This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Improving the Timing Resolution of Positron Emission Tomography Detectors Using Boosted Learning—A Residual Physics Approach

Stephan Naunheim, *Member, IEEE*, Yannick Kuhl, *Member, IEEE*, David Schug, *Member, IEEE*, Volkmar Schulz, *Senior Member, IEEE*, and Florian Mueller, *Member, IEEE*

***Abstract*— Artificial intelligence (AI) is entering medical imaging, mainly enhancing image reconstruction. Nevertheless, improvements throughout the entire processing, from signal detection to computation, potentially offer significant benefits. This work presents a novel and versatile approach to detector optimization using machine learning (ML) and residual physics. We apply the concept to positron emission tomography (PET), intending to improve the coincidence time resolution (CTR). PET visualizes metabolic processes in the body by detecting photons with scintillation detectors. Improved CTR performance offers the advantage of reducing radioactive dose exposure for patients. Modern PET detectors with sophisticated concepts and read-out topologies represent complex physical and electronic systems requiring dedicated calibration techniques. Traditional methods primarily depend on analytical formulations successfully describing the main detector characteristics. However, when accounting for higher-order effects, additional complexities arise matching theoretical models to experimental reality. Our work addresses this challenge by combining traditional calibration with AI and residual physics, presenting a highly promising approach. We present a residual physics-based strategy using gradient tree boosting and physics-guided data generation. The explainable AI framework SHapley Additive exPlanations (SHAPs) was used to identify known physical effects with learned patterns. In addition, the models were tested against basic physical laws. We were able to improve the CTR significantly (more than 20%) for clinically relevant detectors of 19 mm height, reaching CTRs of 185 ps (450–550 keV).**

***Index Terms*— Coincidence time resolution (CTR), explainable artificial intelligence (XAI), gradient boosted decision tree (GBDT), positron emission tomography (PET), residual physics, time-of-flight (TOF).**

Manuscript received 20 January 2023; revised 27 July 2023; accepted 29 September 2023. This work was supported by the German Federal Ministry of Education and Research (BMBF) within the funding Program "Recognizing and Treating Psychological and Neurological Illnesses—Potentials of Medical Technology for a Higher Quality of Life ("Psychische und neurologische Erkrankungen erkennen und behandeln—Potenziale der Medizintechnik für eine höhere Lebensqualität nutzen") under Contract 13GW0621B. *(Corresponding authors: Stephan Naunheim; Volkmar Schulz; Florian Mueller.)*

Stephan Naunheim, Yannick Kuhl, and Florian Mueller are with the Department of Physics of Molecular Imaging Systems, Medical Faculty, Institute for Experimental Molecular Imaging, RWTH Aachen University, 52074 Aachen, Germany (e-mail: stephan.naunheim@pmi.rwth-aachen.de; florian.mueller@pmi.rwth-aachen.de).

David Schug is with the Department of Physics of Molecular Imaging Systems, Medical Faculty, Institute for Experimental Molecular Imaging, RWTH Aachen University, 52074 Aachen, Germany, and also with the Hyperion Hybrid Imaging Systems GmbH, 52074 Aachen, Germany.

Volkmar Schulz is with the Department of Physics of Molecular Imaging Systems, Medical Faculty, Institute for Experimental Molecular Imaging, RWTH Aachen University, 52074 Aachen, Germany, also with Hyperion Hybrid Imaging Systems GmbH, 52074 Aachen, Germany, and also with Fraunhofer MEVIS, Institute for Digital Medicine, and Physics Institute III B, RWTH Aachen University, 52062 Aachen, Germany (e-mail: volkmar.schulz@pmi.rwth-aachen.de).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2023.3323131.

Digital Object Identifier 10.1109/TNNLS.2023.3323131

I. Introduction

ARTIFICIAL intelligence (AI) is finding its way more and more into medical imaging [1], [2], including the research field around positron emission tomography (PET) [3]. In contrast to computed tomography (CT) [4] or magnetic resonance imaging (MRI) [5], PET is a functional imaging technique that does not reproduce anatomical structures but can visualize metabolic processes in the body. PET uses the effect of electron–positron annihilation to obtain information about processes within the object of interest. A radioactive tracer is administered to the patient, accumulating in highly metabolic regions and emitting positrons [6]. These positrons annihilate with the surrounding tissue, producing two γ-photons emitted back-to-back defining a line-of-response (LOR). The γ-photons are subsequently registered in coincidence by a PET scanner (see Fig. 1) equipped with scintillation detectors, which convert the γ-photons into many optical photons in the visible light range that can be measured with a photosensor [7]. Analog silicon photomultipliers (SiPMs) utilizes an application-specific integrated circuit (ASIC) to digitize the signal pulses, whereas digital SiPMs performs digitization at a single photon avalanche diode (SPAD) level. Based on the detection information, especially the spatial and the temporal information, a PET image can be reconstructed. While the application of neural networks in medical imaging usually focuses on the image reconstruction process [8], [9], [10], improvements along the complete imaging chain, from detecting physical signals [11] to processing the resulting data, can improve the resulting image to facilitate medical diagnoses. In this work, we show the application of learning algorithms in the context of residual physics at the detector level to significantly improve the achievable coincidence time resolution (CTR). Especially in medical and physical applications, it is desired to get insights into the inner workings of models to ensure that the algorithms can capture meaningful

© 2023 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

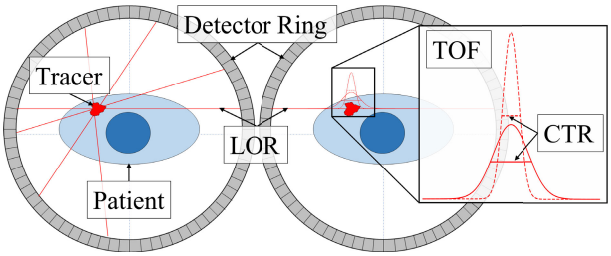


Fig. 1. Principle of PET. A radioactive tracer (red blob) is administered to a patient (bluish). By detecting the created back-to-back γ-photons, defining many LORs (red lines), with the detector ring (grayish), spatial and temporal information can be inferred to reconstruct a PET image. Modern scanners utilize TOF information to estimate the annihilation point on the LOR.

relations. Therefore, we use eXplainable AI (XAI) [12], [13], [14], [15] methods to check whether trained models can replicate simple physical constraints implied by the data generation.

State-of-the-art clinical PET scanners combine high spatial resolution (SR) with precise time-of-flight (TOF) information (see Fig. 1). Including the timing information in the image reconstruction process provokes an improvement in the signal-to-noise ratio (SNR) of the image [16] without increasing the radioactive dose and therefore improves also lesion detectability [17]. Most PET systems [18], [19] utilize segmented scintillator topologies (see Fig. 2) due to the readout simplicity and very good timing performances. Contrary to this, (semi-)monolithic detector concepts spread the light over multiple channels. Recently, they have gained attention [20], [21] as they provide high SR [22], [23], [24] but also offer intrinsic depth of interaction (DOI) capabilities [25], [26], thus reducing parallax errors at reduced costs compared to segmented topologies. While, for example, the γ-positioning strongly profits from the spread detection information, it creates disadvantages for the timing performance due to an enhancement of timewalk effects [27], [28], [29] and jitter in signal-propagation times [30], [31], which deteriorate CTR. Therefore, monolithic detector concepts demand advanced readout algorithms and calibration routines to infer the needed information from the detected optical information. Due to the light-spreading characteristic of (semi-)monolithic detectors, many approaches use machine- or deep-learning techniques, for example, to infer the γ-interaction position within the scintillator volume. This strategy suggests itself since the detected optical photons represent abstract patterns that can easily be recognized by learning algorithms. However, applying machine learning (ML) for time skew calibration and estimation still remains experimentally a hard task since skew effects can vary in their magnitude and also in the incorporated nature on an event basis without the need for a spatial relation. Besides this, supervised learning demands labeled data, which is a priori not accessible without using simulation techniques, and unsupervised learning is often used in the context of clustering and association algorithms [32] unsuitable for the proposed problem. Recently, we proposed an analytical timing calibration technique [33] suitable for traditional segmented and light-sharing-based scintillator topologies. This analytical calibration aims to reduce sequentially major skew effects by using a convex optimization of a matrix equation. When applying the technique, one observes that the skew effects are iteratively reduced. Within each iteration, the experimenter can address different characteristics of the time skews, for example, by choosing a different separation into subvolumes (called voxels) of the scintillation crystals. At a certain number of iterations, we see that the reported correction values $\hat{\vec{c}}$ oscillate around the baseline and that the CTR does not improve further, indicating that the linear formulation of the problem, with $\boldsymbol{M}$ denoting the matrix and $\overrightarrow{\Delta t}$ the estimated mean time difference between the calibration objects

$$\hat{\vec{c}} = \arg\min_{\vec{c}} \|\overrightarrow{\Delta t} - \boldsymbol{M} \cdot \vec{c}\|_2 \tag{1}$$

has limited capability of completely describing the physical situation. This challenge can theoretically be addressed by changing the mathematical formulation representing also the effects of higher order. However, this requires prior knowledge of the precise optical processes [34] taking place in the chosen scintillator topology, to change the mathematical formulation [35]. Furthermore, depending on the readout infrastructure and the detector concept, the problem might depend on numerous variables and parameters [36], [37] which are hard to determine in advance. Hence, covering the effects of higher order can become arbitrarily complicated. In addition, detectors can vary in response such that an optimized representation might only fit the specific detector. A statistical approach using maximum likelihood was presented by van Dam et al. [38], focusing on differences between timestamps. We propose to use an ML approach instead and furthermore utilize a special way of experimental data generation to propose simple prior physical knowledge to the model by shifting a radiation source to different known positions [39]. We intend to apply this technique on top of the conventionally used analytical approach, forcing the algorithm to learn the effects of higher order, which we understand as residual physics [40], [41]. By following this, we free ourselves from precisely modeling and catching all nonlinear effects in the complete scintillation and detection processes. In this work, we employed gradient-boosted decision trees (GBDTs) as learning algorithm since it can handle missing data [42] and allows usage in (near) real-time processing systems [43] due to the simplicity of the model's architecture.

The proposed approach is studied using experimental data acquired with a coincidence setup equipped with a semi-monolithic ($3.9 \times 31.9 \times 19.0$ mm^3) and a one-to-one coupled ($3.9 \times 3.9 \times 19.0$ mm^3) detector array concept. We trained multiple models on the acquired data and studied their performance based on the physics-related learning task, the agreement with theoretical expectations and bias effects, and the obtained CTR values.

