UC Davis
UC Davis Previously Published Works

Title

The reduction of 176Lu background in Lu-based PET scanners using optimized classification

Permalink

Journal

Physics in Medicine and Biology, 65(17)

ISSN

0031-9155

Authors

Wang, Qian
Mikhaylova, Ekaterina
Baikejiang, Reheman
et al.

Publication Date

2020-09-07

DOI

10.1088/1361-6560/aba088

Peer reviewed

The reduction of [176]Lu background in Lu-based PET scanners using optimized classification

Qian Wang[1], **Ekaterina Mikhaylova**[1], **Reheman Baikejiang**[1], **Simon R Cherry**[1,2]

[1]Department of Biomedical Engineering, University of California, Davis, CA, United States

[2]Department of Radiology, University of California, Davis, CA, United States

Abstract

Positron emission tomography (PET) using scanners incorporating lutetium-based (Lu-based) scintillators are widely used in nuclear medicine. However their application in imaging very low (< 100 kBq) activity distributions is quite limited due to the intrinsic [176]Lu radiation emitted from the scintillators. To visualize very low activities, [176]Lu background needs to be reduced or removed. This study proposes a classification method to select background coincidences from true coincidences arising from the source by supervised learning using the optimal classifier as determined by investigating 5 different classifiers: logistic regression, support vector machine, random forest, extreme gradient boosting (XGBoost) and deep neural network. Five energy and time-of-flight (TOF) related features from each coincidence event are extracted to form the training and test set in the classification. The proposed method was verified on a pair of TOF-PET detector modules. Since the measured source coincidences cannot be differentiated from the background events experimentally, simulated source coincidences are used to train the classification model. The simulated feature spectra are therefore compared with those obtained from measurement to verify the feasibility of classifying measured coincidences using a model learned by simulation. XGBoost classifier performed most effectively in classifying the coincidences and provided impressively high classification accuracy (> 99%). It was subsequently tested by imaging point-like source, planar Derenzo and bar phantoms with the pair of TOF-PET detectors. An 89.4% image contrast enhancement for the Derenzo phantom at an activity concentration of 100 Bq/mm^2, and a 52.4% peak-to-valley ratio improvement across the area of bar phantom at a concentration of 25 Bq/mm^2, were observed on the reconstructed images with XGBoost classification applied. The proposed method could extend the usage of Lu-based PET scanners to very low activity detection and imaging and has the potential to be used in a variety of molecular imaging tasks to detect low-level signals.

1. Introduction

Positron emission tomography (PET), a powerful molecular imaging modality, has experienced continuous performance improvements in recent years and is being applied in an increasing number of areas (Du et al., 2018). Due to the inherent sensitivity in

detecting radiotracers, PET has great potential to also be used for low-activity and very low-activity multimodal biomedical imaging applications, including pediatric and obstetric imaging (Gjelsteen et al., 2008; Schäfer et al., 2014), gene expression imaging (Gambhir et al., 1999), in-beam imaging (Pawelke et al., 1997), cell trafficking (Thunemann et al., 2017), ^{90}Y PET imaging (Carlier et al., 2015; Strydhorst et al., 2016) and even single photon emission computed tomography imaging using detectors/systems primarily designed for PET (Yao et al., 2008). However, due to the intrinsic radiation emitted from the small fraction of radioactive ^{176}lutetium (^{176}Lu) residing in widely used Lu-based scintillators, such as lutetium oxyorthosilicate (LSO) and lutetium yttrium orthosilicate (LYSO), the visualization and quantification of stationary or moving very low-activity sources in PET scanners using these materials becomes quite challenging (Conti et al., 2017). Scanners made from scintillators without Lu, like bismuth germanate (BGO), seem to be more suitable for very low activity imaging (Ouyang et al., 2016). However Lu-based PET scanners are commonly used and available, and offer higher overall performance especially in terms of timing resolution, which enables time-of-flight (TOF) measurement (Cherry et al., 2017). Therefore, efforts to develop methods for very low activity imaging on Lu-based PET scanners are important.

