
USING NEURAL NETWORKS TO PREDICT RADIATION DAMAGE TO LEAD TUNGSTATE CRYSTALS AT THE CERN LHC

Buyun Liang^{1,†}, Bhargav Joshi^{2,†}, Taihui Li^{1,†}, Roger Rusack^{2,†}, Ju Sun^{1,†}

¹ Department of Computer Science & Engineering, University of Minnesota, Minneapolis, USA

² Department of Physics, University of Minnesota, Minneapolis, USA

[†] These authors contributed equally to this work
{liang664, joshib, lixx5027, rusack, jusun}@umn.edu

ABSTRACT

The 72,000 lead tungstate crystals in CMS experiment at the CERN Large Hadron Collider are used to measure the energy of electrons and photons produced in the proton-proton collisions. The optical transparency of the crystals degrades slowly with radiation dose due to the beam-beam collisions. The transparency of each crystal is monitored with a laser monitoring system that tracks changes in the optical properties of the crystals due to radiation from the collision products. Predicting the optical transparency of the crystals, both in the short term and in the long term, is a critical question for the CMS experiment. We describe here the public data release, following FAIR principles [1], of the crystal monitoring data collected by the CMS Collaboration between 2016 and 2018. Besides describing the dataset and its access, the problems that can be addressed with it are described, as well as an example solution based on a Long Short Term Memory neural network developed to predict future behaviour of the crystals.

Keywords Time Series Prediction · Machine Learning · FAIR Data · LSTM

1 Introduction

The Compact Muon Solenoid (CMS) experiment at the CERN Large Hadron Collider (LHC) is designed to detect and measure particles produced in the proton-proton collisions at the LHC [2]. One of the components of the CMS detector is an electromagnetic calorimeter (ECAL) made of lead tungstate crystals that is used to measure the energy of electrons and photons produced in the collisions. This was the principle detector used to detect the Higgs boson in the two-photon decay channel in the discovery of the Higgs boson in 2012 [3], in the precision measurement of the Higgs boson mass [4], and in many other measurements made by the CMS Collaboration. Changes in the crystal transparency caused by radiation from the collisions in the LHC lead to an instability on the electron and photon response of the crystals that needs to be tracked and corrected for. Predicting these changes both in the short term and in the long term is a critical question for the CMS experiment. We describe here the public release following FAIR principles [5] of the crystal monitoring data collected by CMS between 2016 and 2018. Besides describing the dataset and its access, and the problem to be addressed with it, we provide links to an example solution based on a Long Short-Term Memory neural network, developed to predict future behaviour of the crystals.

1.1 Electromagnetic Calorimeter (ECAL)

There are 75,848 individual lead tungstate crystals in the ECAL, of these 62,000 are arranged in a barrel surrounding the interaction point where the beams collide, and the barrel is capped by two endcap calorimeters, each consisting of 7,324 crystals [6]. The crystals in CMS measure approximately 22 cm in length with a 2×2 cm² section and weigh ≈ 1.1 kg. Lead tungstate crystals are optically transparent and emit a short pulse of light when they absorb ionizing radiation. Due to the intense radiation while the LHC is in operation, the optical transparency of the crystals degrades over time. This degradation is due to the creation by the radiation of atomic-level impurities in the crystal that act as color centers [7] that absorb light propagating through the crystal. Deep well impurities are stable and persist for years, and the radiation damage is permanent, and shallow impurities are metastable and are short-lived. During beam-beam

collisions the optical transmission is reduced and partially recovers as the meta-stable states decay during periods when there is no radiation.

When the LHC is operating the beam-beam collisions are continuous, with a 'fill' generally lasting for approximately 18 hours, with a four to six hour interval between fills. There are longer intervals when there are no beams for maintenance of the machine and the detectors, which last on the order of a day, and every year there are several-month-long shut downs. During operation the light output of the crystals is reduced as the color centers are created, and between periods of operation the crystals recover as the metastable states decay.

Since detailed knowledge of the crystal's light output due to ionizing radiation is essential for the physics measurements, the transparency of each crystal is carefully monitored with a laser monitoring system which injects pulses of light into each crystal at intervals of approximately every 40 minutes. Predicting the optical transparency of the crystals both in the days and years ahead is a critical question for the CMS experiment. We describe here the public release following FAIR principles of the crystal monitoring data collected by CMS between 2016 and 2018. Besides describing the dataset and its access, and the problem to be addressed with it, we provide links to an example solution based on Long Short-Term Memory neural network, developed to predict future behaviour of the crystals.