II. Related Works

A. Approaches Toward Residual Physics

To the authors' knowledge, the first popular mention of "residual physics" in the context of AI was by Zeng et al. [40]. In their work, they investigated whether a robotic arm can pick up arbitrary objects and throw them into selected target boxes. While the problem of throwing can be described sufficiently well in theory by Newtonian physics, the real-world

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

implementation for arbitrary objects is very challenging due to numerous additional variables that affect the throw.

Similar works have been [44], [45], [46], [47], [48] and are still being published [49] in the context of "hybrid controllers." All of the studies have in common that they exploit the residuals between well-understood idealized physics and actual measurement.

Alternative approaches aiming to combine physics domain knowledge and AI are given by "physics-informed learning" [50], [51], [52], [53], where the utilized loss function is often modified to guide the model to physics-meaningful predictions.

B. Timing Capabilities of PET Detectors

In recent publications [42], [54], [55], [56], it has been shown that (semi-)monolithic detectors can provide good performances. Especially their timing capabilities have been studied under various experimental settings. Van Dam et al. [38] were able to reach sub-200 ps CTR for a monolithic crystal ($24 \times 24 \times 20$ mm^3) using a maximum likelihood approach and a measurement temperature of -20 °C, challenging to implement in a PET system designed for the clinical domain. Sánchez et al. [57] developed a new ASIC (HRFlexToT) with redesigned energy measurement for linear time-over-threshold (ToT) behavior while reducing power consumption and improved timing response and achieved 324 ps CTR for a big monolithic crystal ($25 \times 25 \times 20$ mm^3). In a recent simulation study, Maebe and Vandenberghe [58] reported 141 ps. In their simulation, they used a monolithic detector ($50 \times 50 \times 16$ mm^3) and a convolutional neural network (CNN), while the network's input is given by the digitized waveforms truncated to a window of 3 ns using a step size of 100 ps.

Zhang et al. [55] reported a timing resolution of about 718 ps with thick semi-monoliths ($1.37 \times 51.2 \times 20$ mm^3) digitized with the TOFPET2 ASIC at measurement temperatures of 28 °C. Using energy-weighted averaging of timestamps reported by the TOFPET2 ASIC, Cucarella et al. [59] achieved a CTR of 276 ps for slabs with a volume of $25.4 \times 12 \times 0.95$ mm^3.

In a proof-of-concept study performed by Berg and Cherry [60] using two small lutetium fine silicate crystals ($5 \times 5 \times 10$ mm^3) coupled to a single photomultiplier tube, a timing resolution of about 185 ps was achieved using the CNNs. Onishi et al. [61] proposed a simple method for unbiased TOF estimation by applying a combination of a CNN and an leading edge discriminator (LED) to an oscilloscope equipped with a pair of single scintillation Lutetium–yttrium oxyorthosilicate (LYSO) crystal of dimensions $3 \times 3 \times 10$ mm^3 reaching 159 ps.

III. Learning Algorithm and Materials

A. Gradient-Boosted Decision Trees

While we utilized GBDT in this work, the presented calibration approach is also applicable to different learning architectures, for example, deep neural nets. GBDT is a supervised learning algorithm based on an ensemble of binary decision trees, where each tree is trained on the residuals of the already established ensemble (additive training). In this work, we use the GBDT implementation of XGBoost [62], with the model ϕ

$$\phi = \sum_{k=1}^{K} f_k \tag{2}$$

being given as the superposition of the K trees (weak learners) f_k. Each tree f_k is an element in the CART [63] space Ω

$$f_k \in \Omega. \tag{3}$$

In its design, GBDT is a relatively simple architecture compared to widely used deep neural networks [64], [65], [66]. However, it has proven high predictive power in many applications [67], [68], [69], [70], [71], and due to its simplicity, GBDT allows usage in high-throughput software [43] suitable for complete PET systems or even the application directly on the detector level [72], [73]. Regarding the scope of this work, two hyperparameters of GBDT models are of particular importance, namely the maximal depth d, denoting the maximal number of decisions within an ensemble, and the learning rate lr, measuring the residual influence on the learning of the following tree. The learning rate must be optimized in most cases to find a compromise between training duration and accuracy. A third prominent hyperparameter is the number of trees n of an ensemble. We excluded n from the hyperparameter search in this work since we used an early stopping criterion.

B. SHapley Additive exPlanations

The SHapley Additive exPlanations (SHAP) framework [74], [75] is used as an explainable AI technique to analyze feature importance to search for correlations between physical effects and patterns the model has learned. In particular, in this work, we utilized the TreeExplainer implementation [76] because of the chosen learning architecture. The framework uses mathematical game theory. Each input sample and corresponding prediction is connected by assuming a coalition game. The players in the game are represented by the feature values of the input sample, where each feature influences the model's prediction. These influences are called contributions and are expressed in the same physical unit as the predictions. Contribution values are mathematically either positive or negative, while the model's output is equal to the sum of the contributions. The magnitude of a given contribution indicates the level of its importance.

SHAP uses Shapley values [77], which are a measure to quantify the contribution of a feature regarding the specific model's output. In a mathematical sense, SHAP combines three concepts essential for providing a consistent picture concerning feature importance. Firstly, the SHAP values must satisfy local accuracy, meaning that for a given input sample, the sum of the estimated feature contributions must be equal to the corresponding model's prediction that should be explained. If a feature is missing, it cannot be attributed to importance, which is covered in SHAP using the concept of missingness. Lastly, consistency is required, ensuring that when changing

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

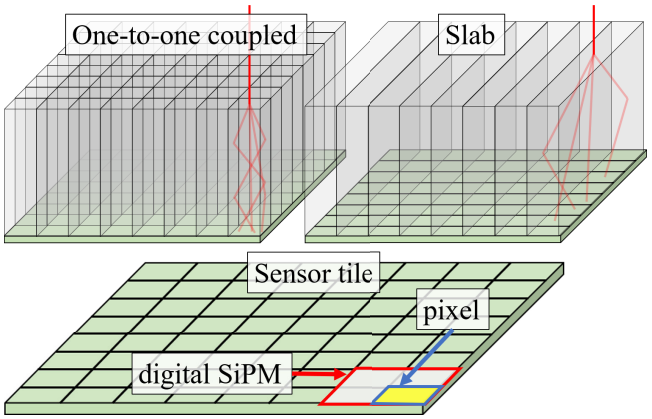


Fig. 2. Used scintillator topologies and photosensors. The incoming γ-photon, as well as a part of the optical photons, are illustrated as red lines. The sensor tile consists of 4 × 4 digital SiPMs (DPC3200-22, PDPC), each one holding four pixels and a twin time-to-digital converter. A triggered SiPM reports four pixel count values and a timestamp.

a model such that a particular feature has a larger impact on the model, the corresponding attribution cannot decrease.

Practically, for each feature value f_k of a given input sample $X = \{f_k\}$, an associated SHAP value $SV(f_k)$ can be computed, reporting a local explanation that connects the feature value with its contribution to the model's output y. By combining many local explanations, one can conclude a global understanding of the model.

C. PET Detectors

The study is conducted using two different detector types, where one detector is based on a one-to-one coupled scintillator design, and the other detector is based on a semi-monolithic scintillator design (see Fig. 2).

Each scintillator concept is glue-coupled (Meltmount, Cargille Laboratories) to a sensor tile holding 4 × 4 digital SiPMs (DPC3200-22, Philips Digital Photon Counting (PDPC), Aachen [78]). Each SiPM is formed by 2 × 2 readout channels (also called pixels) and a twin time-to-digital converter, where one readout channel consists of 3200 SPADs. Each SiPM of a sensor tile works independently and follows a configured acquisition sequence if a predefined internal two-level trigger scheme is fulfilled. After the reception of a trigger, it is checked during the validation phase if the geometrical distribution of discharged SPADs met the configured requirement. If both trigger thresholds are fulfilled, the acquisition is continued. Each triggered SiPM provides information that encloses a timestamp and four-pixel photon count values, called a hit.

Both scintillators use LYSO as scintillation material (Crystal Photonics, Sanford). Concerning the scintillator architecture, an array of 8 × 8 LYSO segments of 4.0 mm pitch and 19.0 mm height is utilized in the one-to-one coupled design. Each segment is wrapped with enhanced specular reflector (ESR) foil and covers the pitch of one pixel.

The semi-monolithic detector concept comprises eight monolithic LYSO slabs, each having a volume of 39 × 31.9 × 19.0 mm^3. Each slab aligns with one row of pixels. ESR foil is located between every second slab and on the lateral walls to reduce light sharing between the trigger and readout regions.

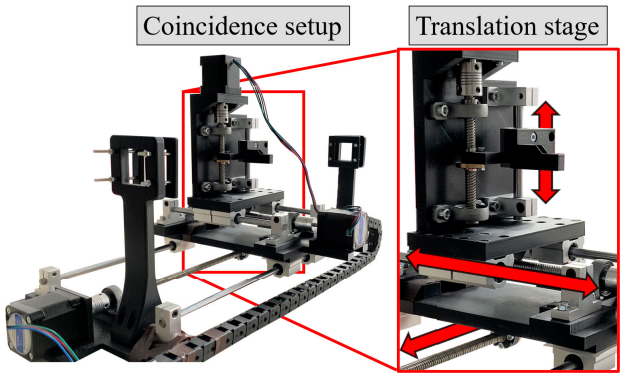


Fig. 3. Used coincidence setup for the acquisition of labeled data. The source mounting is connected to the translation stage system allowing motion along all three axes (indicated as red arrows).

The slab detector is able to provide intrinsic DOI information due to its monolithic characteristics. Not all SiPMs that are partly covered by a slab might be triggered and send hit data corresponding to a γ-interaction due to the independent operation of the SiPMs.

D. Coincidence Setup

The experimental setup comprises a source mounting, in addition to the detectors, and is located in a tempered dark box. The source mounting is connected to a programmable translation stage system, allowing motion in all three spatial axes (see Fig. 3). The distance between the detector surfaces is given to be 435 mm. The precision of the translation stage considering the complete measurement range is given to be 10 μm which translates regarding a coincidence measurement to an uncertainty in the time domain of about 0.067 ps. The source mounting is equipped with a ^{22}Na source with an activity of approximately 12 MBq and a diameter of 0.5 mm. Coincidences are acquired by utilizing flood irradiation and moving the source to various positions between the detectors.

