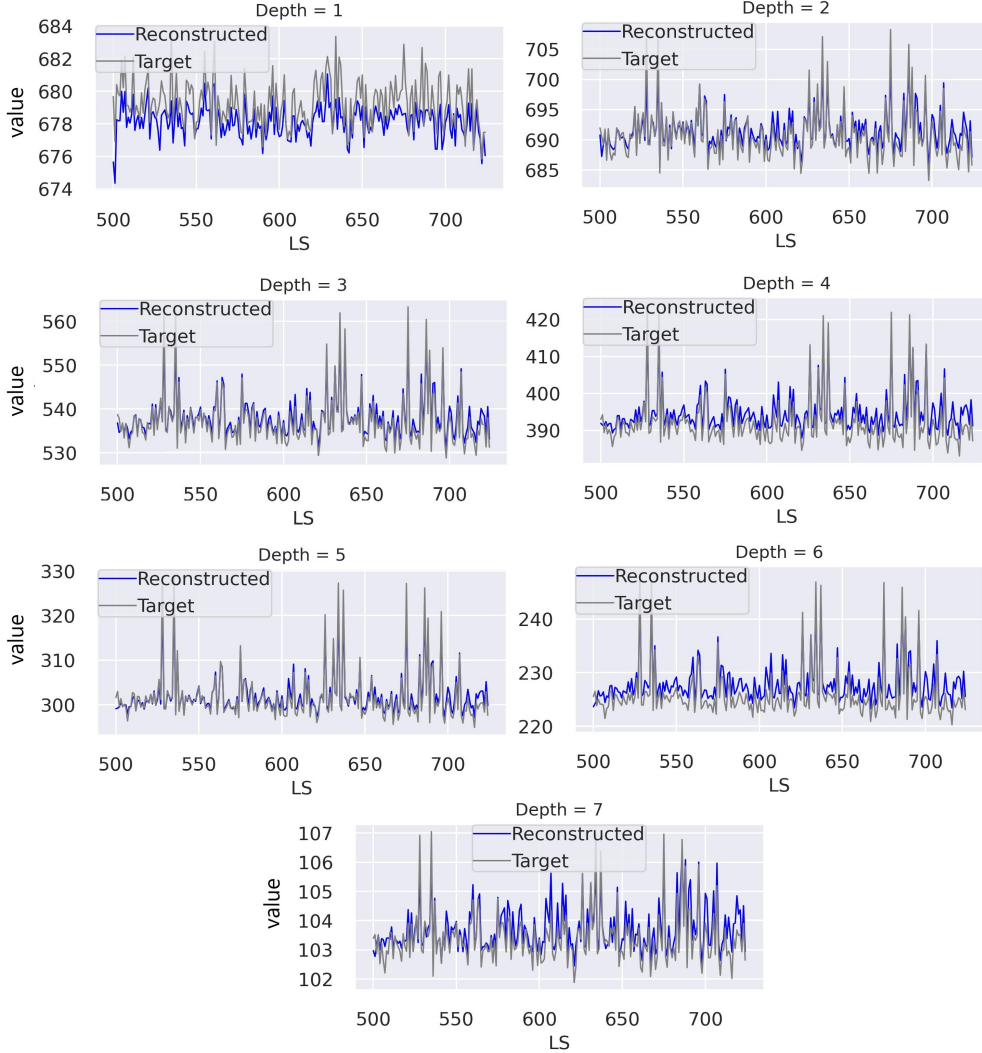
Figure 8: ST digi-occupancy maps reconstruction on samples from the test dataset (*RunId: 325170, LS=[500, 750]*). The figure illustrates the total digi-occupancy per LS across the seven depths ($\hat{\gamma}_l$). The proposed GraphSTAD AE operates on ST γ map data, and we present the curves corresponding to the γ_l per LS only to demonstrate the capability of the AE in handling the fluctuation across the sequence of lumisections.

5.1.1 Detection of Dead and Hot Channels

We have evaluated the AD accuracy on dead–$\gamma_a = 0$–and hot–$\gamma_a = 200\%\gamma_h$–channels on the 10K maps–5K maps for each anomaly type. We have investigated the AD performance on transient channel anomalies that are short-lived in isolated maps (see Table 1) and persisting anomalies that encroach consecutive maps in a time window (see Table 2). The model has achieved high accuracy with good localization of the faulty channels–0.99 precision when capturing 99% of the 335K faulty channels. Time-persistent anomalies are easier to detect–the FPR generally improves by 13%-23% and 28%-40% for the dead and hot anomalies, respectively–compared to the short-lived anomalies on isolated LSs. We have observed that most FPs occur on channels with low expected γ, where the model achieves relatively lower reconstruction accuracy. The performance is not entirely unexpected, as we trained the AE to minimize a global MSE loss function (19). The reconstruction errors become relatively higher for channels with low γ ranges that limit AD effectiveness in distinguishing the anomalies when capturing 99% of the time-persistent dead channels using (8).

We have monitored roughly 31.28M HE sensor channels–of which 335K (1.07%) are simulated abnormal channels–from the 5K maps on the isolated map evaluation in Table 1. The monitored channels grow to 156M with 1.68M (1.07%) anomalies for the evaluation of time-persistent anomalies in Table 2 using time window five maps resulting in 25K maps.

Table 1: AD on dead and hot channel anomalies on isolated digi-occupancy maps.

Anomaly Type	Captured Anomalies	P	R	F_1	FPR
Dead Channel ($R_D = 0\%$)	99%	0.999	0.99	0.995	6.722×10^{-6}
	95%	1.000	0.95	0.974	3.102×10^{-6}
	90%	1.000	0.90	0.947	2.068×10^{-6}
Hot Channel ($R_D = 200\%$)	99%	0.999	0.99	0.994	9.113×10^{-6}
	95%	1.000	0.95	0.974	1.939×10^{-6}
	90%	1.000	0.90	0.947	1.196×10^{-6}

[*] P – Precision, R – Recall, F1 – F1-score, and FPR – False Positive Rate

Table 2: AD on time-persistent dead and hot channel anomalies.

Anomaly Type	Captured Anomalies	P	R	F1	FPR
Dead Channel ($R_D = 0\%$)	99%	0.999	0.99	0.995	7.691×10^{-6}
	95%	1.000	0.95	0.974	2.715×10^{-6}
	90%	1.000	0.90	0.947	1.616×10^{-6}
Hot Channel ($R_D = 200\%$)	99%	0.999	0.99	0.995	5.461×10^{-6}
	95%	1.000	0.95	0.974	1.357×10^{-6}
	90%	1.000	0.90	0.947	7.756×10^{-7}

5.1.2 Detection of Degrading Channels

Table 3 presents the AD accuracy of time-persistent degrading channels simulated with $R_D = [80\%, 60\%, 40\%, 20\%, 0\%]$; the $R_D = 0\%$ corresponds to a dead channel. We injected the generated channel faults into 1K maps for each decay factor. We have monitored around 156M channels–of which 1.74M (1.11%) are abnormal channels–in the total of 25K digi-occupancy maps–5K maps per in the time window. The AD system has demonstrated promising potential in detecting degraded channel anomalies. The FPR to capture 99% of the anomaly is 2.988%, 0.155%, 0.022%, 0.002%, and 0.001% when channels operate at 80%, 60%, 40%, 20%, and 0% of their expected capacity, respectively.