The principle of operation of the crystal calorimeter, which is designed to detect electrons and photons with energies of 1 GeV or more, is as follows: When a high energy photon (> 100 MeV) is incident on the crystal electron and positron pairs are produced, these in turn interact with the atomic nuclei and radiate (bremsstrahlung) photons. These in turn produce more electron-positron pairs, which in turn radiate photons, though always with less energy. This sets up a cascade where photons and electrons-positrons pairs are produced that continues until the photons have insufficient energy to create an electron-positron pair (< 1 MeV). During this process the electrons and positrons produced propagate in the crystal and ionize the atoms causing scintillation light (optical photons) to be produced that can be detected with a photodetector coupled to the crystal. The operating principle of the CMS ECAL is that the amount of ionizing radiation is linearly proportional to the energy of the incident electron or photon, and hence the light output is proportional to the energy of the incident particle.

The cascade is a stochastic process, thus the energy measured with the crystals is measured with a resolution that can be parameterized as:

$$\frac{\sigma_E}{E} = \frac{S\%}{\sqrt{E}} \oplus \frac{N\%}{E} \oplus C\% \quad (1)$$

where the first term on the right is the *stochastic* term, which is the contribution to the resolution from the natural fluctuations in the cascade; second is the *noise* term due electronic noise, and third is the *constant* term that accounts for energy leakage and other signal losses. The parameters, S , N and C have been measured with a prototype in a dedicated particle beam. For electrons with p_T larger than 10 GeV, the energy resolution is better than 1%.

The radiation incident on the crystals from beam-beam collisions, which reduces their transparency, is not uniform across the detector, and is approximately proportional to the direct distance, radius, of the crystals from the beams. Thus the degradation is similar for crystals in the barrel ECAL, while it increases towards the center of the endcaps. A convenient approximation to the production angle is the pseudorapidity (η) [6], which is equal to zero at the center of the barrel and increases to ± 3.0 at the smallest radius part of the endcaps. This difference in the level of radiation damage in the crystals due to different radiation levels can be seen in Figure 1.

2 Data Challenge

The details of the color center formation under radiation are poorly understood. It is thought that they are primarily determined by the atomic-level defects in the crystals when they were first grown and, thus, determined by the individual history of each crystal. From this it follows that the best way to predict their future behavior can be learnt from prior response to radiation, or by examining past behavior under radiation of a crystal, one can learn how that crystal will behave in the future.

The interest of the CMS collaboration is to predict how the ensemble of crystals will respond to future irradiation cycles with damage and recovery. In particular there is interest in predicting by how much a group of crystals will have recovered from irradiation after a duration of the order one day. Another point of interest is to predict the performance of the crystals in the barrel calorimeter – the endcaps will be replaced in 2029 – at the end of operation of the High-Luminosity LHC in 2040.

In the example given below the short-term problem is tackled. The long term problem is an open challenge.

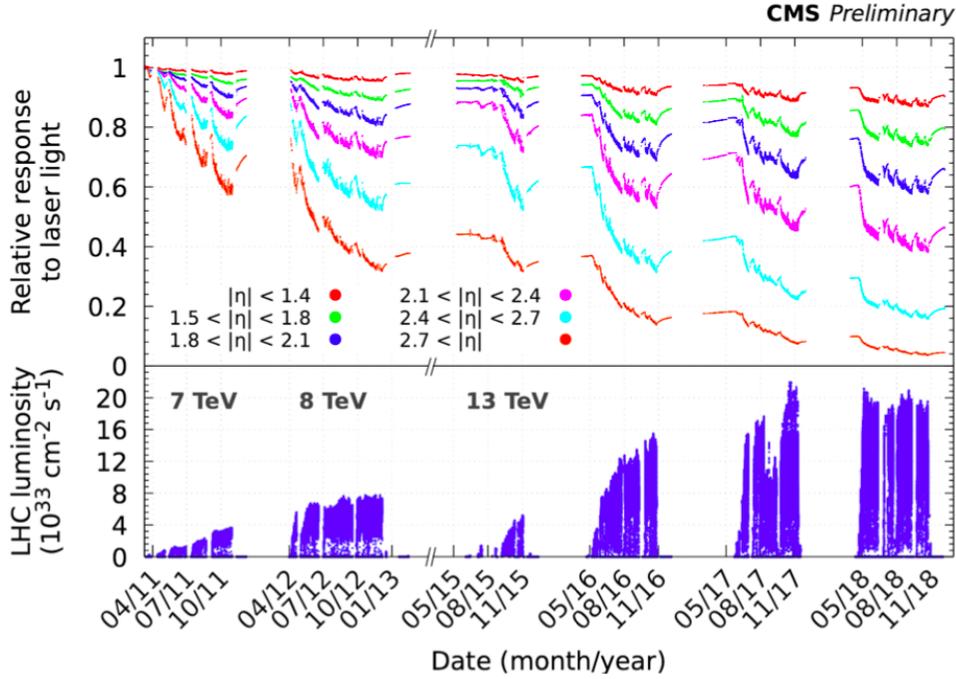


Figure 1: Relative response of crystal to laser light in different η regions between the start of operations in 2011 and 2018. The crystals in barrel ($|\eta| < 1.44$) all have a relatively small light loss, while the crystals closest to the beam in endcaps ($|\eta| > 2.7$) show a large degradation in transparency.