IV. Experiments

A. Data Acquisition

The proposed calibration technique uses supervised ML and therefore demands labeled data. The labels are generated by moving the radiation source to specific positions between the facing detectors and measuring coincidences. Thanks to the known source position, one can calculate the expected time difference $\mathbb{E}(\{\Delta t\})$, and because of the different path lengths, the γ-photons have to travel until reaching the detector. The source was moved to 47 different z-positions (see Fig. 4) with a step size of 5 mm ranging from −130 mm to 100 mm, while at each z-position, a grid of 5 × 5 positions in the xy-plane with a step size of 6 mm was utilized to acquire coincidences. Both, x and y positions ranged from −12 mm to 12 mm. At each grid point, a measurement time of 600 s was set, resulting in a total measurement time of about 8 d. The acquired measurement data consisting of 6.82×10^8 coincidences ($\approx 5.80 \times 10^5$ coincidences per position) is finally used to form three datasets for training, validation, and testing during the model-building process comprising 3.29×10^8, 1.56×10^8, and 1.97×10^8 input samples,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

respectively. We decided to evaluate the final CTR performance (see Section V-B), using data from a measurement conducted on a different day using the same conditions and detectors to prove the predictive power and generalized applicability of the trained models. This dataset comprises 4.20×10^6 coincidences acquired with the radiation source located near the iso-center of the setup, as it is usually done for CTR evaluations. To allow a clean separation in the naming, the dataset used for the CTR performance evaluation is called the performance dataset, while the three other datasets remain in the usual naming (training, validation, and testing).

During both acquisitions, the sensor tile reported a constant temperature of 2.1 °C for the one-to-one coupled detector and 0.0 °C for the slab detector. Both sensor tiles were operated in the first-photon trigger [79]. The excess voltage was adjusted to 2.8 V, while the validation pattern was set to scheme 16 (0x55:AND) demanding on average 54 ± 19 optical photons [80].

B. Data Preprocessing and Preparation

1) Coincidence Clustering: Data associated with one γ-interaction has to be clustered due to the independent readout of the DPCs. A cluster window of 40 ns is reasoned by the timestamp difference distribution of the hits' uncorrected timestamps to combine all hits into a cluster correlated to the same γ-interaction. Measured raw data were corrected for saturation effects, and the time-to-digital converters of each DPC were linearly calibrated against each other, assuming a uniform distribution of triggers regarding a clock cycle [31]. Clusters with less than 400 or more than 4000 detected optical photons were rejected for noise reduction since the noncalibrated photopeak of the 511 keV γ-photons was located at 2300 and 2800 for the slab and one-to-one coupled detector, respectively. Coincidences were grouped on the cluster level using a sliding coincidence window of 10 ns considering the first timestamp of two clusters.

2) Position and Energy Estimation: A subset of the features used during the proposed time skew calibration is given by the γ-interaction position inside the scintillator volume and by the deposited and calibrated energy in units of kilo electronvolt. To acquire the positioning and energy information of each event, dedicated calibrations already established in previous works [22], [25] were performed.

While the γ-positioning in the one-to-one coupled detector is given by the pixel's position showing the highest photon count, the semi-monolithic slab detector requires a calibration procedure to estimate the 3-D interaction location. For this purpose, GBDT [22], [25], [62] models are trained based on data acquired with an external reference using a fan-beam setup [81], which irradiates the scintillator at known positions. While the positioning resolution of the one-to-one coupled detector is given to be 2 mm, the slab detector's resolution is in the planar direction 2.5 mm and in the DOI direction 3.3 mm. The positioning resolution is determined by the full width at half maximum (FWHM) of the positioning error distribution [25].

The energy value associated with a γ interaction is estimated using a 3-D-dependent energy calibration utilizing an averaged light pattern. The crystal volume is divided into $n_x \times n_y \times n_{\text{doi}}$ voxels, where for each voxel, the mean number of detected optical photons is estimated, based on γ-events, whose interaction positions were located inside the voxel volume. The slab detector is divided into $8 \times 8 \times 4$ voxels, while the one-to-one coupled detector is divided into $8 \times 8 \times 1$ voxels. The energy resolution of the one-to-one coupled detector was evaluated at 10.4%, while the energy resolution of the slab was estimated to be 11.3%.

3) Analytical Timing Calibration: The first part of the calibration is given by performing an analytical calibration, which has been studied in previous publications [31], [82], [83] and relies on well-known mathematical principles like convex optimization. In this work, our calibration formalism [33] was used. However, the principle of exploiting residual physics remains functional for every other analytical calibration.

During the calibration process, multiple subcalibrations are conducted, where in each subcalibration, different hyperparameters are applied such that one tries to address many aspects of time skew effects. The same convex optimization process is used within each subcalibration to find suitable corrections $\hat{\vec{c}}$

$$\hat{\vec{c}} = \arg\min_{\vec{c}} ||\overrightarrow{\Delta t} - \boldsymbol{M} \cdot \vec{c}||^2 \tag{4}$$

with $\vec{c}$ and $\overrightarrow{\Delta t}$ denoting the calibration channel vector and the mean time difference vector, respectively, and $\boldsymbol{M}$ encoding different channel combinations in the form of a matrix. After some number N of performed subcalibrations, a convergence of the detector CTR value as well as the estimated corrections $\{\hat{c}_k\} \in \hat{\vec{c}}$ is visible

$$\text{CTR}_i \rightarrow \text{const. for } i \rightarrow N \tag{5}$$

$$\{\hat{c}_k\}_i \rightarrow 0 \text{ ps for } i \rightarrow N \tag{6}$$

with i denoting the number of applied subcalibrations. For this work, we used three subcalibration iterations (based on time-to-digital converter (TDC) regions, readout channels, and voxels as described in [33]) mainly addressing fixed skews due to differences in the signal propagation and time jitter introduced by the scintillator itself. At this point, it becomes inconvenient to add more and more subcalibrations since the benefit decreases strongly.

C. Residual Timing Calibration

We propose to use a data-driven approach on top of the conventionally used technique to explore new corrections that have not been covered by the analytical formulation and improve the CTR. A suitable way of doing this is by using AI to search for patterns in the acquired coincidence data. We decided to employ the supervised algorithm GBDT (see Section III-A), which was also used during the γ-positioning (see Section IV-B2). Using a supervised approach demands labeled data (input samples and corresponding target values known as labels) to train a model. However, for the proposed problem of non-static time skew effects labeling is difficult, since it is a priori not clear how many and how strong the worsening effects are pronounced in each measured

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

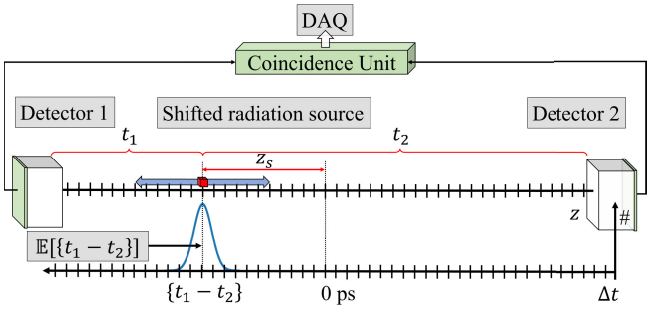


Fig. 4. Scheme of the labeling process to acquire data that can be used for supervised learning. The radiation source (red cube) is shifted to different positions z_s along the centered coordinate system z. Varying source positions lead to different travel times t_1 and t_2 of the γ-photons. The expected time difference $\mathbb{E}[\{t_1 - t_2\}]$ is used as label for the learning process.

coincidence. Using an analytical estimator to generate the ground truth would limit the capabilities of the trained model to the chosen estimator. To solve the problem of labeling, we propose to shift the radiation source to different positions and measure coincidences between the facing detectors. The γ-photons travel varying path lengths to the detectors resulting in different expected time differences per source position.

The different path lengths of the γ-photons (see Fig. 4) lead to different travel times t_1 and t_2. One can conclude the expected time difference $\mathbb{E}[\{t_1 - t_2\}]$, which is subsequently used as label y

$$y = \mathbb{E}[\{t_1 - t_2\}] \approx \frac{-2z_s}{c_{\text{air}}} \tag{7}$$

with c_{air} denoting the speed of light in air and z_s denoting the source offset under the assumption that the coordinate system z is located at the iso-center of the setup (see Fig. 4). For Gaussian distributions, the expectation value $\mathbb{E}$ is identical to the mean value of the distributions. Data acquired with the aforementioned scheme is further processed and finally used to train GBDT models. The input features $\mathcal{F}$ can be grouped into three categories: purely slab detector-related features $\mathcal{F}^s$, purely one-to-one detector-related features $\mathcal{F}^o$, and features associated with both detector concepts $\mathcal{F}^{so}$. While $\mathcal{F}^{so}$ consists only of the difference Δt_{meas} between the first timestamps from slab and one-to-one coupled detector, respectively, $\mathcal{F}^s$ and $\mathcal{F}^o$ can be separated into the subsets timestamp information $\mathcal{F}_T^{s/o}$, energy information $\mathcal{F}_E^{s/o}$, and position information $\mathcal{F}_{\text{Pos}}^{s/o}$ (see Fig. 5).

Since the detector-specific feature sets $\mathcal{F}^s$ and $\mathcal{F}^o$ are symmetrical in their content, we will explain the specific features in a generalized way in the following. The subset timestamp information $\mathcal{F}_T^{s/o}$ contains the four (three) first timestamp values reported by the slab (one-to-one coupled) detector. A tradeoff between available information and needed memory reasons for the choice of the different number of used timestamps. For both detectors, the cumulative distribution of the number of generated timestamps per cluster was analyzed and determined to the value matching roughly 80% of all clusters. Let $\mathcal{T}_j$ be the set of timestamps provided within a cluster j by the photodetector

$$\mathcal{T}_j = \{t_{j,0}, t_{j,1}, \ldots, t_{j,i}, \ldots\} \tag{8}$$

with $t_{j,i}$ denoting the ith timestamp of cluster j. Since the photosensor consecutively reports the timestamp values throughout the measurement, they need to be processed after the coincidence search to be suitable for feeding into an ML algorithm. Therefore, the very earliest timestamp $t_{j,0}$ of a cluster j is subtracted from the following timestamps $t_{j,i}$ of this cluster:

$$\tilde{t}_{j,i} = t_{j,i} - t_{j,0} \tag{9}$$

with $\tilde{t}_{j,i}$ denoting the processed timestamp i of cluster j employed as input.