The relatively lower precision at the $R_D = 80\%$ indicates that there are still a few anomalies challenging to catch even though the FPR is very low considering the accurate classification of numerous TN healthy channels (see Fig. 9); The channels operating at $R_D = 80\%$ are mostly inliers–overlapping with the healthy operating ranges–and detecting such anomalies is difficult when the expected γ of the channel is low. The significant improvement of the FPR by 88% and 95% when the amount of the captured anomaly is reduced to 95% and 90%, respectively, demonstrates a small percentage of the channels causes the performance drop at $R_D = 80\%$. Fig. 10 illustrates the overlap regions on the distribution of the reconstruction errors of the healthy and faulty channels at the various R_D values.

5.1.3 Performance Comparison with Benchmark Models

We have quantitatively compared alternative benchmark models to validate the capability of the GraphSTAD (see Fig. 11). The benchmark AE models employ a similar architecture as the GraphSTAD AE but with different layers. The results demonstrate that the integration of the GNN has a significant performance improvement from 1.6 to 3.9 times in the FPR. The temporal models–with RNN–have achieved a 3 to 5-fold boost over the non-temporal spatial AD model when capturing severely degraded channels. For subtle or inlier anomalies–e.g., when channels deteriorate by 20% at $R_D = 80\%$–the GraphSTAD has a substantial 25 times amelioration over the non-temporal model. Incorporating temporal modeling and GNN has enhanced degrading channel detection performance.

Table 3: AD on time-persistent degrading channels.

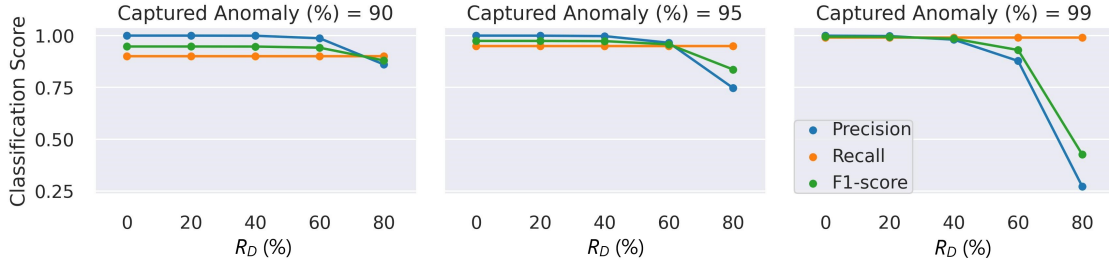| Anomaly Type	Health Rate	FPR (90%)	FPR (95%)	FPR (99%)
Degrading Channel	80%	1.636×10^{-3}	3.614×10^{-3}	2.988×10^{-2}
	60%	1.329×10^{-4}	3.834×10^{-4}	1.550×10^{-3}
	40%	8.405×10^{-6}	2.764×10^{-5}	2.242×10^{-4}
	20%	2.263×10^{-6}	5.173×10^{-6}	2.505×10^{-5}
	0%	9.699×10^{-7}	1.778×10^{-6}	6.142×10^{-6}

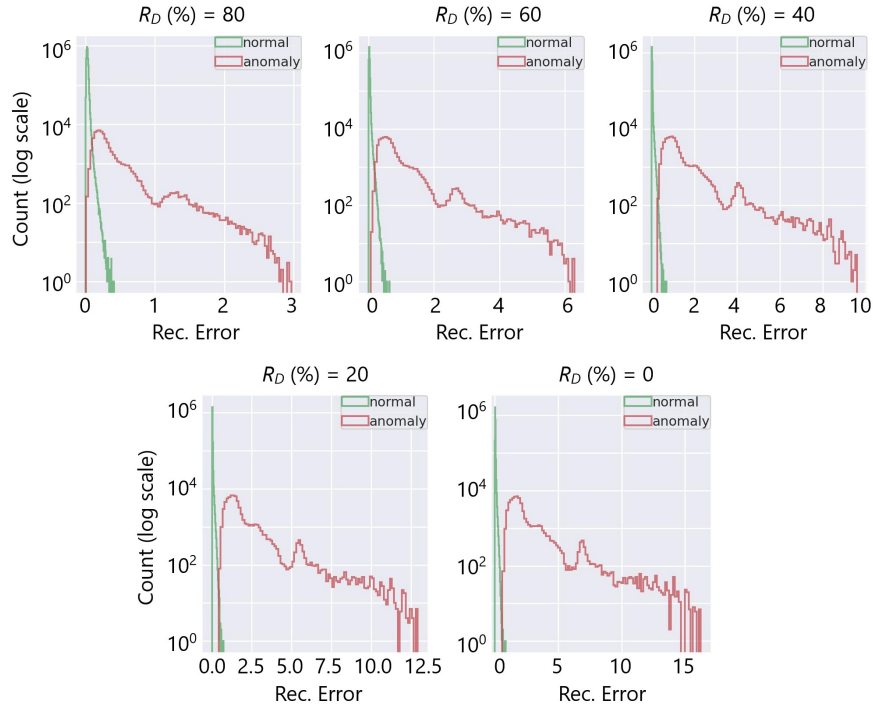Figure 9: AD classification performance on time-persistent degrading channels.

Figure 10: Reconstruction error distribution of healthy and anomalous channels at different health rates. The overlap region decreases substantially as the channel deterioration increases (left to right).