3 Datasets

The response to the injected laser for every crystal taken is measured approximately every 40 minutes and stored in a offline database. The data corresponding to the laser response of all 75848 crystals from 2016 through 2018 was extracted from the database. Each entry has a timestamp corresponding to the time when the measurement was taken and the current intensity of the beam-beam collisions – the instantaneous luminosity – provided by the CMS beam radiation, instrumentation, and luminosity (BRIL) group. This is a measure of the radiation level as the crystals.

The dataset consists of the following elements corresponding to each measurement.

- **xtal_id**: Crystal Identification number within ECAL ranging from [0, 75848].
- **start_ts**: Start of the Interval of Validity (IOV). An Interval of Validity corresponds to a time during which a measurement is taken for a single crystal. In other words, each IOV contains one and only one measurement per crystal.
- **stop_ts**: End of the Interval of Validity (IOV).
- **laser_datetime**: Timestamp of the measurement for a given crystal within an IOV. The timestamp lies between the start of IOV and end of IOV.
- **calibration**: APD/PD ratio taken at laser_datetime. This value is used to quantify the transparency of the crystal at the time of measurement.
- **time**: Time corresponding to the luminosity measurement (obtained from BRIL) closest to the time when the laser measurement was taken.
- **int_deliv_inv_ub**: Approximate integrated luminosity delivered up to the measurement in the units of micro barn inverse.

In order to ensure the FAIR-ness of the publication of the dataset, it has been published [8] on Zenodo¹ platform, which was launched in May 2013 as a part of OpenAIRE project, in partnership with CERN. The dataset consists of 26 files in

¹<https://zenodo.org/>

tar gzip format, each file consisting of up to 10 csv files. Each csv file contains measurements of up to 360 crystals. The files corresponding to the +z side of the ECAL are labelled as "plus" and those corresponding to -z side are labelled "minus". The list of $i\eta$ rings along with their position in terms of pseudo-rapidity (η) and azimuthal angle (ϕ) in each tar file are included in form of json files under the metadata section in Zenodo.

4 Machine Learning Solutions

An artificial Neural Networks (NN) is a Machine Learning (ML) algorithm that mimic the biological structure and functioning of neurons in a brain. A neural network that does not involve any cyclic connections is called as a Feedforward neural network (FNN). A Deep neural networks (DNN) are the structures that consist of many stages of inter-connected neurons. Deep Neural Networks are powerful on handling hard learning tasks such as object identification and speech recognition. However, it requires the task inputs and outputs to be encoded into vectors with fixed dimensionality [9].

Many problems like machine translation and speech recognition naturally have sequential structure, as their input and output lengths are not known a-priori [9]. The same architectures have also been implemented to solve time series prediction problems. A class of neural networks called Recurrent Neural Networks (RNNs), which are a type of FNNs that pass the data sequentially between different nodes. This architecture allows the network to learn and retain the past knowledge when processing data points from a given data series. RNNs have some shortcomings—in particular, the vanishing gradient problem [10] while training the network. Several other architectures have been developed in the past to address these issues.

4.1 Long Short Term Memory (LSTM) Models

LSTM models are a type of RNNs that include feedback components. RNNs are good at tracking arbitrary long-term dependencies in a sequence, but have a tendency to be unstable during training. LSTMs solve the vanishing-gradient problem through an additive gradient structure. The LSTM cell is as shown in Figure 2, which is the key part of the Seq2Seq model. For each element in the input sequence, each layer of LSTM computes the following functions:

$$\begin{aligned}
 \mathbf{i}_t &= \sigma(\mathbf{W}_{ii}\mathbf{x}_t + \mathbf{b}_{ii} + \mathbf{W}_{hi}\mathbf{h}_{t-1} + \mathbf{b}_{hi}) \\
 \mathbf{f}_t &= \sigma(\mathbf{W}_{if}\mathbf{x}_t + \mathbf{b}_{if} + \mathbf{W}_{hf}\mathbf{h}_{t-1} + \mathbf{b}_{hf}) \\
 \mathbf{g}_t &= \tanh(\mathbf{W}_{ig}\mathbf{x}_t + \mathbf{b}_{ig} + \mathbf{W}_{hg}\mathbf{h}_{t-1} + \mathbf{b}_{hg}) \\
 \mathbf{o}_t &= \sigma(\mathbf{W}_{io}\mathbf{x}_t + \mathbf{b}_{io} + \mathbf{W}_{ho}\mathbf{h}_{t-1} + \mathbf{b}_{ho}) \\
 \mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t \\
 \mathbf{h}_t &= \mathbf{o}_t \odot \tanh \mathbf{c}_t
 \end{aligned} \tag{2}$$

where $\mathbf{h}_t, \mathbf{c}_t$ are hidden and cell state at time t , \mathbf{x}_t is the input at time t , $\mathbf{i}_t, \mathbf{f}_t, \mathbf{g}_t, \mathbf{o}_t$ are the input, forget, cell and output gates, respectively. σ is the sigmoid function and \odot is the Hadamard product [11].