Furthermore, the origin of the respective timestamps is used and represented by their SiPM ID number. Besides this, information about the cluster's timestamp spread (the difference between the first and the last timestamps) and the number of timestamps (equals the number of hits) in the cluster is utilized. The subset energy information $\mathcal{F}_E^{s/o}$ contains information about the deposited energy as estimated energy value in keV, and as raw photon counts that have been detected on the corresponding SiPMs. The γ-positioning set $\mathcal{F}_{\text{Pos}}^{s/o}$ holds information about the interaction position of the γ-photon within the scintillator volume. While this is given as a 3-D position for the semi-monolithic case, the one-to-one coupled design provides only planar (2-D) information.

To find suitable hyperparameters regarding the learning task, a grid search was conducted considering the maximal tree depth d and the learning rate lr, with

$$d \in \{12, 15, 18, 20\}, \text{ and} \tag{10}$$

$$lr \in \{0.1, 0.3, 0.5\}. \tag{11}$$

During the model-building process, the maximal number of estimators n within an ensemble was set to $n = 500$, where the final number of used estimators was defined by the built-in early stopping criterion considering ten early stopping rounds to suppress possible overfitting. The learning task is performed using XGBoost's default squared error loss function [62].

D. MAE Evaluation and Linearity of Predictions

The mean absolute error (MAE) is used to evaluate the performance of a trained GBDT model based on the testing data

$$\text{MAE}(z_s) = \frac{\sum_{i=0}^{N} |y_i(z_s) - \hat{y}_i(z_s)|}{N} \tag{12}$$

with $y_i(z_s)$ denoting the label of sample i belonging to the source position z_s, and $\hat{y}_i(z_s)$ denoting the corresponding model prediction. We utilize information about the test data prediction distributions to verify their Gaussian shape using a goodness-of-fit approach and to validate that the linearity condition given by (7) is fulfilled. This validation ensures that the trained models obey the physical principle and do not compress the time differences since it would artificially improve the CTR. Therefore, linear regression is performed for each trained GBDT model and each grid point (x_s, y_s) in a range from -75 mm to 45 mm, considering the fit mean value of the prediction distributions μ_s and the associated source

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

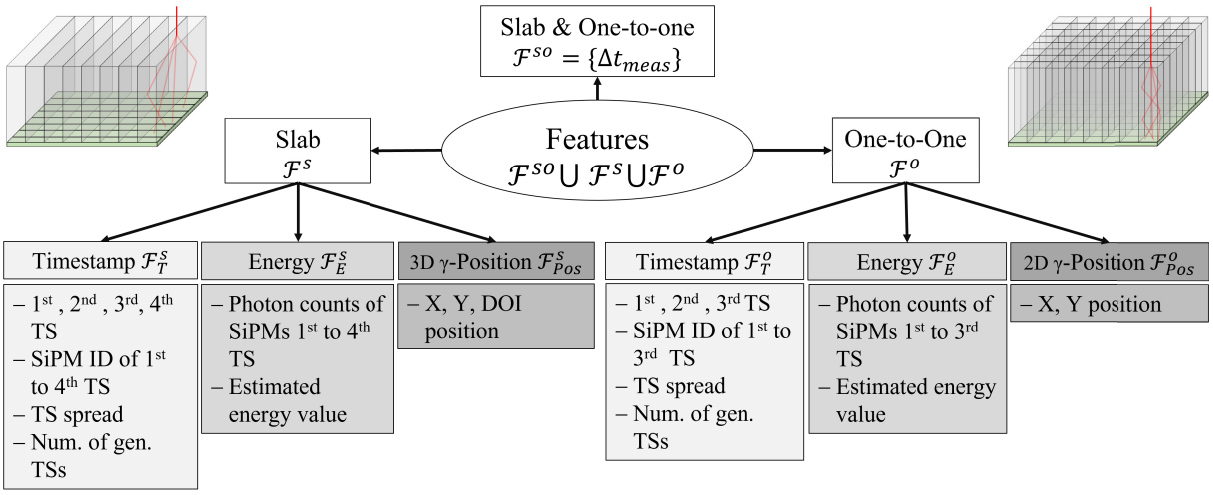


Fig. 5. Overview of the used features $\mathcal{F}$ to train the GBDT models. There are three different feature sets, given by purely slab detector-related features $\mathcal{F}^{s}$, purely one-to-one coupled detector-related features $\mathcal{F}^{o}$, and features associated with both detector concepts $\mathcal{F}^{so}$. $\mathcal{F}^{so}$ consists only of the difference Δt_{meas} between the first timestamps from slab and one-to-one coupled detector, respectively. The sets $\mathcal{F}^{s}$ and $\mathcal{F}^{o}$ are symmetrical in their content and can again be grouped into the subset's timestamp information $\mathcal{F}^{s/o}_{T}$, energy information $\mathcal{F}^{s/o}_{E}$, and position information $\mathcal{F}^{s/o}_{\text{Pos}}$. Information about the processed timestamps (denoted as TS), the SiPM IDs of those timestamps, the timestamp spread (difference between the first and last timestamp of a cluster), and the number of generated timestamps is given. The latter equals also the number of hits within the cluster. Besides this, the photon counts of the corresponding SiPMs, the calibrated energy value, and the spatial interaction position are used.

position z_s. We assumed a linear dependence following:

$$\mu_{\text{s}}(z_s|\varepsilon, b) = \frac{-2}{c_{\text{air}}} \cdot \varepsilon \cdot z_s + b \tag{13}$$

while in theory

$$\varepsilon \stackrel{!}{=} 1. \tag{14}$$

All fitting procedures are performed using SciPy's ODR package [84]. The uncertainty $\sigma_{\mu_{\text{s}}}$ on μ_{s} was based on the uncertainty on the mean reported by the fit procedure. Furthermore, an uncertainty on the translation stage position was given to be the same for all source positions $\sigma_{z_s} = 0.1$ mm. Finally, the global linearity performance is given by the averaged ε value for each model.

E. CTR Performance

To evaluate the timing performance, the FWHM of the predicted time difference distribution is estimated by fitting a Gaussian function. The error on the estimated timing resolution is given by the uncertainty on the fit σ-parameter of the Gaussian. The input data is given by the performance dataset. The CTR is estimated for unfiltered data, for coincidences within a large energy window from 300 to 700 keV, and coincidences within a smaller energy window from 450 to 550 keV.

F. SHAP Analysis

Due to computational costs, the SHAP analysis was performed for the model showing the best MAE and CTR performance using a subset of 23 500 samples of the performance data. The analysis was done without applying any filters.

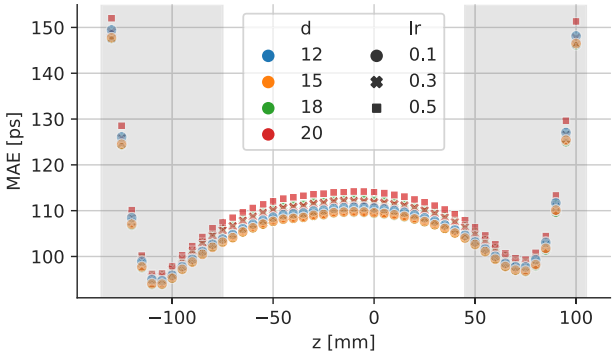


Fig. 6. Progression of the MAE for each source position z_s contained in the test dataset. No energy filter or restrictions on the measured light distribution were applied. Models utilizing a small learning rate show the best performances. Predictions located in the grayish areas ($z \notin [-75, 45]$) mm are excluded from the linearity analysis since the MAE progression indicates the starting of the transition into the artifact-dominated region for these points.

V. Results

A. MAE Evaluation and Linearity of Predictions

The MAE performance (see Fig. 6) is similar for all chosen hyperparameter configurations. The distribution shows a symmetrical behavior around the median value of $\tilde{z} = -15$ mm, with a slight skewness that can be observed going from negative offset positions toward positive ones. While the prediction quality strongly decreases at the borders of the presented data, the models' predictions work very well in the central region. In general, one observes that models with a lower learning rate perform slightly better than those with a learning rate equal to or higher than 0.3. Furthermore, Table I reveals that the MAE is reduced by restricting the allowed energy of the test data. The model with hyperparameter configuration ($d = 18, lr = 0.1$) achieved the best MAE performance. For the linearity analysis, we excluded the predictions located

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

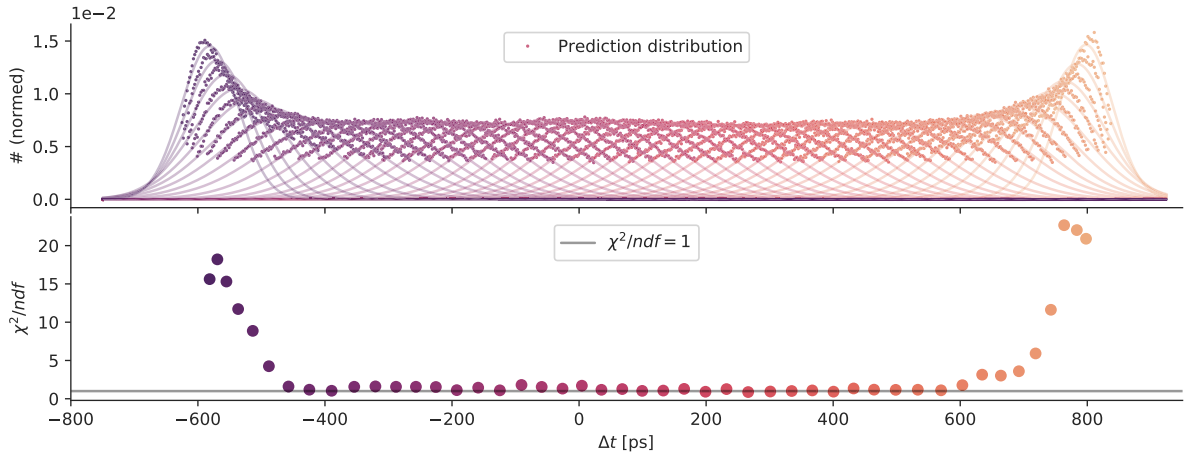


Fig. 7. Depiction of the distribution of predictions of the model using the hyperparameters ($d = 18, lr = 0.1$) using all 47 source positions along the z-axis. The different source positions z_s are encoded in color. The upper plot shows the different histograms and Gauss fits. The lower plot shows the goodness-of-fit per number of degrees of freedom (χ^2/ndf) value for each position. For a large central region, the predictions are in very good agreement with a Gaussian function. When moving toward the edges of the data, the distribution becomes skewed and deviates from the Gaussian shape. No energy filter or restrictions on the measured light distribution were applied.