5.1.4 Detection of Real Anomalies in the HCAL

The GraphSTAD system has spotted five real faulty HE channels in collision data *RunId=324841* using the digi-occupancy maps The faulty channels are located at $[i\eta = 17, i\phi = 71, depth = 3]$, $[i\eta = 18, i\phi = 71, depth = 3]$, $[i\eta = 18, i\phi = 71, depth = 4]$, $[i\eta = 18, i\phi = 71, depth = 5]$, and $[i\eta = 28, i\phi = 71, depth = 4]$, and have impacted 52 consecutive LSs (see Fig. 12). Fig. 12 and Fig. 13 illustrate the detected faults fall into the dead channel category except in the last *LS=57* where the channels operated in a degraded state–the γ is lower than expected. Detecting degraded channels is challenging since the γ reading is non-extreme like in dead and hot channels, and the γ drop overlaps with other false down-spikes (see *LS > 57* in Fig. 12). The down-spikes in the digi-occupancy for *LS > 57* are due to non-linearity in the LHC–changes in collision run settings (see Fig. 12b); our normalizing regression model has successfully handled the fluctuation during prepossessing before causing false-positive alerts (see Fig. 12a). Fig. 14 and Fig. 15 portray the spatial anomaly scores during death and degraded status of the faulty channels; the high anomaly scores localized at the faulty channels demonstrate the GraphSTAD AD performance at a channel-level granularity. The existing production DQM system of the CMS–uses rule-based and statistical methods–has also reported these abnormal channels at run-level analysis; the results are only available at the end of the run after analyzing all the LSs for the run Tuura et al. [2010]. Our approach is adaptive to variability in the digi-occupancy maps and provides anomaly localization that detects faulty–including non-extreme degrading–channels per lumisection granularity.
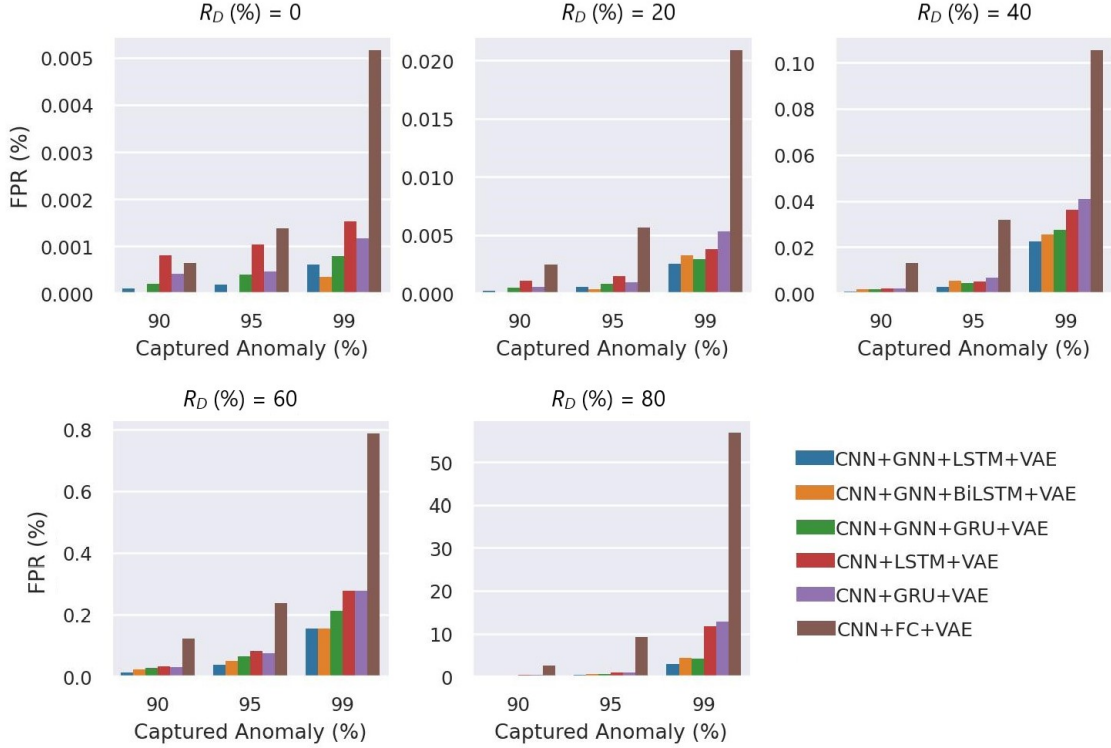
CNN: convolutional neural network, GNN: graph neural network, BiLSTM: bidirectional LSTM, GRU: gated recurrent unit, and VAE: variational AE.

Figure 11: Comparison with benchmark models on detecting time-persistent degrading channels. The GraphSTAD model achieves a significantly lower FPR.
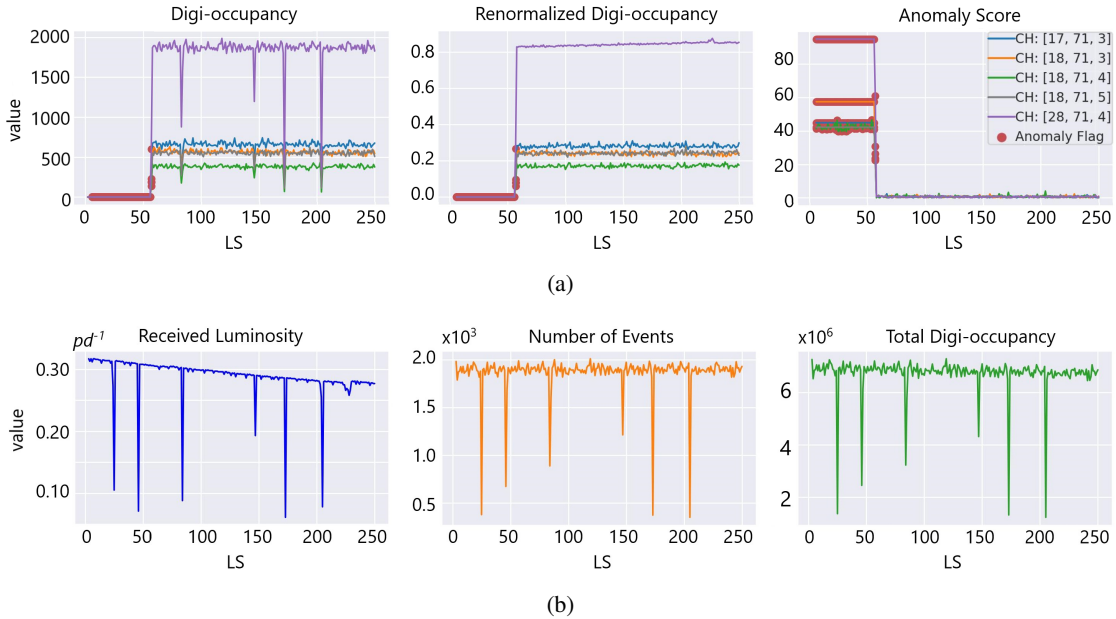
Figure 12: Detected real faulty channels on digi-occupancy maps at *LS=[6, 57]* of *RunId=324841*. (**a**) The digi-occupancy dropped to near zero for the faulty channels (left and middle plots), resulting in high anomaly scores (right). Dead (*LS=[6, 56]*) and degraded channel anomalies (*LS=57*) were captured on the highlighted LSs (red). (**b**) collision run settings and the total digi-occupancy per LS.

(a)