In order to use Lu-based PET scanners for imaging the spatiotemporal distribution of radiotracers administered at very low (< 100 kBq) activities, the intrinsic radiation must be reduced before image reconstruction. Most reported methods for reducing ^{176}Lu background adjust the energy window (Yoshida et al., 2014). ^{176}Lu decays into ^{176}Hf by emitting an electron whose end point energy is 596 keV with a branching ratio of 99.6% or an electron whose end point energy is 195 keV with a branching ratio of 0.4%, with a cascade of gamma photons with energies of 401 keV (0.4% branching ratio, only occurs when the 195-keV end-point electron is emitted), 307 keV (94%), 202 keV (78%) and 88 keV (15%) (Conti et al., 2017; Alva-Sánchez et al., 2018). The three common gamma photons have different probabilities of escaping the original crystal and the probability also depends on the crystal-size. As small crystals are typically used in PET detectors, the 307-keV and 202-keV gamma photons are energetic enough that they have a high probability to escape the original crystal in which they were produced. For instance, the probabilities of 307-keV and 202-keV gamma photons escaping $1{\times}1{\times}1$ cm^3 LYSO crystals are 0.618 and 0.346, respectively (Alva-Sánchez et al., 2018), and these can then be detected by another crystal within the timing window opened by the original electron, forming true background coincidences. The intrinsic (background) coincidences consist of intrinsic true coincidences that mostly arise from the pairing of an electron emitted during ^{176}Lu decay in a crystal with a prompt gamma photon from the same decay that interacts in another crystal, as well as intrinsic random coincidences that involve two ^{176}Lu gamma photons, two electrons, an electron and a (^{176}Lu or 511 keV) gamma photon, or an ^{176}Lu and a 511 keV gamma photon from different decays. The intrinsic true background coincidences can be effectively eliminated by narrowing the energy window. However, for intrinsic random coincidences, the deposited energy of the particles from ^{176}Lu decay can vary from 88 keV to higher than 1 MeV due to the joint detection of conversion electrons and gamma photons that do not escape the original crystal and the multiple scattering of 511 keV gamma photons in crystals induces further results in a spread of energies

(Alva-Sánchez et al., 2018). Therefore, the intrinsic random coincidences cannot be fully removed by narrowing the energy window. The number of coincidences that remain can still be comparable, or even greater in number, than true coincidences from a very low-activity source. For instance, it is reported that ~1200 intrinsic random coincidences are obtained per second compared to only 0.174 counts/s/MBq from an ^{90}Y source (true unscattered coincidences) for ^{90}Y PET imaging on the Biograph mCT scanner (Siemens Healthineers, Knoxville, TN) using a narrow energy window 435-650 keV applied (Strydhorst et al., 2016). Simultaneously, true coincidences are also reduced by narrowing the energy window, hindering the detection of a very low activity source. Moreover, the conventional "singles rate" and "delayed window" random correction approaches are based on substracting the statistically estimated random coincidences from measured coincidences for a given detector pair (Cherry and Dahlbom, 2006). In a very low-activity situation, the estimated random coincidences fluctuate spatiotemporally, and these statistical fluctuations make it difficult to robustly extract the small number of true coincidence events.

Similarly, statistics-based methods are employed for background reduction, for instance, subtracting a time-averaged background of a long blank scan from the measured coincidences or sinograms. However, this does not account for attenuation and scatter of the intrinsic background signal by the object. This approach is insufficient for intrinsic background removal in the presence of a very low-activity source.

For some scanners that are able to provide time-of-flight (TOF) information, this also can be used in combination with the energy window to more effectively reduce the ^{176}Lu background (Yoshida et al., 2014). The TOF information works best for localized sources such as in in-beam imaging, and will be less effective when the source activity is distributed across the whole field of view (FOV) (Budinger, 1983; Conti and Bendriem, 2019).

Focusing on the specific application of very low activity imaging, this study proposes a classification methodology which exploits a classification model established by training the energy- and TOF-related features which are present within detected coincidence events to distinguish events according to their origin, source or background, while leaving the energy window intact. Different classification methods/classifiers and parameters are examined to determine which of them results in the highest classification accuracy (*i.e.* background reduction efficiency). The proposed approach is experimentally validated by imaging different phantoms. Reconstructed images preprocessed with and without classification are compared to demonstrate the effect of the approach.