TABLE I
OVERALL MAE PERFORMANCE FOR DIFFERENT HYPERPARAMETER CONFIGURATIONS AND ENERGY WINDOWS OF THE TEST DATA

Model (d, lr)	MAE [ps] all	MAE [ps] $[300, 700]$ keV	MAE [ps] $[450, 550]$ keV
(12, 0.1)	107.86	85.17	80.18
(12, 0.3)	107.50	84.90	79.75
(12, 0.5)	108.32	85.61	80.41
(15, 0.1)	106.92	84.40	79.32
(15, 0.3)	107.79	85.07	79.91
(15, 0.5)	108.84	85.99	80.80
(18, 0.1)	106.87	84.23	79.09
(18, 0.3)	108.40	85.42	80.23
(18, 0.5)	109.47	86.23	80.93
(20, 0.1)	107.30	84.45	79.24
(20, 0.3)	109.02	85.66	80.30
(20, 0.5)	110.76	87.07	81.76

outside an interval of ±60 mm around the median (grayish areas in Fig. 6) to be able to give an unbiased evaluation of the performance in the large central region of the data.

Fig. 7 shows exemplarily the distribution of the predictions considering the complete data range for the model (18, 0.1) in combination with the goodness-of-fit per number of degrees of freedom (χ^2/ndf) for a Gaussian function. Both distributions are symmetrical. The model can infer the expected time difference on a coincidence basis according to the input data. Considering the goodness-of-fit, the shapes of the predicted distributions are in very good agreement with the expected Gaussian. A substantial deviation from the Gaussian shape is observed when moving toward the far left and far right source positions. A part of the linearity analysis for model (18, 0.1) is exemplarily depicted for the position $(x_s, y_s) = (12,0)$ mm in Fig. 8. The global ε performance for each model is shown in Fig. 9. The estimated ε parameters of all trained models are within a 3σ-interval in agreement with the theoretical value of $\varepsilon = 1$.

B. CTR Performance

The CTR performances of the trained models, as well as the performance of applying only the analytical corrections,

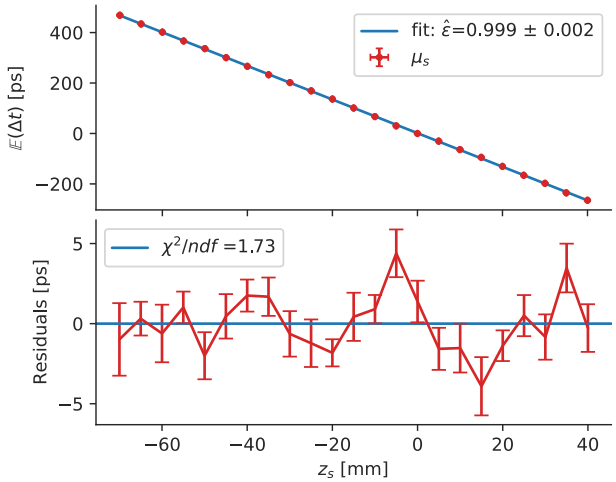


Fig. 8. Linear regression and residual plot of the linearity analysis of model (18, 0.1) for (x_s, y_s) = (12,0) mm. The Gaussian fitting procedure gives the uncertainty on μ_s, while the uncertainty on the z_s position is assumed to be 0.1 mm.

are listed in Table II. As one can see, the best CTR was achieved by the model (18, 0.1), which also performed best regarding the MAE evaluation. The model improved the CTR by about 50 ps down to (185 ± 2) ps for an energy window from 450 to 550 keV. Except for the models having a max. depth of $d = 12$, all other models yield an improved CTR performance when using lower learning rates. A comparison of the time difference distributions before and after using the model (18, 0.1) is depicted in Fig. 10. Regarding the shape of the emerging distribution especially coincidences in the tails of the distribution have been recovered to smaller time differences.

C. SHAP Analysis

The model (18, 0.1) was chosen for the analysis using SHAP [74], [75], [76] since it provided the best performance regarding MAE and CTR. The mean absolute contributions of the different feature sets $\mathcal{F}$ are depicted in Fig. 11. The most

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

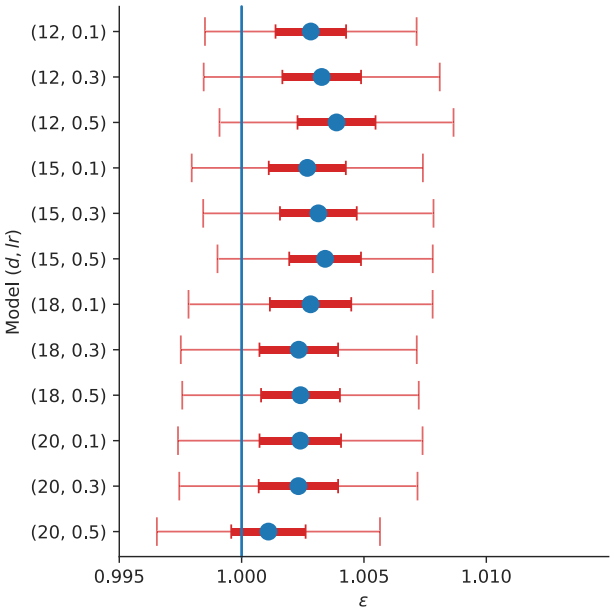


Fig. 9. Global linearity analysis for each trained model. The ε-values are based on the test dataset applying an energy window of 300–700 keV. The thick errorbars represent 1σ-deviation, while the thin errorbars illustrate 3σ-deviation.

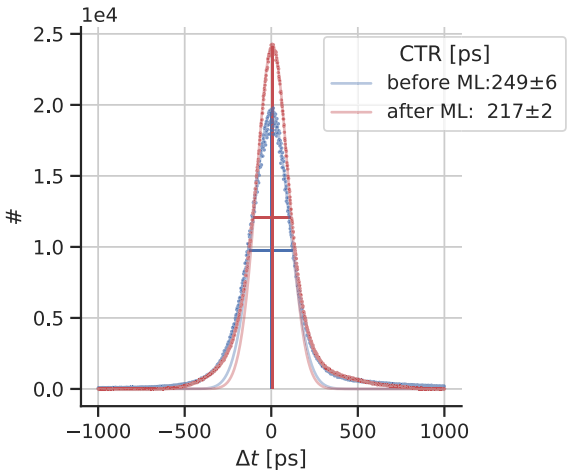


Fig. 10. Time difference distributions before and after using the proposed ML time skew calibration. The model (18, 0.1) was used. No energy windows or restrictions regarding the light distribution are applied.

important feature set is $\mathcal{F}^{so}$, which consists of the measured time difference Δt_{meas}. Besides this substantial contribution, timestamp information $\mathcal{F}_T^{s/o}$, and energy information $\mathcal{F}_E^{s/o}$ also seem to be crucial for good model performance. The feature group $\mathcal{F}_{\text{Pos}}^{s}$ shows a slightly higher contribution compared to $\mathcal{F}_{\text{Pos}}^{o}$, due to the introduction of DOI information. The specific contributions of the planer coordinates, however, differ only marginally for the slab and the one-to-one coupled detector. Furthermore, one observes a similar behavior comparing the feature sets of the slab and the one-to-one coupled detector.

When looking at the progression of the SHAP values $SV(\Delta t_{\text{meas}})$ in dependence on the number of detected optical photons for the SiPM providing the first timestamp ($\#\text{OP}_0^{s/o}$), one observes different developments. Fig. 12(a) shows a clear separation between different SHAP values for a given feature

TABLE II

CTR PERFORMANCE OF THE TRAINED MODELS BASED ON THE PERFORMANCE DATASET. THE RESULTS OF APPLYING ONLY THE ANALYTICAL TIMING CALIBRATION IS DENOTED AS "BEFORE ML"

Model (d, lr)	CTR [ps] all	$[300, 700]$ keV	$[450, 550]$ keV
before ML	249 ± 6	238 ± 5	235 ± 5
(12, 0.1)	230 ± 2	208 ± 2	197 ± 2
(12, 0.3)	224 ± 2	203 ± 2	189 ± 2
(12, 0.5)	225 ± 2	204 ± 1	191 ± 2
(15, 0.1)	223 ± 2	203 ± 2	190 ± 2
(15, 0.3)	225 ± 2	206 ± 2	193 ± 2
(15, 0.5)	232 ± 2	211 ± 2	201 ± 3
(18, 0.1)	217 ± 2	198 ± 2	185 ± 2
(18, 0.3)	226 ± 2	206 ± 2	195 ± 3
(18, 0.5)	233 ± 2	214 ± 2	201 ± 2
(20, 0.1)	220 ± 2	201 ± 2	188 ± 2
(20, 0.3)	225 ± 2	204 ± 2	190 ± 2
(20, 0.5)	236 ± 2	216 ± 1	203 ± 2

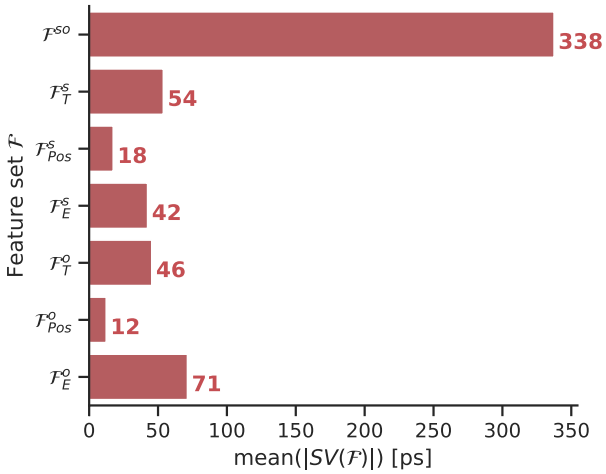


Fig. 11. Mean absolute SHAP values mean($|SV(\mathcal{F})|$) estimated from a subset of the performance dataset for the different feature sets $\mathcal{F}$ explained in Fig. 5. The strongest contribution comes from the shared feature set $\mathcal{F}^{so}$, which consists of the difference between the first timestamps Δt_{meas}. Furthermore, detector-specific information about the timestamps ($\mathcal{F}_T^{s/o}$) and energy information ($\mathcal{F}_E^{s/o}$) are of great importance.

value of Δt_{meas} depending on the number of detected optical photons. This is not observed for the slab detector since, in Fig. 12(b), the strong separation regarding the number of detected optical photons is not given.