4.2 Seq2Seq Model

A Sequence to sequence (Seq2Seq) model is an architecture which combines two or more LSTMs. It consists of two parts—the encoder and the decoder, each of which is built by using separate LSTMs. This type of model has been developed for automatic language translation, where a sentence from one language is translated to another language [9]. The encoder is used to process each token in the input sentence, and encode all the input sequence information into a fixed length vector. The transform vector, known as context vector, is a vector in a latent space and it encapsulate the whole meaning of the input sequence. The decoder reads the context vector and predicts the target sequence token by token.

Figure 2 shows the basic architecture of the encoder-decoder network used for this problem. The encoder block consists of LSTM units connected in series which takes in a set of calibration values and the luminosity differences between those calibration values as an input. All the information of input sequence is encapsulated into internal states \mathbf{h}_t (hidden state) and \mathbf{c}_t (cell state). The decoder block is another block of LSTM units connected in series. The final states ($\mathbf{h}_t, \mathbf{c}_t$) of the encoder are used as the initial states ($\mathbf{h}_0, \mathbf{c}_0$) to the decoder, which is the context vector used to predict the target sequence. The decoder network also takes inputs along with the initial states to predict the target sequence. The input to the decoder varies according to the method used for training.

The Seq2Seq model can be trained using teacher forcing method, where the decoder is trained using the target output (ground truth output) instead of the output generated by the decoder in the previous step of the sequence. However,

during the evaluation step, the decoder generates the output sequentially using the outputs generated in the previous step. Using teacher forcing during the training has shown to improve the training process.

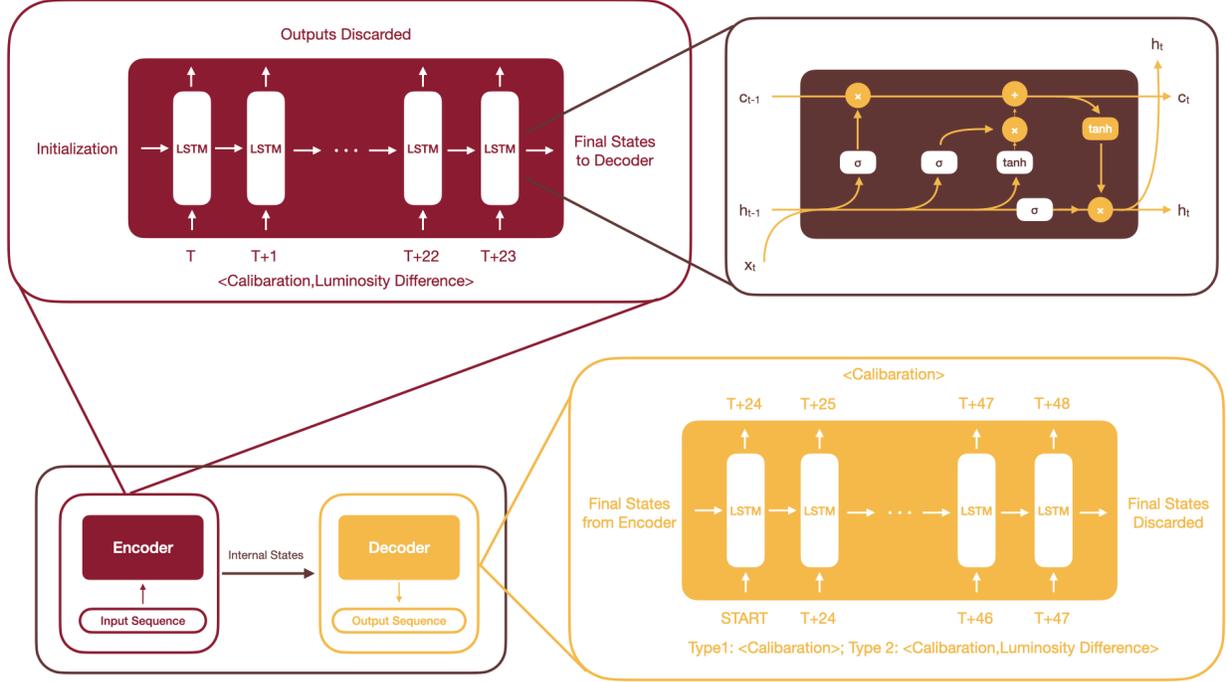


Figure 2: Seq2Seq model for used for predicting future calibration values (**lower left**). The Encoder block (**upper left**) and Decoder block (**lower right**) are a set of sequentially connected LSTM units (**upper right**).