(b)
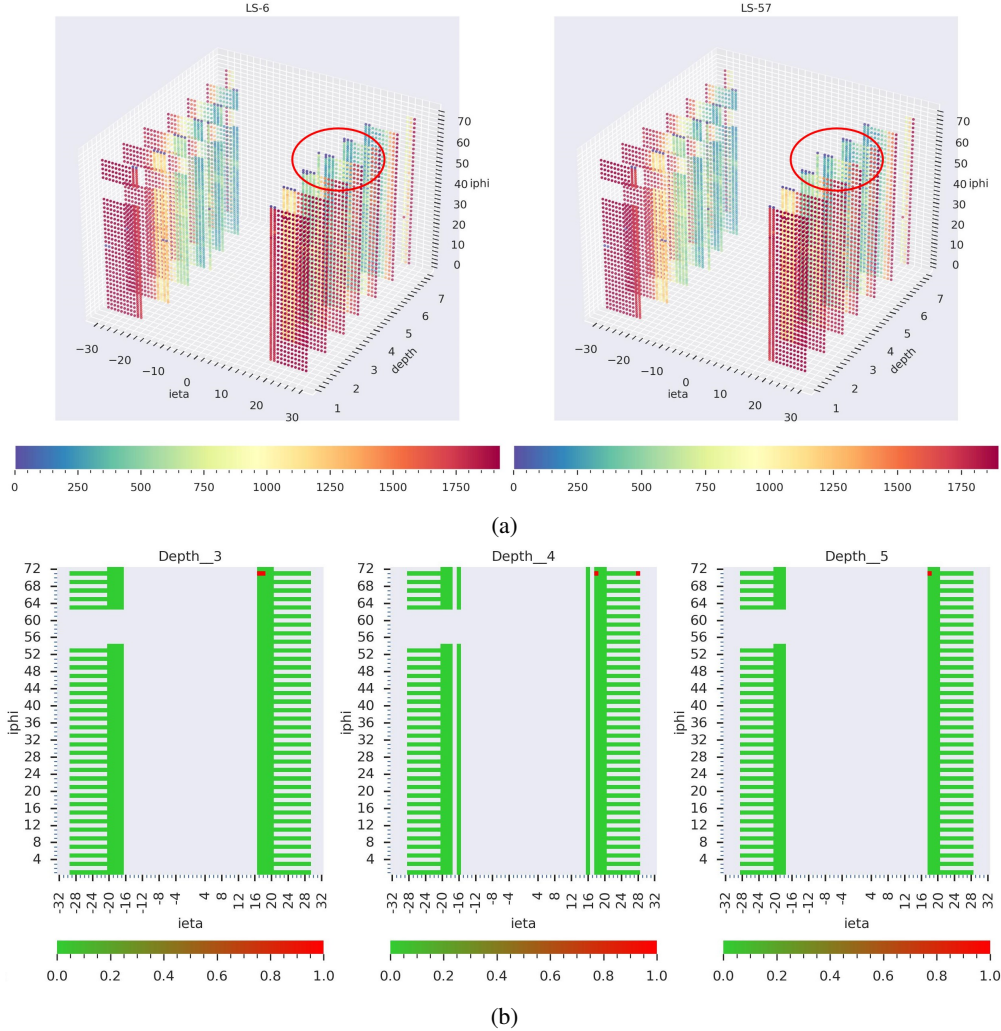
Figure 13: Detection of real faulty channels from *RunId=324841* collision run data. a) the 3D digi-occupancy maps with faulty channels–dead on the left at *LS=6* and degraded on the right at *LS=57*, and b) visualization of the channel anomaly flags on the 2D map per the depth axes–red for anomaly and green for healthy.

5.2 Model Computational Cost

We developed the models with PyTorch and trained them on four GPUs of NVIDIA Tesla V100 SXM3 32GB and Intel(R) Xeon(R) Platinum 8168 CPU 2.70GHz. We utilized a time window $T = 5$ and batch size $B = 8$ for training, and the dimension of a batch is $[B \times T \times N_{i\eta} \times N_{i\phi} \times N_d \times N_f]$. The training time of the GraphSTAD AE model is approximately 45 seconds per epoch. The training iteration epoch 200 achieves good accuracy with a one-cycle learning rate schedule Smith and Topin [2019]. The median inference time of the GraphSTAD on a single GPU is roughly 0.05 seconds with a standard deviation of 0.006 seconds. The integration of the GNN makes the inference relatively slower compared to the benchmark models. The processing cost is, nonetheless, within an acceptable range for the CMS production requirement, as the input digi-occupancy map is generated at each lumisection with a time interval of 23 seconds.

6 Conclusion

Our study presents a semi-supervised anomaly detection system for the hadronic calorimeter's data quality monitoring system using spatio-temporal digi-occupancy maps. We extend the synergy of temporal deep learning developments for the CMS experiment. Our approach addresses modeling challenges, such as digi-occupancy map renormalization, learning non-Euclidean spatial behavior, and degrading channel detection. To overcome these challenges, we propose the
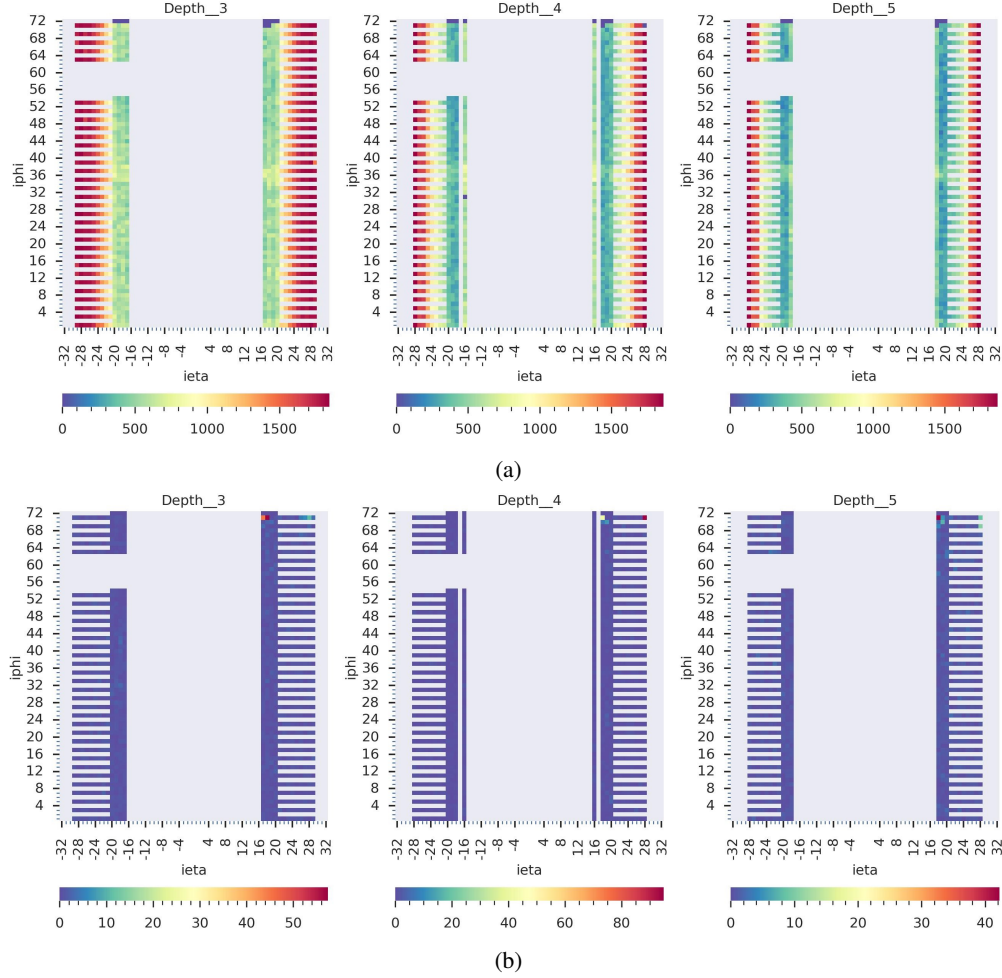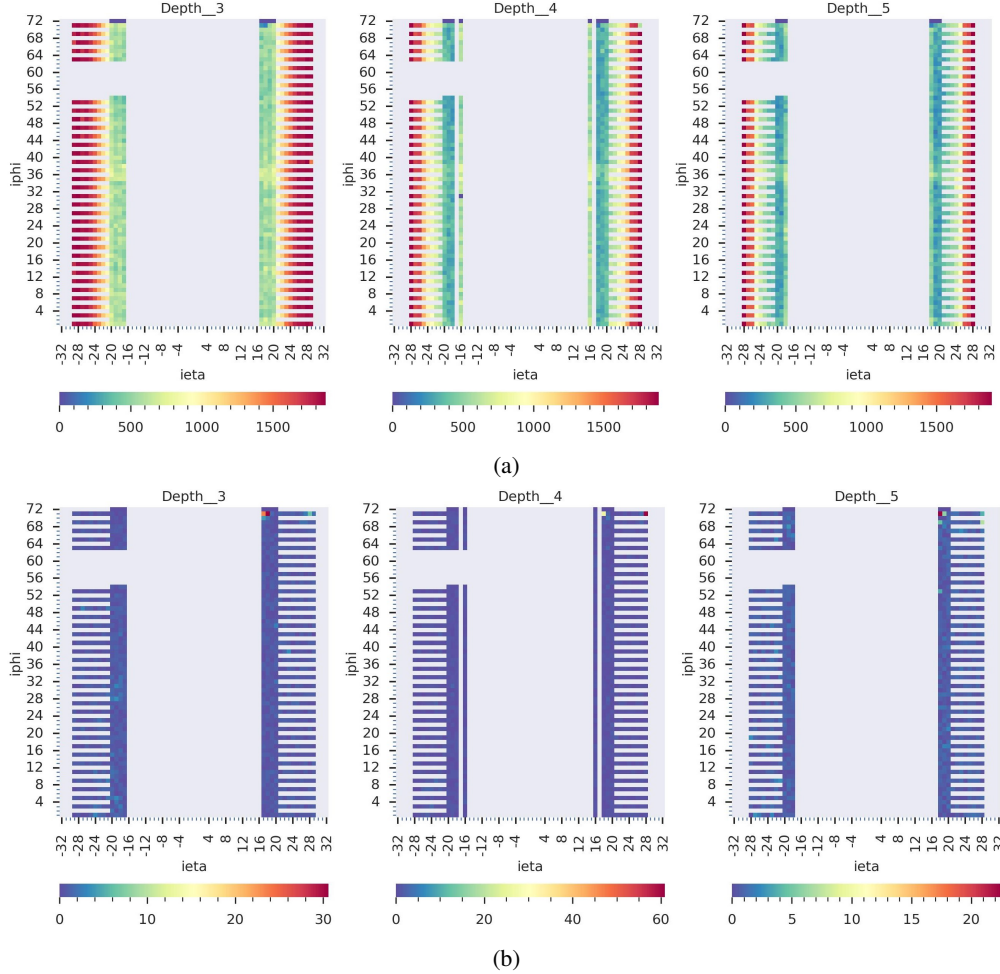
Figure 14: The detected real dead channels at the *LS=6* from the *RunId=324841*: a) the raw 2D digi-occupancy maps at the *depth* axes of the faulty channels, and b) the corresponding anomaly score maps. The GraphSTAD localizes the anomaly scores on the faulty dead channels.