VI. DISCUSSION

All models have been trained successfully. The predictions follow a Gaussian function for a large area of the trained data range, as it can be seen in Figs. 6 and 7, as well as in Table I. When moving to the borders of the presented data, the models' outputs deviate from the expected shape, and the prediction quality decreases. This effect is known for many ML algorithms and can be reasoned by the inability to extrapolate to values outside the training range. In future studies, we want to address this issue with different strategies, with using a higher sampling rate at the edges being one of them.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

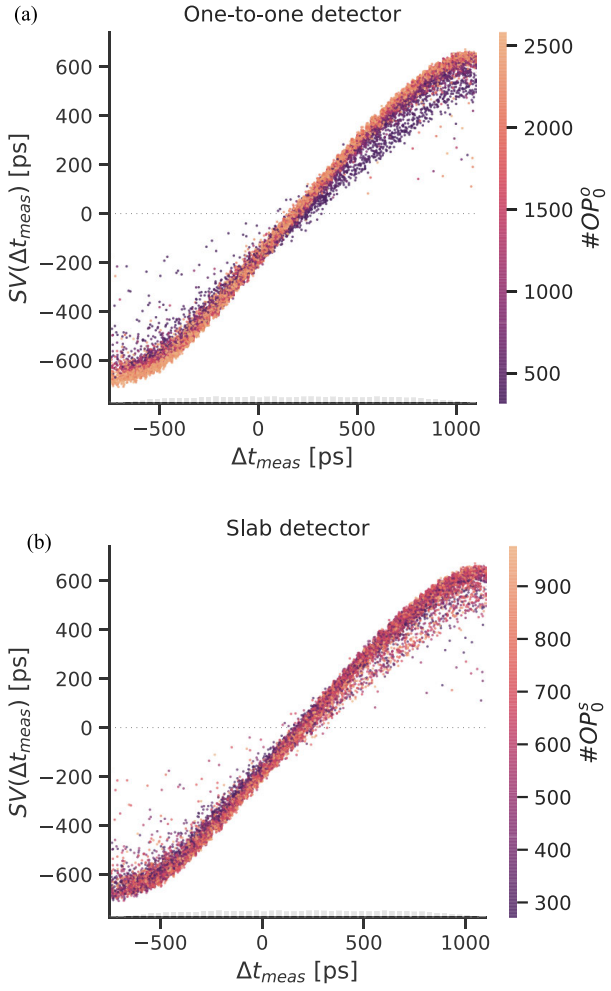


Fig. 12. Progression of the SHAP values $SV(\Delta t_{\text{meas}})$ in dependence of the feature value Δt_{meas} itself. The number of optical photons detected on the SiPM providing the first timestamp ($\#\text{OP}_0^{s/o}$) is encoded in the color (a) for the one-to-one coupled detector and (b) for the slab detector.

Within the central region, where the models show stable behavior, the means of the prediction distributions follow the expected linear relation of (13) (see Fig. 8). No systematic deviation from the linear relation between offset position z_s and predicted mean time difference μ_{s} could be observed, indicating that the trained models are capable of learning the given physical problem. The averaged ε-values are slightly bigger than the expected value of $\varepsilon_{\text{theo}} = 1$ ($\max(\{\varepsilon_i - \varepsilon_{\text{theo}}\}) \leq 3.9 \times 10^{-3}$), which consequently enlarges time differences, and therefore produces an overestimation of the determined CTR values, such that a rescaled resolution might be even better than the here reported one. To compensate for this effect, one could introduce a scaling function $s(\mu_{\text{s}})$ which would correct the slope to the desired value of $\varepsilon = 1$ for a given mean time difference. Since the observed effect is insignificant and the estimated slope factors ε agree for all models with the theoretical value within a 3σ-interval, this procedure is unnecessary for the GBDT models used in this work.

All trained models can improve the achievable CTR values, such that sub-200 ps resolution could be reached for an energy window from 300 keV to 700 keV (see Table II). Minding the shape of the emerging distribution, especially coincidences in the tails of the distribution have been recovered to more minor time differences, indicating that the model can learn physical effects and correct those. This observation underlies the capability of this new approach and shows that the timing resolution can be improved beyond the usage of purely analytical calibrations.

We used explainable AI (XAI) techniques to understand which quantities the models are relying on. The analysis of the SHAP values of the model $(18, 0.1)$ reveals that the reported timestamp difference Δt_{meas} mainly, but also timing and energy information is of great importance. This observation agrees with human intuition, since Δt_{meas} would represent a human's first estimator if one tried to solve the task given to the model. Furthermore, the results clearly indicate that the model is learning timewalk effects for the one-to-one coupled detector [see Fig. 12(a)], since for a given feature value Δt_{meas}, the SHAP value is increased or decreased depending on whether a high or a low number of optical photons has been detected. If a timestamp is affected by a timewalk, the exact moment of timestamping is delayed due to low deposited energy. In conclusion, the importance of this timestamp has to be decreased since it would enlarge the reported time difference and worsen the CTR. This observation does not occur in the same clearness for the slab detector (see Fig. 12(b)). However, for the one-to-one coupled detector, the vast majority of information is contained in one channel, whereas, for the semi-monolithic case, the information is spread across multiple channels, making it hard to display the effect in the chosen visualization. There is still an indication that also for the slab detector, timewalk effects are caught by the model since the feature set using energy-related quantities shows a high absolute SHAP value (see Fig. 11) and that both tails of the time difference distribution are reduced.

VII. Conclusion and Outlook

In this work, we demonstrated a new approach based on the combination of residual physics and ML to address real-world physics-based problems. We applied the concept to detector calibration. We hope the work highlights the potential for applications of learning systems along all computing steps of complex acquisition and processing systems and, thus, may inspire future research.

Since the formalism settles on previously linear corrected timestamps [33], it can be seen as a first approach toward residual physics in timing calibration. All models could be trained successfully and are in a 3σ agreement with the underlying physical relation. The first results indicate that this new calibration strategy has provoked a strong improvement in the achievable CTR reaching from 238 ps down to 198 ps for an energy window from 300 keV to 700 keV, and from 235 ps even down to 185 ps for a smaller energy window from 450 keV to 550 keV. The SHAP analysis offers a strong indication that the proposed technique has the capability to build physics-informed models.

All results are based on experimentally acquired data from two clinically relevant detector arrays. This work and the corresponding promising first results represent a proof-of-concept

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

for future time skew calibration techniques relying on AI. Nevertheless, several studies have to be performed before an application to a complete PET system is possible. The presented technique is currently implemented for a pair of detectors utilizing digital SiPMs. Research toward systems of multiple detectors will be addressed in future works. Besides this, we want to explore the performance of the concept in different environmental settings (e.g., higher measurement temperatures and different readouts), potentially enlarging the learning system's importance. Furthermore, the reduction and study of the influence of the needed data acquisition time and the bias effects toward the edges of the training data is mandatory for possible usage in a clinical scanner. A possible method to address this point would be an artificial enlargement of the available training data found on only a few measured data points. However, we expect the measurement time to increase weaker than linearly with the number of detectors since one source position can be used for many detectors.

REFERENCES

[1] J. B. Wagner, "Artificial intelligence in medical imaging," *Radiologic Technol.*, vol. 90, no. 5, pp. 489–501, May 2019.

[2] M. G. Carneiro and L. Zhao, "Organizational data classification based on the importance concept of complex networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3361–3373, Aug. 2018, doi: 10.1109/TNNLS.2017.2726082.

[3] J. Zhang, Z. Cui, C. Jiang, S. Guo, F. Gao, and D. Shen, "Hierarchical organ-aware total-body standard-dose PET reconstruction from low-dose PET and CT images," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, May 9, 2023, 10.1109/TNNLS.2023.3266551.

[4] T. Buzug, *Computed Tomography*. Berlin, Germany: Springer, 2008, doi: 10.1007/978-3-540-39408-2.

[5] M. T. Vlaardingerbroek and J. A. Den Boer, *Magnetic Resonance Imaging*. Berlin, Germany: Springer, 2003, doi: 10.1007/978-3-662-05252-5.

[6] S. S. Gambhir, "Molecular imaging of cancer with positron emission tomography," *Nat. Rev. Cancer*, vol. 2, no. 9, pp. 683–693, 2002, doi: 10.1038/nrc882.

[7] G. Muehllehner and J. S. Karp, "Positron emission tomography," *Phys. Med. Biol.*, vol. 51, no. 13, pp. R117–R137, Jun. 2006, doi: 10.1088/0031-9155/51/13/R08.

[8] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, Dec. 2016, pp. 241–246, doi: 10.1109/ICDMW.2016.0041.

[9] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017, doi: 10.1109/TIP.2017.2713099.

[10] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of deep learning and reinforcement learning to biological data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2063–2079, Jun. 2018, doi: 10.1109/TNNLS.2018.2790388.

[11] M. van den Ende, I. Lior, J.-P. Ampuero, A. Sladen, A. Ferrari, and C. Richard, "A self-supervised deep learning approach for blind denoising and waveform coherence enhancement in distributed acoustic sensing data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 7, pp. 3371–3384, Jul. 2023, doi: 10.1109/TNNLS.2021.3132832.

[12] W. Samek, T. Wiegand, and K.-R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," 2017, *arXiv:1708.08296*, doi: 10.48550/arXiv.1708.08296.

[13] W. Samek and K.-R. Müller, "Towards explainable artificial intelligence," in *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning* (Lecture Notes in Computer Science), W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, and K.-R. Müller, Eds. Cham, Switzerland: Springer, 2019, pp. 5–22, doi: 10.1007/978-3-030-28954-61.

[14] A. Barredo Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.

[15] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4793–4813, Nov. 2021, doi: 10.1109/TNNLS.2020.3027314.

[16] M. Conti and B. Bendriem, "The new opportunities for high time resolution clinical TOF PET," *Clin. Translational Imag.*, vol. 7, no. 2, pp. 139–147, Apr. 2019, doi: 10.1007/s40336-019-00316-5.

[17] S. Surti, "Update on time-of-flight PET imaging," *J. Nucl. Med.*, vol. 56, no. 1, pp. 98–105, Jan. 2015, doi: 10.2967/jnumed.114.145029.