5 Training

For training AI models, PyTorch packages were used. The code is maintained in a public Github repository (<https://github.com/FAIR-UMN/FAIR-UMN-ECAL>). Conda environments are provided so that the users of the datasets can use any API of their liking. The project details can also be found in <https://fair-umn.github.io/FAIR-UMN-Docs>.

5.1 Data Pre-Processing

The difference between subsequent entries in the dataset is the integrated luminosity delivered between the two consecutive measurements. Before training the networks, the measured calibration values and the luminosity differences in the training dataset are normalized to unity using the *StandardScaler* from the sklearn library. Next, to obtain input X and true output Y_{true} used for model training, we performed the following steps:

- Define input length L_E (e.g., $L_E = 24$), corresponding to the number of LSTM units in the encoder, and the output length L_D (e.g., $L_D = 24$), corresponding to the number of LSTM units in the decoder for each individual sample. The Seq2Seq model will be trained to use a sequence of calibration values and luminosity differences of length L_E and learn to predict the next L_D calibration values.
- In order to avoid any overlap between the prediction sequences, a separation stride L_S is set to be the same as L_D . Hence, the total number of samples is

$$N_{sample} = \frac{N - L_E - L_D}{L_S} + 1,$$

where N is the total number of entries in the training dataset. For each individual sample, the input is a sequence starting from T to $T + L_E - 1$ and the output is a sequence starting from $T + L_E$ to $T + L_E + L_D - 1$.

- In PyTorch, the LSTM module takes a 3D tensor as the input whose dimensions are given by (sequence, batch, features). In this problem, the input to encoder ($X_{encoder}$), the input to decoder ($X_{decoder}$) and the output of the

decoder (Y_{decoder}) can be represented as below:

$$\begin{aligned} X_{\text{encoder}} &\in \mathbb{R}^{L_E \times N_{\text{sample}} \times N_E}, \\ X_{\text{decoder}} &\in \mathbb{R}^{L_D \times N_{\text{sample}} \times N_D}, \\ Y_{\text{decoder}} &\in \mathbb{R}^{L_D \times N_{\text{sample}} \times 1}, \end{aligned}$$

where N_E and N_D are the number of features used in encoder and decoder respectively. In this study, N_E and N_D are set to be 2, which represents features calibration value and luminosity difference. But the more features such as the difference in the timestamps between two entries can be added if needed.

5.2 Training Seq2Seq Model

The Seq2Seq model was built using the PyTorch library. Both encoder and decoder blocks were setup with 1024 hidden layers. The number of LSTM cells in the encoder and the decoder were varied to scan for the optimal input and output lengths. The LSTMs were initialized with a sigmoid activation for input, forget and output gates and a hyper-tangent activation for the cell gates. The Mean Squared Error (MSE) loss function along with the Adam [12] optimizer are used to train the model. The model is trained for 200 epochs with a batch size of 128 and a learning rate of 10^{-3} . A higher number of epochs (3000) were used to check if the model performance improves, but it was found to converge after about 200 epochs. All trainings and predictions were performed on machine with Intel Xeon Silver 4214R@2.40GHz, and Nvidia RTX A6000 graphics card with 48 GB memory.

With a large amount of data points, there are several options available for training a model. A single model can be developed for each of the individual crystals. On the other hand, the data points of different crystals that are at equal distances from the center of ECAL, i.e., within one $i\eta$ -ring of the ECAL, can be combined together. The assumption is that these crystals receive an equal amount of radiation dose because of the radial symmetry and hence will have similar behavior in laser response over a course of time. Then this model trained with a larger data points would be able to predict calibrations for all the crystals in the corresponding $i\eta$ -ring. In addition to changing the number of crystals, three different strategies were used which are given as follows:

1. Recursive: In this setting, as shown in Figure 3 (left), we feed the token from Y_{pred} from the previous time step as the input to the current time step.
2. Teacher Forcing: In this setting, as shown in Figure 3 (right), we feed the token from Y_{true} (instead of the token from Y_{pred}) from the previous time-step as the input to the current time step.
3. Mixed: In this setting, the previous two strategies can be combined in different ratios. For example, a mixed training with a teacher forcing ratio of 0.7 means only 70% of batches in the decoder training uses teacher forcing strategy.

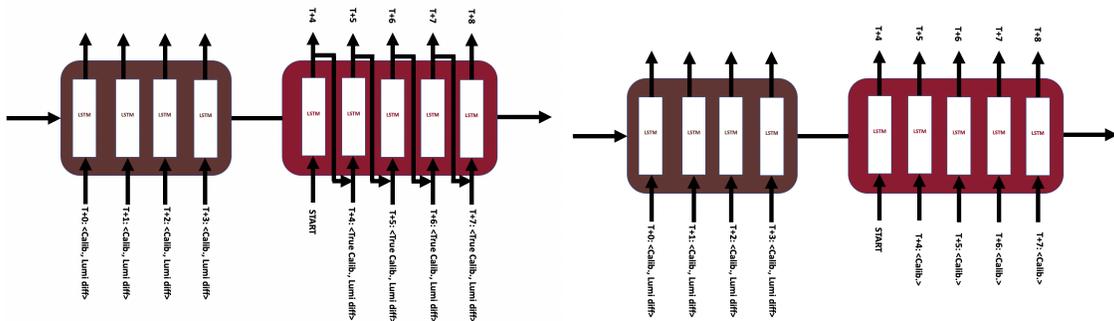


Figure 3: (left): Seq2Seq model with recursive training; (right): Se2Seq model with teacher forcing.