2. Methods

2.1. Classification methods

In order to reduce the intrinsic background, we established a platform embedded with five effective and commonly-used classifiers, including logistic regression (LR), support vector machine (SVM), random forest (RF), extreme gradient boosting (XGBoost) and deep neural network (DNN). In this study coincidences are classified into two classes: background coincidences and coincidences induced by the source, *i.e.* source coincidences. Five features of each individual coincidence event are extracted from the list mode data recorded: the

energy deposited in each of the two detectors, the average and the difference of the energies, and the difference of the time stamps of the two photons detected in the coincidence event. The input data for training each classification model is composed of coincidence features and corresponding class labels that are already known and are randomly split into training and validation data/coincidences. 80% of the input data is the training data used to fit the classification model, and 20% of the input data is the validation data used to provide an evaluation of the model fit on the training data while tuning the model hyperparameters. Unlike background coincidences, source coincidences cannot be measured independently as the background is always involved in the measurement. Therefore, the source coincidences used in the training set come from Monte Carlo simulation. Monte Carlo simulation of the acquisition of coincidence events with Lu-based PET detectors is explained in Section 2.3. The background coincidence component can come from either measurement or simulation and measured background coincidences are chosen for this study. The test data is prepared for evaluating the established classification model and contains the features of coincidences to be classified, however the class labels of the coincidences may be known or unknown. The test datasets with known class labels are used for estimating the classification accuracy. The class labels of measured coincidences are unknown in the test dataset and are the target of the classification. All targeted coincidences are identified with the classifier to filter out background coincidences before image reconstruction.

All classification algorithms, learning and prediction are implemented in Python. Scikit-learn (http://scikit-learn.org/stable/index.html) (Pedregosa et al., 2011), an efficient machine learning tool in Python, is used for LR (class: sklearn.linear_model.LogisticRegression), RF (class: sklearn.ensemble.RandomForestClassifier) and SVM (class: sklearn.svm.SVC) classifications. XGBoost classification uses a Python XGBoost package (https://xgboost.readthedocs.io/en/latest//) (Chen and Guestrin, 2016; Friedman, 2001). A Python deep learning library Keras is employed for deep neural network (https://keras.io/) (Chollet, 2015). Keras is a high-level neural network application programming interface, written in Python and capable of running on top of TensorFlow, CNTK, or Theano.

In advance of learning the classification model, features are processed with principle component analysis (PCA) for each classification method to search for the optimal combination of features and discard redundant features in order to improve the performance of classification. A grid search is carried out to find the optimal combination of classification parameters.

In LR, the most basic linear classifier, parameters to be optimized with the grid search include (1) inverse of regularization strength, (2) solver for the optimization problem (all solvers provided by class: LogisticRegression are evaluated: 'saga', 'sag', 'liblinear', 'newton-cg', 'lbfgs') and (3) when fitting an estimator repeatedly on the same dataset, whether to use attributes of the existing model to initialize the new model.

In SVM, an effective and memory-efficient classifier in high dimensional spaces, a Gaussian radial basis function (RBF) is chosen and variables considered in the grid search are (1) RBF kernel parameter gamma, and (2) penalty parameter of the error term.

RF is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. Parameters of RF classifier interrogated during the grid search account for: (1) number of trees in the forest, (2) max depth of the tree, (3) number of features to consider when looking for the best split, (4) minimum number of samples required to be at a leaf node, and (5) whether bootstrap samples are used when building trees.

XGBoost is a scalable end-to-end tree boosting system, implements gradient boosting decision tree algorithm and is characterized by the excellent execution speed and model performance (Chen and Guestrin, 2016). The tree booster is used and parameters in the grid search encompass: (1) maximum depth of a tree, (2) minimum loss reduction required to make a further partition on a leaf node of the tree, (3) subsample ratio of the training instance, (4) step size shrinkage used in update to prevent overfitting, (5) subsample ratio of columns when constructing each tree, (6) L1 regularization term on weights, and (7) L2 regularization term on weights.