[18] J. van Sluis et al., "Performance characteristics of the digital biograph vision PET/CT system," *J. Nucl. Med.*, vol. 60, no. 7, pp. 1031–1036, Jul. 2019, doi: 10.2967/jnumed.118.215418.

[19] G. A. Prenosil et al., "Performance characteristics of the biograph vision quadra PET/CT system with a long axial field of view using the NEMA NU 2-2018 standard," *J. Nucl. Med.*, vol. 63, no. 3, pp. 476–484, Mar. 2022, doi: 10.2967/jnumed.121.261972.

[20] S. Krishnamoorthy et al., "Performance evaluation of the MOLECUBES β-CUBE—A high spatial resolution and high sensitivity small animal PET scanner utilizing monolithic LYSO scintillation detectors," *Phys. Med. Biol.*, vol. 63, no. 15, Jul. 2018, Art. no. 155013, doi: 10.1088/1361-6560/aacec3.

[21] A. Gonzalez-Montoro et al., "A new brain dedicated PET scanner with 4D detector information," *Bio-Algorithms Med.-Syst.*, vol. 18, no. 1, pp. 107–119, Dec. 2022, doi: 10.2478/bioal-2022-0083.

[22] F. Müller, D. Schug, P. Hallen, J. Grahe, and V. Schulz, "Gradient tree boosting-based positioning method for monolithic scintillator crystals in positron emission tomography," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 2, no. 5, pp. 411–421, Sep. 2018, doi: 10.1109/TRPMS.2018.2837738.

[23] W. He et al., "A preliminary study on 3D position reconstruction of monolithic crystal readout," *Radiat. Detection Technol. Methods*, vol. 5, no. 1, pp. 102–109, Mar. 2021, doi: 10.1007/s41605-020-00225-6.

[24] M. Decuyper, M. Stockhoff, S. Vandenberghe, and R. Van Holen, "Artificial neural networks for positioning of gamma interactions in monolithic PET detectors," *Phys. Med. Biol.*, vol. 66, no. 7, Mar. 2021, Art. no. 075001, doi: 10.1088/1361-6560/abebfc.

[25] F. Müller, D. Schug, P. Hallen, J. Grahe, and V. Schulz, "A novel DOI positioning algorithm for monolithic scintillator crystals in PET based on gradient tree boosting," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 4, pp. 465–474, Jul. 2019, doi: 10.1109/TRPMS.2018.2884320.

[26] F. Loignon-Houle et al., "DOI estimation through signal arrival time distribution: A theoretical description including proof of concept measurements," *Phys. Med. Biol.*, vol. 66, no. 9, Apr. 2021, Art. no. 095015, doi: 10.1088/1361-6560/abf604.

[27] R. Vinke et al., "Time walk correction for TOF-PET detectors based on a monolithic scintillation crystal coupled to a photosensor array," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 621, nos. 1–3, pp. 595–604, Sep. 2010, doi: 10.1016/j.nima.2010.05.034.

[28] J. Du, J. P. Schmall, M. S. Judenhofer, K. Di, Y. Yang, and S. R. Cherry, "A time-walk correction method for PET detectors based on leading edge discriminators," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 1, no. 5, pp. 385–390, Sep. 2017, doi: 10.1109/TRPMS.2017.2726534.

[29] S. Xie et al., "Optical simulation and experimental assessment with time–walk correction of TOF–PET detectors with multi-ended readouts," *Sensors*, vol. 21, no. 14, p. 4681, Jul. 2021, doi: 10.3390/s21144681.

[30] C. J. Thompson, M.-L. Camborde, and M. E. Casey, "A central positron source to perform the timing alignment of detectors in a PET scanner," in *Proc. IEEE Symp. Conf. Rec. Nucl. Sci.*, Oct. 2004, pp. 2361–2365, doi: 10.1109/NSSMIC.2004.1462731.

[31] D. Schug et al., "Data processing for a high resolution preclinical PET detector based on Philips DPC digital SiPMs," *IEEE Trans. Nucl. Sci.*, vol. 62, no. 3, pp. 669–678, Jun. 2015, doi: 10.1109/TNS.2015.2420578.

[32] H. U. Dike, Y. Zhou, K. K. Deveerasetty, and Q. Wu, "Unsupervised learning based on artificial neural network: A review," in *Proc. IEEE Int. Conf. Cyborg Bionic Syst. (CBS)*, Oct. 2018, pp. 322–327, doi: 10.1109/CBS.2018.8612259.

[33] S. Naunheim, Y. Kuhl, T. Solf, D. Schug, V. Schulz, and F. Mueller, "Analysis of a convex time skew calibration for light sharing-based PET detectors," *Phys. Med. Biol.*, vol. 68, no. 2, 2022, Art. no. 025013, doi: 10.1088/1361-6560/aca872.

[34] M. Korzhik, G. Tamulaitis, and A. N. Vasilév, *Physics of Fast Processes in Scintillators* (Particle Acceleration and Detection). Cham, Switzerland: Springer, 2020, doi: 10.1007/978-3-030-21966-6.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[35] F. Daví, "A brief overview of existence results and decay time estimates for a mathematical modeling of scintillating crystals," *Math. Methods Appl. Sci.*, vol. 44, no. 18, pp. 13833–13854, Dec. 2021, doi: 10.1002/mma.7660.

[36] S. Seifert et al., "A comprehensive model to predict the timing resolution of SiPM-based scintillation detectors: Theory and experimental validation," *IEEE Trans. Nucl. Sci.*, vol. 59, no. 1, pp. 190–204, Feb. 2012, doi: 10.1109/TNS.2011.2179314.

[37] J. W. Cates, R. Vinke, and C. S. Levin, "Analytical calculation of the lower bound on timing resolution for PET scintillation detectors comprising high-aspect-ratio crystal elements," *Phys. Med. Biol.*, vol. 60, no. 13, pp. 5141–5161, Jun. 2015, doi: 10.1088/0031-9155/60/13/5141.

[38] H. T. van Dam, G. Borghi, S. Seifert, and D. R. Schaart, "Sub-200 ps CRT in monolithic scintillator PET detectors using digital SiPM arrays and maximum likelihood interaction time estimation," *Phys. Med. Biol.*, vol. 58, no. 10, pp. 3243–3257, May 2013, doi: 10.1088/0031-9155/58/10/3243.

[39] S. Naunheim, T. Solf, Y. Kuhl, D. Schug, and V. Schulz, "Towards 200 ps CRT in DOI-capable semi-monolithic PET-detectors for clinical applications," in *Proc. IEEE Nuclear Sci. Symp. Med. Imag. Conf.* Milano, Italy: IEEE, Oct. 2021.

[40] A. Zeng, S. Song, J. Lee, A. Rodriguez, and T. Funkhouser, "TossingBot: Learning to throw arbitrary objects with residual physics," *IEEE Trans. Robot.*, vol. 36, no. 4, pp. 1307–1319, Aug. 2020, doi: 10.1109/TRO.2020.2988642.

[41] Q. Zhang, M. H. B. Mahbod, C.-B. Chng, P.-S. Lee, and C.-K. Chui, "Residual physics and post-posed shielding for safe deep reinforcement learning method," *IEEE Trans. Cybern.*, early access, Jun. 14, 2022, doi: 10.1109/TCYB.2022.3178084.

[42] F. Mueller, S. Naunheim, Y. Kuhl, D. Schug, T. Solf, and V. Schulz, "A semi-monolithic detector providing intrinsic DOI-encoding and sub-200 ps CRT TOF-capabilities for clinical PET applications," *Med. Phys.*, vol. 49, no. 12, pp. 7469–7488, Dec. 2022, doi: 10.1002/mp.16015.

[43] C. Wassermann et al., "High throughput software-based gradient tree boosting positioning for PET systems," *Biomed. Phys. Eng. Exp.*, vol. 7, no. 5, Aug. 2021, Art. no. 055023, doi: 10.1088/2057-1976/ac11c0.

[44] P. Abbeel, M. Quigley, and A. Y. Ng, "Using inaccurate models in reinforcement learning," in *Proc. 23rd Int. Conf. Mach. Learn.* New York, NY, USA: Association for Computing Machinery, Jun. 2006, pp. 1–8, doi: 10.1145/1143844.1143845.

[45] P. Pastor et al., "Learning task error models for manipulation," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2013, pp. 2612–2618, doi: 10.1109/ICRA.2013.6630935.

[46] J. C. G. Higuera, D. Meger, and G. Dudek, "Adapting learned robotics behaviours through policy adjustment," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2017, pp. 5837–5843, doi: 10.1109/ICRA.2017.7989686.

[47] T. Silver, K. R. Allen, J. Tenenbaum, and L. Kaelbling. (Dec. 2018). *Residual Policy Learning*, doi: 10.48550/arXiv.1812.06298.

[48] T. Johannink et al., "Residual reinforcement learning for robot control," in *Proc. Int. Conf. Robot. Automat. (ICRA)*, May 2019, pp. 6023–6029, doi: 10.1109/ICRA.2019.8794127.

[49] A. Kloss, S. Schaal, and J. Bohg, "Combining learned and analytical models for predicting action effects from sensory data," *Int. J. Robot. Res.*, vol. 41, no. 8, pp. 778–797, Jul. 2022, doi: 10.1177/0278364920954896.

[50] S. Fu, S. Zhong, L. Lin, and M. Zhao, "A novel time-series memory auto-encoder with sequentially updated reconstructions for remaining useful life prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 7114–7125, Dec. 2022, doi: 10.1109/TNNLS.2021.3084249.

[51] C. Oszkinat, S. E. Luczak, and I. G. Rosen, "Uncertainty quantification in estimating blood alcohol concentration from transdermal alcohol level with physics-informed neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 8094–8101, Oct. 2022, doi: 10.1109/TNNLS.2022.3140726.

[52] Y. Li, S. He, Y. Li, Y. Shi, and Z. Zeng, "Federated multiagent deep reinforcement learning approach via physics-informed reward for multimicrogrid energy management," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jan. 3, 2023, doi: 10.1109/TNNLS.2022.3232630.

[53] J. Hua, Y. Li, C. Liu, P. Wan, and X. Liu, "Physics-informed neural networks with weighted losses by uncertainty evaluation for accurate and stable prediction of manufacturing systems," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 7, 2023, doi: 10.1109/TNNLS.2023.3247163.

[54] X. Zhang et al., "Performance of long rectangular semi-monolithic scintillator PET detectors," *Med. Phys.*, vol. 46, no. 4, pp. 1608–1619, Apr. 2019, doi: 10.1002/mp.13432.