6 Results

To quantify the performance of the LSTM or Seq2Seq model, we use Mean Absolute Percentage Error (MAPE) as the metric, which is defined as

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|, \quad (3)$$

where A_t is the actual value and F_t is the forecast value.

To determine an appropriate input length (L_E) and output length (L_D) for the Seq2Seq model, the models were trained using different sequence lengths with the recursive training strategy. We set $L_E = L_D$ in this example. As shown in Figure 6, sequence length equals 24 gives the lowest MAPE. Hence, this value will be used for all the trainings described in this section.

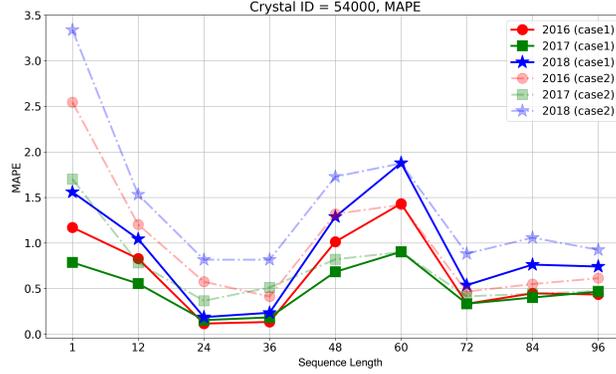


Figure 4: Seq2Seq model trained with different sequence lengths.

6.1 Different Training Strategies

For the purpose of this study, the data taken in the year 2016 corresponding to single crystal with ID 54000, from $i\eta$ -ring 66, were used to train a Seq2Seq model (Model-S). Another model (Model-R) was trained using the data taken in the year 2016 corresponding to all 360 crystals in $i\eta$ -ring 66, with ID ranging from 54000 to 54359. The response of the two models was evaluated on different crystal from the same $i\eta$ -ring 66 using the data taken in the year of 2017 and 2018.

For making these predictions, two cases were used:

1. Case 1: The ground truth is provided as the model input at each prediction window (Figure 5 (left)).
2. Case 2: In this case only the first input is provided and the model would recursively “reuses” the predictions from its previous prediction window as its input (Figure 5 (right)). to make predictions and then evaluate their performance separately. Therefore, in terms of learning, this case is more challenging than Case 1 as we use less prior information.

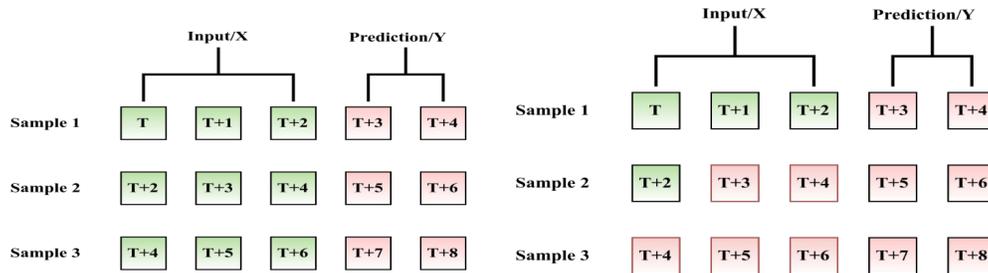


Figure 5: **(left)**: Case 1: without using prediction as the input of the next round prediction; **(right)**: Case 2: using prediction as the input of the next round prediction

The input to each LSTM cell in the decoder contains the luminosity delivered (ΔL_i) between the current and the next timestamp, and the previous calibration value. Typically, during training, the output from the previous LSTM cell in the decoder is used as an input to the next LSTM as a current calibration value, along with the ΔL_i . However, in case of teacher forcing, true calibration value for input is used instead of the output from the previous LSTM cell. Teacher

forcing ratio can be used to define the fraction of batches in the decoder training which would get true calibration values as input during training.