GraphSTAD model, which combines convolutional, graph, and temporal learning networks to capture spatio-temporal behavior and achieve robust localization of anomalies at a channel granularity on high spatial data. The AD performance evaluation has demonstrated the efficacy of the proposed system for channel monitoring. Our proposed AD system will facilitate monitoring and diagnostics of faults in the frontend particle hit sensing hardware and software system of the calorimeter. It will enhance the accuracy and automation of the existing DQM system–providing instant anomaly alerts on a broader range of channel faults in real-time and offline; the improved monitoring of the calorimeter will result in the collection of high-quality physics data. The methods and approaches discussed in this study are domain-agnostic and can be adopted in other spatio-temporal fields–particularly when the data exhibits regular and irregular spatial characteristics.

Acknowledgment

17

(a)

(b)

Figure 15: The detected real degraded channels at the *LS=57* from the *RunId=324841*: a) raw 2D digi-occupancy maps at the *depth* axes of the faulty channels, and b) the corresponding anomaly visualization maps. The GraphSTAD localizes the anomaly scores on the faulty degraded channels with strength proportional to anomaly severity–lower scores in the color bars than the dead channels.

Bibliography

Lyndon Evans and Philip Bryant. LHC machine. *Journal of instrumentation*, 3(08):S08001, 2008.

Rolf-Dieter Heuer. The future of the Large Hadron Collider and CERN. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1961):986–994, 2012.

Virginia Azzolini, Dmitrijus Bugelskis, Tomas Hreus, Kaori Maeshima, Menendez Javier Fernandez, Antanas Norkus, Patrick James Fraser, Marco Rovere, Marcel Andre Schneider, et al. The data quality monitoring software for the CMS experiment at the lhc: past, present and future. In *EPJ Web of Conferences*, volume 214, page 02003. EDP Sciences, 2019.

Lassi Tuura, A Meyer, Ilaria Segoni, and G Della Ricca. CMS data quality monitoring: systems and experiences. In *Journal of Physics: Conference Series*, volume 219, page 072020. IOP Publishing, 2010.

Federico De Guio and The CMS Collaboration. The CMS data quality monitoring software: experience and future prospects. In *Journal of Physics: Conference Series*, volume 513, page 032024. IOP Publishing, 2014.

Virginia Azzolin, Michael Andrews, Gianluca Cerminara, Nabarun Dev, Colin Jessop, Nancy Marinelli, Tanmay Mudholkar, Maurizio Pierini, Adrian Pol, and Jean-Roch Vlimant. Improving data quality monitoring via a partnership of technologies and resources between the CMS experiment at CERN and industry. In *EPJ Web of Conferences*, volume 214, page 01007. EDP Sciences, 2019.

Adrian Alan Pol, Virginia Azzolini, Gianluca Cerminara, Federico De Guio, Giovanni Franzoni, Maurizio Pierini, Filip Sirokỳ, and Jean-Roch Vlimant. Anomaly detection using deep autoencoders for the assessment of the quality of the data acquired by the CMS experiment. In *EPJ Web of Conferences*, volume 214, page 06008. EDP Sciences, 2019a.

Adrian Alan Pol, Gianluca Cerminara, Cécile Germain, Maurizio Pierini, and Agrima Seth. Detector monitoring with artificial neural networks at the CMS experiment at the CERN Large Hadron Collider. *Computing and Software for Big Science*, 3(1):3, 2019b.

Oleksandr Viazlo and The CMS Collaboration. Non-uniformity in HE digi-occupancy distributions. Private communications, 2022.

Maciej Wielgosz, Matej Mertik, Andrzej Skoczeń, and Ernesto De Matteis. The model of an anomaly detector for hilumi LHC magnets based on recurrent neural networks and adaptive quantization. *Engineering Applications of Artificial Intelligence*, 74: 166–185, 2018.