[55] C. Zhang et al., "A thick semi-monolithic scintillator detector for clinical PET scanners," *Phys. Med. Biol.*, vol. 66, no. 6, Mar. 2021, Art. no. 065023, doi: 10.1088/1361-6560/abe761.

[56] M. Freire et al., "Position estimation using neural networks in semi-monolithic PET detectors," *Phys. Med. Biol.*, vol. 67, no. 24, Dec. 2022, Art. no. 245011, doi: 10.1088/1361-6560/aca389.

[57] D. Sánchez et al., "HRFlexToT: A high dynamic range ASIC for time-of-flight positron emission tomography," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 6, no. 1, pp. 51–67, Jan. 2022, doi: 10.1109/TRPMS.2021.3066426.

[58] J. Maebe and S. Vandenberghe, "Simulation study on 3D convolutional neural networks for time-of-flight prediction in monolithic PET detectors using digitized waveforms," *Phys. Med. Biol.*, vol. 67, no. 12, Jun. 2022, Art. no. 125016, doi: 10.1088/1361-6560/ac73d3.

[59] N. Cucarella, J. Barrio, E. Lamprou, C. Valladares, J. M. Benlloch, and A. J. Gonzalez, "Timing evaluation of a PET detector block based on semi-monolithic LYSO crystals," *Med. Phys.*, vol. 48, no. 12, pp. 8010–8023, Dec. 2021, doi: 10.1002/mp.15318.

[60] E. Berg and S. R. Cherry, "Using convolutional neural networks to estimate time-of-flight from PET detector waveforms," *Phys. Med. Biol.*, vol. 63, no. 2, Jan. 2018, Art. no. 02LT01, doi: 10.1088/1361-6560/aa9dc5.

[61] Y. Onishi, F. Hashimoto, K. Ote, and R. Ota, "Unbiased TOF estimation using leading-edge discriminator and convolutional neural network trained by single-source-position waveforms," *Phys. Med. Biol.*, vol. 67, no. 4, Feb. 2022, Art. no. 04NT01, doi: 10.1088/1361-6560/ac508f.

[62] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining* New York, NY, USA: Association for Computing Machinery, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[63] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. New York, NY, USA: Taylor & Francis, Oct. 2017, P. 358, doi: 10.1201/9781315139470.

[64] J. Tang, C. Deng, and G.-B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 4, pp. 809–821, Apr. 2016, doi: 10.1109/TNNLS.2015.2424995.

[65] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021, doi: 10.1109/TNNLS.2020.2978386.

[66] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.

[67] Y. Xia, C. Liu, Y. Li, and N. Liu, "A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring," *Exp. Syst. Appl.*, vol. 78, pp. 225–241, Jul. 2017, doi: 10.1016/j.eswa.2017.02.017.

[68] L. Torlay, M. Perrone-Bertolotti, E. Thomas, and M. Baciu, "Machine learning–XGBoost analysis of language networks to classify patients with epilepsy," *Brain Informat.*, vol. 4, no. 3, pp. 159–169, Sep. 2017, doi: 10.1007/s40708-017-0065-7.

[69] M. Aaboud et al., "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector," *Phys. Lett. B*, vol. 784, pp. 173–191, Sep. 2018, doi: 10.1016/j.physletb.2018.07.035.

[70] Z. Zhang and C. Jung, "GBDT-MO: gradient-boosted decision trees for multiple outputs," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 3156–3167, Jul. 2021, doi: 10.1109/TNNLS.2020.3009776.

[71] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Inf. Fusion*, vol. 81, pp. 84–90, May 2022, doi: 10.1016/j.inffus.2021.11.011.

[72] M. Shepovalov and V. Akella, "FPGA and GPU-based acceleration of ML workloads on Amazon cloud—A case study using gradient boosted decision tree library," *Integration*, vol. 70, pp. 1–9, Jan. 2020, doi: 10.1016/j.vlsi.2019.09.007.

[73] A. Alcolea and J. Resano, "FPGA accelerator for gradient boosting decision trees," *Electronics*, vol. 10, no. 3, p. 314, Jan. 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/3/314, doi: 10.3390/electronics10030314.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[74] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2017, pp. 4766–4775, doi: 10.48550/arXiv.1705.07874.

[75] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent individualized feature attribution for tree ensembles," 2018, *arXiv:1802.03888*, doi: 10.48550/arXiv.1802.03888.

[76] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," *Nature Mach. Intell.*, vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: 10.1038/s42256-019-0138-9.

[77] L. S. Shapley, "17. A value for N-person games," in *Contributions to the Theory of Games (AM-28)*, vol. 2, H. W. Kuhn and A. W. Tucker, Eds. Princeton, NJ, USA: Princeton Univ. Press, Dec. 1953, pp. 307–318, doi: 10.1515/9781400881970-018.

[78] T. Frach, G. Prescher, C. Degenhardt, and B. Zwaans, "The digital silicon photomultiplier—System architecture and performance evaluation," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, Oct. 2010, pp. 1722–1727, doi: 10.1109/NSSMIC.2010.5874069.

[79] V. Tabacchini, V. Westerwoudt, G. Borghi, S. Seifert, and D. R. Schaart, "Probabilities of triggering and validation in a digital silicon photomultiplier," *J. Instrum.*, vol. 9, no. 6, Jun. 2014, Art. no. P06016, doi: 10.1088/1748-0221/9/06/P06016.

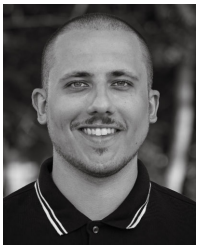
[80] Philips Digital Photon Counting. (2016). *Tile-TEK User Manual*.

[81] R. Hetzel, F. Mueller, J. Grahe, A. Honné, D. Schug, and V. Schulz, "Characterization and simulation of an adaptable fan-beam collimator for fast calibration of radiation detectors for PET," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 4, no. 5, pp. 538–545, Sep. 2020, doi: 10.1109/TRPMS.2020.2990651.

[82] A. B. Mann, S. Paul, A. Tapfer, V. C. Spanoudaki, and S. I. Ziegler, "A computing efficient PET time calibration method based on pseudoinverse matrices," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf.*, Oct. 2009, pp. 3889–3892, doi: 10.1109/NSSMIC.2009.5401925.

[83] P. D. Reynolds, P. D. Olcott, G. Pratx, F. W. Y. Lau, and C. S. Levin, "Convex optimization of coincidence time resolution for a high-resolution PET system," *IEEE Trans. Med. Imag.*, vol. 30, no. 2, pp. 391–400, Feb. 2011, doi: 10.1109/TMI.2010.2080282.

[84] (2008). *Orthogonal Distance Regression (Scipy.odr)—SciPy v1.9.3 Manual*.



Stephan Naunheim (Member, IEEE) received the M.Sc. degree (Hons.) in physics from RWTH Aachen University, Aachen, Germany, in 2021. He is currently pursuing the Ph.D. degree with the Department of Physics of Molecular Imaging Systems (PMI), part of the Institute for Experimental Molecular Imaging (ExMI), RWTH Aachen University.

During his studies, he specialized in particle physics and machine learning. His research interests include fundamental and applied machine-learning techniques, with a particular focus on connecting artificial learning systems with prior domain knowledge. Furthermore, he is interested in explainable AI (XAI) techniques to understand computational intelligence better.

Yannick Kuhl (Member, IEEE) received the M.Sc. degree (Hons.) in physics from RWTH Aachen University, Aachen, Germany, in 2021. He is currently pursuing the Ph.D. degree with the Department of Physics of Molecular Imaging Systems (PMI), part of the Institute for Experimental Molecular Imaging (ExMI), RWTH Aachen University.

His research interests include PET system design and development as well as integration with MRI, with a focus on photodetector design and calibration using machine-learning techniques.

David Schug (Member, IEEE) received the Dr. rer. nat. degree (Hons.) in physics from RWTH Aachen University, Aachen, Germany, in 2015.

From 2006 to 2011, he studied physics at RWTH Aachen University. Since 2016, he has been the Group Leader of the PET detector physics group of Physics of Molecular Imaging Systems (PMI) headed by Prof. Volkmar Schulz. The PET detector physics group focuses on developing new PET detector concepts starting with scintillator designs to photosensor readout topologies and the required calibration and data processing techniques up to information transfer to image reconstruction frameworks. In 2019, he Co-Founded and has since then been the Co-Director of Hyperion Hybrid Imaging Systems GmbH, a spin-off of PMI with the goal to commercialize the technologies developed at PMI. His research interests include PET detectors and systems with focus on sensor data processing, required calibration techniques and how machine learning can be applied to the large variety of problems encountered in medical imaging.

Volkmar Schulz (Senior Member, IEEE) received the electrical engineering degree from Bielefeld University, Bielefeld, Germany, in 1993, the electrical engineering degree from the University of Paderborn, Paderborn, Germany, in 1996, and the doctorate degree in electrical engineering from the University of Paderborn in 2001.

In 2000, he initiated his industrial research at Philips Research Laboratories, Hamburg, Germany, with a focus on MRI systems. In 2005, he specialized in PET and MRI integration, ascending to the role of Principal Scientist. In 2013, he became a Full Professor at the Department of Physics of Molecular Imaging Systems (PMI), the Institute for Experimental Molecular Imaging (ExMI), RWTH Aachen University. His PMI develops full PET systems, including hardware such as mechanics and electronics, firmware, and software, often in the context of being operated together with a magnetic resonance tomography system. In 2019, he Co-Founded Hyperion Hybrid Imaging Systems and currently holds a position there. His primary research areas encompass medical imaging and machine learning.

Florian Mueller (Member, IEEE) received the Dr.rer.nat. degree (Hons.) in physics from RWTH Aachen University, Aachen, Germany, in 2023.

He is currently leading the System Design and Integration Group with the Department of Physics of Molecular Imaging Systems (PMI), part of the Institute for Experimental Molecular Imaging (ExMI), RWTH Aachen University. The group revolves around critical considerations such as the scalability of components and algorithms, with a particular focus on the challenges associated with integrating multimodal imaging systems, especially PET/MRI devices. His doctoral research focused on the development of calibration methods and algorithms tailored for light-sharing detectors in positron emission tomography (PET). His current research interests include machine learning, the intersection of learning systems with present knowledge, and dedicated calibration techniques and their application to advance large-scale and high-performance PET and PET/MRI scanners.