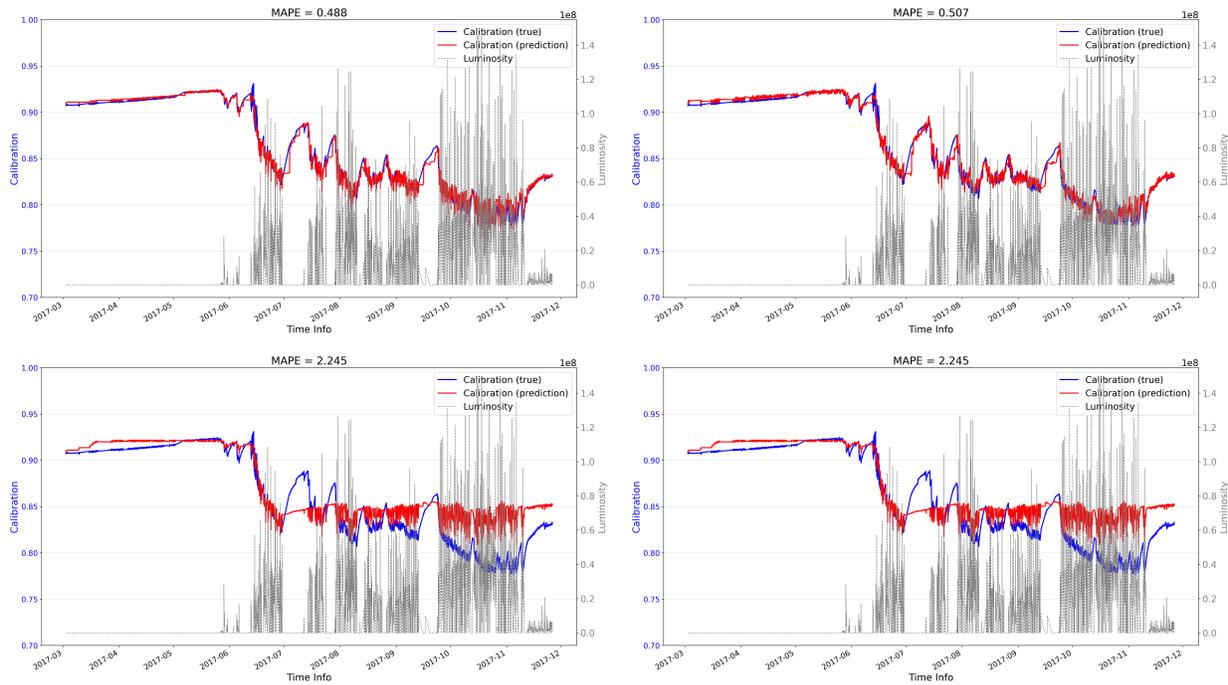


Figure 6: A demonstration of the true calibration curve vs the predicted calibration curve. **(top left)** Case 1 predictions using Model-S (mixed; teacher forcing ratio = 0.5): the response of the crystal ID 54300 in 2017. **(top right)** Case 1 predictions using Model-R (mixed; teacher forcing ratio = 0.5): the response of the crystal ID 54300 in 2017. **(bottom left)** Case 2 predictions using Model-S (mixed; teacher forcing ratio = 0.5): the response of the crystal ID 54300 in 2017. **(top right)** Case 2 predictions using Model-R (mixed; teacher forcing ratio = 0.5): the response of the crystal ID 54300 in 2017.

The demonstration of our predicted calibration curve is shown in Figure 6: Both Model-S and Model-R in Case 1 can successfully predict the calibration in future time steps with low MAPE. However, both Model-S and Model-R in Case 2 give worse predictions after certain time steps.

Year	Prediction	Model-S(M)	Model-R(R)	Model-R(M)	Model-R(T)
2016	Case 1	0.194	0.168	0.180	0.191
2017	Case 1	0.223	0.228	0.234	0.263
2018	Case 1	0.291	0.323	0.330	0.391
2016	Case 2	0.888	0.516	0.577	0.530
2017	Case 2	0.836	0.680	0.713	0.673
2018	Case 2	1.24	1.216	1.147	1.327

Table 1: Average MAPE from prediction on all 360 crystals. M: mixed strategy of teacher forcing and recursive (teacher forcing ratio= 0.5); R: recursive strategy; T: teacher-forcing strategy.

Furthermore, the MAPE has been evaluated for all the 360 crystals (crystal ID:34000-34359) using all three years (2016, 2017 and 2018) of data and the distribution is shown in Figure 7. Also, the corresponding average MAPE among all predictions is shown in Table 1. As shown in both Figure 7 and Table 1 Model-R with recursive, mixed, and teacher forcing strategy have different behavior: the recursive version leads to lower MAPE than the teacher forcing version, which indicates a potential overfitting when using the teacher forcing strategy. Also, in Case 1 prediction, mixed strategy has MAPE between recursive and teacher-forcing; while in Case 2, mixed strategy get worse prediction on 2016 and 2017 data, but get better prediction on 2018 data.