Dan Xu, Yan Yan, Elisa Ricci, and Nicu Sebe. Detecting anomalous events in videos by learning deep representations of appearance and motion. *Computer Vision and Image Understanding*, 156:117–127, 2017.

Yunpeng Chang, Zhigang Tu, Wei Xie, Bin Luo, Shifu Zhang, Haigang Sui, and Junsong Yuan. Video anomaly detection with spatio-temporal dissociation. *Pattern Recognition*, 122:108213, 2022.

Weixin Luo, Wen Liu, Dongze Lian, Jinhui Tang, Lixin Duan, Xi Peng, and Shenghua Gao. Video anomaly detection with sparse coding inspired deep neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(3):1070–1084, 2019.

Mahmudul Hasan, Jonghyun Choi, Jan Neumann, Amit K Roy-Chowdhury, and Larry S Davis. Learning temporal regularity in video sequences. In *Proceedings of Computer Vision and Pattern Recognition*, pages 733–742. IEEE, 2016.

Peng Wu, Jing Liu, Mingming Li, Yujia Sun, and Fang Shen. Fast sparse coding networks for anomaly detection in videos. *Pattern Recognition*, 107:107515, 2020.

Daniel Hsu. Anomaly detection on graph time series. *arXiv preprint arXiv:1708.02975*, 2017.

Waseem Ullah, Amin Ullah, Tanveer Hussain, Zulfiqar Ahmad Khan, and Sung Wook Baik. An efficient anomaly recognition framework using an attention residual LSTM in surveillance videos. *Sensors*, 21(8):2811, 2021.

Jingtao Hu, En Zhu, Siqi Wang, Xinwang Liu, Xifeng Guo, and Jianping Yin. An efficient and robust unsupervised anomaly detection method using ensemble random projection in surveillance videos. *Sensors*, 19(19):4145, 2019.

Zorana Banković, David Fraga, José M Moya, and Juan Carlos Vallejo. Detecting unknown attacks in wireless sensor networks that contain mobile nodes. *Sensors*, 12(8):10834, 2012.

Gangqiang Zhang, Wei Zheng, Wenjie Yin, and Weiwei Lei. Improving the resolution and accuracy of groundwater level anomalies using the machine learning-based fusion model in the North China plain. *Sensors*, 21(1):46, 2020.

Gowtham Atluri, Anuj Karpatne, and Vipin Kumar. Spatio-temporal data mining: a survey of problems and methods. *ACM Computing Surveys*, 51(4):1–41, 2018.

Leyan Deng, Defu Lian, Zhenya Huang, and Enhong Chen. Graph convolutional adversarial networks for spatiotemporal anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems*, 33(6):2416–2428, 2022.

Leo Tišljarić, Sofia Fernandes, Tonči Carić, and João Gama. Spatiotemporal road traffic anomaly detection: a tensor-based approach. *Applied Sciences*, 11(24):12017, 2021.

Lin Jiang, Hang Xu, Jinhai Liu, Xiangkai Shen, Senxiang Lu, and Zhan Shi. Anomaly detection of industrial multi-sensor signals based on enhanced spatiotemporal features. *Neural Computing and Applications*, pages 1–13, 2022.

David Ahmedt-Aristizabal, Mohammad Ali Armin, Simon Denman, Clinton Fookes, and Lars Petersson. Graph-based deep learning for medical diagnosis and analysis: past, present and future. *Sensors*, 21(14):4758, 2021.

Javier Duarte and Jean-Roch Vlimant. Graph neural networks for particle tracking and reconstruction. In *Artificial Intelligence for High Energy Physics*, pages 387–436. World Scientific, 2022.

Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. Graph neural networks in particle physics. *Machine Learning: Science and Technology*, 2(2):021001, 2020.

Shah Rukh Qasim, Jan Kieseler, Yutaro Iiyama, and Maurizio Pierini. Learning representations of irregular particle-detector geometry with distance-weighted graph networks. *The European Physical Journal C*, 79(7):1–11, 2019.

J Arjona Martínez, Olmo Cerri, Maria Spiropulu, JR Vlimant, and M Pierini. Pileup mitigation at the Large Hadron Collider with graph neural networks. *The European Physical Journal Plus*, 134(7):333, 2019.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*, 2013.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Ettore Focardi. Status of the CMS detector. *Physics Procedia*, 37:119–127, 2012.

Nadja Strobbe. The upgrade of the CMS Hadron Calorimeter with Silicon photomultipliers. *Journal of Instrumentation*, 12(1): C01080, 2017.

Valdas Rapsevicius, CMS DQM Group, et al. CMS run registry: data certification bookkeeping and publication system. In *Journal of Physics: Conference Series*, volume 331, page 042038. IOP Publishing, 2011.

Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: a survey. *arXiv preprint arXiv:1901.03407*, 2019.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Mulugeta Weldezgina Asres, Grace Cummings, Pavel Parygin, Aleko Khukhunaishvili, Maria Toms, Alan Campbell, Seth I Cooper, David Yu, Jay Dittmann, and Christian W Omlin. Unsupervised deep variational model for multivariate sensor anomaly detection. In *International Conference on Progress in Informatics and Computing*, pages 364–371. IEEE, 2021.

Yifan Guo, Weixian Liao, Qianlong Wang, Lixing Yu, Tianxi Ji, and Pan Li. Multidimensional time series anomaly detection: A GRU-based gaussian mixture variational autoencoder approach. In *Asian Conference on Machine Learning*, pages 97–112. Proceedings of Machine Learning Research, 2018.

Jinwon An and Sungzoon Cho. Variational autoencoder based anomaly detection using reconstruction probability. *Special Lecture on IE*, 2(1):1–18, 2015.

Gavneet Singh Chadha, Arfyan Rabbani, and Andreas Schwung. Comparison of semi-supervised deep neural networks for anomaly detection in industrial processes. In *17th International Conference on Industrial Informatics*, volume 1, pages 214–219. IEEE, 2019.

Matthew D Zeiler, Dilip Krishnan, Graham W Taylor, and Rob Fergus. Deconvolutional networks. In *Proceedings of Computer Vision and Pattern Recognition*, pages 2528–2535. IEEE, 2010.

Leslie N Smith and Nicholay Topin. Super-convergence: very fast training of neural networks using large learning rates. In *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, volume 11006, pages 369–386. SPIE, 2019.

The CMS-HCAL Collaboration

A. Gevorgyan[1], A. Petrosyan[1], A. Tumasyan[1], G.A. Alves[2], C. Hensel[2], W.L. Aldá Júnior[3], W. Carvalho[3], J. Chinellato[3,f], C. De Oliveira Martins[3], D. Matos Figueiredo[3], C. Mora Herrera[3], H. Nogima[3], W.L. Prado Da Silva[3], E.J. Tonelli Manganote[3], A. Vilela Pereira[3], M. Finger[4], M. Finger Jr.[4], G. Adamov[5], Z. Tsamalaidze[5,g], K. Borras[6,y], A. Campbell[6], F. Engelke[6,y], D. Krücker[6], I. Martens[6], L. Wiens[6,y], M. Csanád[7], A. Feherkuti[7], S. Lökös[7,v], G. Pásztor[7], O. Surányi[7], G.I. Veres[7], V. Hegde[8], K. Kothekar[8], S. Pandey[8], S. Sharma[8], S.B. Beri[9], B. Bhawandeep[9], R. Chawla[9], A. Kalsi[9], A. Kaur[9], M. Kaur[9], G. Walia[9], S. Bhattacharya[10], S. Ghosh[10], S. Nandan[10], A. Purohit[10], M. Sharan[10], S. Banerjee[11], S. Bhattacharya[11], S. Chatterjee[11], P. Das[11], M. Guchait[11], S. Jain[11], S. Kumar[11], M. Maity[11], G. Majumder[11], K. Mazumdar[11], M. Patil[11], T. Sarkar[11], S. Sekmen[12,y], A. Juodagalvis[13], D. Agyel[14], F. Boran[14], S. Damarseckin[14], Z.S. Demiroglu[14], F. Dölek[14], I. Dumanoglu[14,ee], E. Eskut[14], G. Gokbulut[14], Y. Guler[14,ff], E. Gurpinar Guler[14,ff], C. Işik[14], E.E. Kangal[14], O. Kara[14], A. Kayis Topaksu[14], U. Kiminsu[14], G. Onengut[14], K. Ozdemir[14,gg], E. Pinar[14], A. Polatoz[14], A.E. Simsek[14], B. Tali[14,hh], U.G. Tok[14], S. Turkcapar[14], E. Uslan[14], I.S. Zorbakir[14], B. Bilin[15,y], G. Karapinar[15,ii], A. Murat Guler[15], K. Ocalan[15,jj], M. Yalvac[15,kk], M. Zeyrek[15], B. Akgun[16], I.O. Atakisi[16,ll], E. Gülmez[16], M. Kaya[16,ll], O. Kaya[16,mm], S. Tekten[16,nn], E.A. Yetkin[16,dd], T. Yetkin[16,qq], A. Cakir[17], K. Cankocak[17,ee], S. Sen[17,oo], O. Aydilek[18], S. Cerci[18,hh], B. Hacisahinoglu[18], I. Hos[18,pp], B. Isildak[18,qq], B. Kaynak[18], S. Ozkorucuklu[18], O. Potok[18], H. Sert[18], C. Simsek[18], D. Sunar Cerci[18,hh], C. Zorbilmez[18], A. Boyarintsev[19], B. Grynyov[19], L. Levchuk[20], V. Popov[20], P. Sorokin[20], H. Flacher[21], S. Abdullin[22], B. Caraway[22], J. Dittmann[22], K. Hatakeyama[22], A.R. Kanuganti[22], B. McMaster[22], M. Saunders[22], J. Wilson[22], A. Buccilli[23,q], P. Bunin[23,z], S.I. Cooper[23], C. Henderson[23,l], C.U. Perez[23], P. Rumerio[23,t], C. Cosby[24], Z. Demiragli[24], D. Gastler[24], E. Hazen[24], J. Rohlf[24], M. Hadley[25], U. Heintz[25], T. Kwon[25], E. Laird[25], G. Landsberg[25], K.T. Lau[25], X. Yan[25], D. Yu[25,cc], Z. Mao[25], J.W. Gary[26], G. Karapostoli[26,bb], O.R. Long[26], R. Bhandari[27], R. Heller[27], D. Stuart[27], J. Yoo[27,j], Y. Chen[28,n], J. Duarte[28], J.M. Lawhorn[28], M. Spiropulu[28], A. Apresyan[29], A. Apyan[29,c], S. Banerjee[29,d], F. Chlebana[29], Y. Feng[29], J. Freeman[29], D. Green[29], K.H.M. Kwok[29], J. Hirschauer[29], U. Joshi[29], D. Lincoln[29], S. Los[29], C. Madrid[29], N. Pastika[29], K. Pedro[29], W.J. Spalding[29], S. Tkaczyk[29], S. Linn[30], P. Markowitz[30], V. Hagopian[31], T. Kolberg[31], G. Martinez[31], O. Viazlo[31], M. Hohlmann[32], R. Kumar Verma[32], D. Noonan[32], F. Yumiceva[32,e], M. Alhusseini[33], B. Bilki[33], D. Blend[33], K. Dilsiz[33,rr], L. Emediato[33], R.P. Gandrajula[33], M. Herrmann[33], O.K. Köseyan[33], J.-P. Merlo[33], A. Mestvirishvili[33,aa], M. Miller[33], H. Ogul[33,ss], Y. Onel[33], A. Penzo[33], D. Southwick[33], E. Tiras[33,tt], J. Wetzel[33], A. Al-bataineh[34,s], J. Bowen[34,o], C. Le Mahieu[34], J. Marquez[34], W. McBrayer[34], M. Murray[34], M. Nickel[34], S. Popescu[34,r], C. Smith[34], Q. Wang[34], K. Kaadze[35], D. Kim[35], Y. Maravin[35], A. Mohammadi[35,d], J. Natoli[35], D. Roy[35], L.K. Saini[35,f], E. Adams[36], A. Baden[36], O. Baron[36], A. Belloni[36], A. Bethani[36], Y-M Chen[36], S.C. Eno[36], C. Ferraioli[36,i], T. Grassi[36], N.J. Hadley[36], R.G. Kellogg[36], T. Koeth[36], Y. Lai[36], S. Lascio[36], A.C. Mignerey[36], S. Nabili[36], C. Palmer[36], C. Papageorgakis[36], M. Seidel[36,u], L. Wang[36], K. Wong[36], M. D'Alfonso[37], M. Hu[37], B. Crossman[38], J. Hiltbrand[38], M. Krohn[38], J. Mans[38], M. Revering[38], N. Strobbe[38], A. Heering[39], Y. Musienko[39,z], R. Ruchti[39], M. Wayne[39], W. Chung[40], G. Kopp[40], K. Mei[40], C. Tully[40], A. Bodek[41], P. de Barbaro[41], C. Fallon[41], M. Galanti[41,y], A. Garcia-Bellido[41], A. Khukhunaishvili[41], C-L Tan[41], R. Taus[41], D. Vishnevskiy[41], M. Zielinski[41], B. Chiarito[42], J.P. Chou[42], S.A. Thayil[42], H. Wang[42], N. Akchurin[43], J. Damgov[43], F. De Guio[43,w], S. Kunori[43], K. Lamichhane[43], S.W. Lee[43], T. Mengke[43], S. Muthumuni[43], S. Undleeb[43], I. Volobouev[43], Z. Wang[43], A. Whitbeck[43], G. Cummings[44], S. Goadhouse[44], J. Hakala[44], R. Hirosky[44], D. Winn[45], V. Alexakhin[46], V. Andreev[46], Y. Andreev[46], M. Azarkin[46], A. Belyaev[46], S. Bitioukov[46], E. Boos[46], O. Bychkova[46], M. Chadeeva[46], V. Chekhovsky[46], R. Chistov[46], M. Danilov[46], A. Demianov[46], A. Dermenev[46], M. Dubinin[46,k], L. Dudko[46], D. Elumakhov[46], V. Epshteyn[46], Y. Ershov[46], A. Ershov[46], V. Gavrilov[46], I. Golutvin[46,a†], A. Gribushin[46], A. Kalinin[46,m], A. Kaminskiy[46], A. Karneyeu[46], L. Khein[46], M. Kirakosyan[46], V. Klyukhin[46], O. Kodolova[46,b], V. Krychkine[46], A. Kurenkov[46], A. Litomin[46], N. Lychkovskaya[46], V. Makarenko[46], P. Mandrik[46], P. Moisenz[46,a†], S. Obraztsov[46], A. Oskin[46], P. Parygin[46,x], V. Petrov[46], S. Petrushanko[46], S. Polikarpov[46], E. Popova[46,x], V. Rusinov[46], R. Ryutin[46], V. Savrin[46], D. Selivanova[46], V. Smirnov[46], A. Snigirev[46], A. Sobol[46], A. Stepennov[46,p], E. Tarkovskii[46], A. Terkulov[46], D. Tlisov[46,a†], I. Tlisova[46], R. Tolochek[46], M. Toms[46,h], A. Toropin[46], S. Troshin[46], A. Volkov[46], B. Yuldashev[46], A. Zarubin[46], A. Zhokin[46]