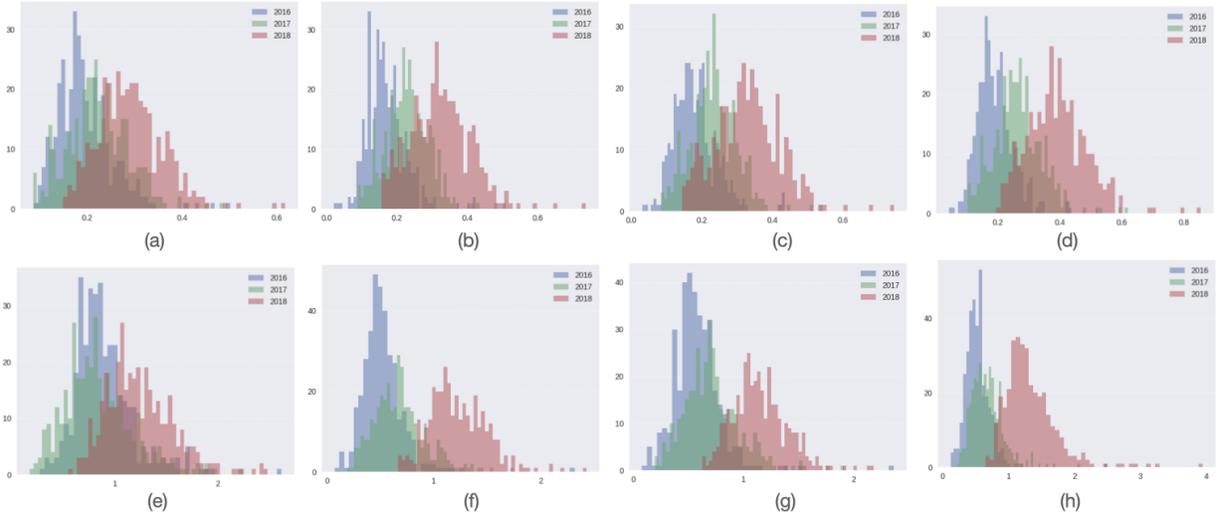


Figure 7: MAPE histograms from prediction on year 2016, 2017 and 2018. (a)(b)(c)(d) use Case 1 prediction strategy, while (e)(f)(g)(h) use Case 2 prediction strategy. (a) and (e) use model trained on single crystal 54000 of 2016. (b)(c)(d)(f)(g)(h) use model trained on all crystals in ring 66 (crystal ID: 54000-54359) of 2016. (b)(f) use recursive strategy in the encoder; (c)(g) use mixed strategy (teacher forcing ratio = 0.5) in the encoder; (d)(h) use teacher forcing strategy in the encoder.

As shown in both Figure 7 (a)(c)(e)(g) and Table 1, Model-S (M) is worse than Model-R (M) in Case 2 prediction but better than Model-R (M) in Case 1 prediction (2017 and 2018). Besides that, Case 1 prediction would be recommended, as it always has better performance than Case 2 prediction.

7 Summary

Recurrent Neural Networks such as LSTMs and Seq2Seq models are able to make decent predictions of the laser response of ECAL crystals in the future. The effects of degradation in the crystal transparency as a function of luminosity is captured by the models along with the recovery of the crystals during absence of any radiations. The dataset of the laser response has been made public and our intention is that these results will generate interest in the wider High Energy Physics and Computer Science community. We hope that this dataset is used for development and bench-marking the performance of different machine learning models that are designed for solving time series problems.

Acknowledgments

This work has been supported by the Department of Energy Office of Advanced Scientific Computing under grant number DE-SC0021395. We would like to thank the CMS collaboration for permitting us to use the laser data. We would also like to thank our colleagues from FAIR4HEP group for discussions and their invaluable inputs and suggestions to writing this paper.

References

- [1] Mark D. Wilkinson, Susanna-Assunta Sansone, Erik Schultes, Peter Doorn, Luiz Olavo Bonino da Silva Santos, and Michel Dumontier. A design framework and exemplar metrics for FAIRness. *Scientific Data*, 5(1):180118, May 2018.
- [2] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004–S08004, aug 2008.
- [3] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716:30, 2012.

- [4] A. M. et al Sirunyan. Measurements of higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{s} = 13$ tev. *Journal of High Energy Physics*, 2021(7):27, Jul 2021.
- [5] Yifan Chen, E. A. Huerta, Javier Duarte, Philip Harris, Daniel S. Katz, Mark S. Neubauer, Daniel Diaz, Farouk Mokhtar, Raghav Kansal, Sang Eon Park, Volodymyr V. Kindratenko, Zhizhen Zhao, and Roger Rusack. A FAIR and AI-ready higgs boson decay dataset. *Scientific Data*, 9(1), feb 2022.
- [6] The CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.
- [7] Xiangdong Qu, Liyuan Zhang, and Ren yuan Zhu. Radiation induced color centers and light monitoring for lead tungstate crystals. *IEEE Transactions on Nuclear Science*, 47(6):1741–1747, 2000.
- [8] Bhargav Joshi and Roger Rusack. Laser response in ecal crystals in cms detector, March 2022.
- [9] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27, 2014.
- [10] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. 2012.
- [11] Lstm-pytorch 1.12 documentation.
- [12] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.