[1]Yerevan Physics Institute, Yerevan, Armenia
[2]Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro, Brazil
[3]Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
[4]Charles University, Prague, Czech Republic
[5]Georgian Technical University, Tbilisi, Georgia
[6]Deutsches Elektronen-Synchrotron, Hamburg, Germany
[7]MTA-ELTE Lendület CMS Particle and Nuclear Physics Group, Eötvös Loránd University, Budapest, Hungary
[8]Indian Institute of Science Education and Research (IISER), Pune, India
[9]Panjab University, Chandigarh, India
[10]Saha Institute of Nuclear Physics, HBNI, Kolkata, India
[11]Tata Institute of Fundamental Research-B, Mumbai, India
[12]Kyungpook National University, Daegu, Korea
[13]Vilnius University, Vilnius, Lithuania
[14]Çukurova University, Physics Department, Science and Art Faculty, Adana, Turkey
[15]Middle East Technical University, Physics Department, Ankara, Turkey
[16]Bogazici University, Istanbul, Turkey

[17]Istanbul Technical University, Istanbul, Turkey

[18]Istanbul University, Istanbul, Turkey

[19]Institute for Scintillation Materials of National Academy of Science of Ukraine, Kharkiv, Ukraine

[20]National Science Centre, Kharkiv Institute of Physics and Technology, Kharkiv, Ukraine

[21]University of Bristol, Bristol, United Kingdom

[22]Baylor University, Waco, Texas, USA

[23]The University of Alabama, Tuscaloosa, Alabama, USA

[24]Boston University, Boston, Massachusetts, USA

[25]Brown University, Providence, Rhode Island, USA

[26]University of California, Riverside, Riverside, California, USA

[27]University of California, Santa Barbara - Department of Physics, Santa Barbara, California, USA

[28]California Institute of Technology, Pasadena, California, USA

[29]Fermi National Accelerator Laboratory, Batavia, Illinois, USA

[30]Florida International University, Miami, USA

[31]Florida State University, Tallahassee, Florida, USA

[32]Florida Institute of Technology, Melbourne, Florida, USA

[33]The University of Iowa, Iowa City, Iowa, USA

[34]The University of Kansas, Lawrence, Kansas, USA

[35]Kansas State University, Manhattan, Kansas, USA

[36]University of Maryland, College Park, Maryland, USA

[37]Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[38]University of Minnesota, Minneapolis, Minnesota, USA

[39]University of Notre Dame, Notre Dame, Indiana, USA

[40]Princeton University, Princeton, New Jersey, USA

[41]University of Rochester, Rochester, New York, USA

[42]Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA

[43]Texas Tech University, Lubbock, Texas, USA

[44]University of Virginia, Charlottesville, Virginia, USA

[45]Fairfield University, Fairfield, USA

[46]Authors affiliated with an institute or an international laboratory covered by a cooperation agreement with CERN.

[a]†Deceased

[b]Also at Yerevan State University, Yerevan, Armenia

[c]Now at Brandeis University, Waltham, USA

[d]Now at University of Wisconsin-Madison, Madison, USA

[e]Now at Northrop Grumman, Linthicum Heights, USA

[f]Now at Gallagher Basset, Schaumburg, USA

[g]Also at Tbilisi State University, Tbilisi, Georgia

[h]Now at Karlsruhe Institute of Technology, Karlsruhe, Germany

[i]Now at Windfall Data, Novato, USA

[j]Now at Korea University, Seoul, Korea

[k]Also at California Institute of Technology, Pasadena, California, USA

[l]Now at University of Cincinnati, Cincinnati, USA

[m]Now at University of Maryland, College Park, Maryland, USA

[n]Now a Massachusetts Institute of Technology, Cambridge, USA

[o]Now at Baker University, Baldwin City, USA

[p]Now at University of Cyprus , Cyprus

[q]Now at Bond, San Francisco, USA

[r]Also at IFIN-HH, Bucharest, Romania

[s]Now at Yarmouk University, Irbid, Jordan

[t]Also at Università di Torino, Torino, Italy

[u]Now at Riga Technical University, Riga, Latvia

[v]Also at Karoly Robert Campus, MATE Institute of Technology, Gyongyos, Hungary

[w]Now at INFN Sezione di Milano-Bicocca, Milano, Italy

[x]Now at University of Rochester, Rochester, New York, USA

[y]Also at CERN, European Organization for Nuclear Research, Geneva, Switzerland

[z]Also at an institute or an international laboratory covered by a cooperation agreement with CERN

[aa]Also at Georgian Technical University, Tbilisi, Georgia

[bb]Now at National Technical University of Athens, Greece

[cc]Now at University of Nebraska, USA

[dd]Also at Istanbul Bilgi University, Istanbul, Turkey

[ee]Also at Near East University, Research Center of Experimental Health Science, Mersin, Turkey

[ff]Also at Konya Technical University, Konya, Turkey

[gg]Also at Izmir Bakircay University, Izmir, Turkey

[hh] Also at Adiyaman University, Adiyaman, Turkey
[ii] Also at Istanbul Gedik University, Istanbul, Turkey
[jj] Also at Necmettin Erbakan University, Konya, Turkey
[kk] Also at Bozok Universitetesi Rektörlügü, Yozgat, Turkey
[ll] Also at Marmara University, Istanbul, Turkey
[mm] Also at Milli Savunma University, Istanbul, Turkey
[nn] Also at Kafkas University, Kars, Turkey
[oo] Also at Hacettepe University, Ankara, Turkey
[pp] Also at Istanbul University - Cerrahpasa, Faculty of Engineering, Istanbul, Turkey
[qq] Also at Yildiz Technical University, Istanbul, Turkey
[rr] Also at Bingol University, Bingol, Turkey
[ss] Also at Sinop University, Sinop, Turkey
[tt] Also at Erciyes University, Kayseri, Turkey