In DNN, the initializer (Glorot uniform, normal and uniform initializers), optimizer (RMSprop, Adam and Adagrad optimizers), and number of hidden layers were optimized, in addition to the following parameters: (1) number of epochs to train the model, (2) dropout rate, and (3) batch size. Rectified linear unit (ReLu) activation is chosen for input and hidden layers while sigmoid activation is used in the output layer. Because of the two-class classification, a binary cross-entropy loss function is used.

These classifiers and their parameters are optimized to achieve the highest accuracy of classifying coincidences in the validation set. Selected classifier and parameters are further validated by experimental imaging studies.

2.2. Detectors and measurements

In this study, the PET detectors used for measurement are TOF-PET EVA-KIT1 modules (Hamamatsu Photonics) composed of two opposing modules. Each module contains a 12×12 array of lutetium fine silicate (LFS) crystals with individual crystal dimensions of $4.14\times4.14\times20$ mm^3 [cross-sectional$\times$length], coupled to a multi-pixel photon counter (MPPC, dimension: 4×4 mm^2, 12×12 array, micropixel size: 75 μm), as shown in figure 1(a). The module has 12.9% energy resolution and 284 ps timing resolution based on our measurements (Mikhaylova et al., 2019). For these studies, the two modules were placed 10 cm apart and the source was positioned at the center of the FOV.

Three acrylic phantoms filled with very low-activity [^{18}F]FDG solution and positioned at the center of the FOV between the two TOF PET modules were used to validate the proposed background reduction method: (1) a disk (1 cm in diameter and 1.5 mm thick) with a 1.5-mm-diameter hole at the center filled with 40 Bq [^{18}F]FDG to mimic a point source; (2) a planar Derenzo phantom (2.6 mm thick) drilled with holes arranged in equilateral-triangular arrays with varying hole diameters: 1.5, 2, 2.5, 3, 3.5, 4 mm with the the center-to-center distance between holes being double the hole diameter, and filled with a solution of [^{18}F]FDG leading to an activity concentration of 100 Bq/mm^2 (hole activity / area) in each of the holes. The total activity in the phantom was 20 kBq; (3) a

planar bar phantom (2.6 mm thick) drilled with two arrays of parallel bars with widths of 2 and 4 mm with the center-to-center distance between bars being double the bar width. The activity of the [^{18}F]FDG solution in the bars was 25 Bq/mm^2. The total estimated activity in this phantom was 20 kBq. Concentrations are quoted in activity per unit area since the phantoms are planar and only a single projection image was acquired. The acquisition time for each phantom was 10 minutes. ~5.0×10^4, 2.1×10^5 and 2.0×10^5 coincidence events were acquired for the point source, planar Derenzo and bar phantom, respectively. Blank scans were repeated 5 times with each scan being 10 minutes to estimate the intrinsic background level of the TOF-PET modules and 3.1×10^4 background coincidence events were measured for a 10-minute scan on average. Therefore, the ratio between the source and background coincidences for the point source, planar Derenzo and bar phantom are ~1:1.6, 5.8:1 and 5.8:1, respectively. After classification, all images were reconstructed using 3D list mode expectation maximization (LMEM) with 20 iterations and each reconstructed image had $25\times50\times50$ voxels with a voxel size of 1 mm (25 slices of 50×50-pixel images parallel to the detector module surfaces). No random, scatter, attenuation correction, data normalization, or point-spread-function modeling were applied.

The data was saved in list mode, and the interaction positions, energy deposition and time stamps stored. As the energy deposited in crystals was recorded in analog to digital conversion (ADC) units, it was first converted to actual energy by calibration. 4 different radionuclides providing 6 energy peaks (^{22}Na for 511 and 1274 keV, ^{57}Co for 122 keV, ^{137}Cs for 662 keV, ^{176}Lu for 202 and 307 keV) were imaged for the energy calibration. The peak positions in ADC units were found by separately fitting a Gaussian model to the ADC histogram around each peak with known energy values (figure 2a) and therefore 6 energy-ADC points were obtained. A relationship between ADCs and energies was established by cubic spline extrapolation for every crystal, one of which is shown in figure 2b. With the established relationship between ADCs and energies, energies were calibrated. The distribution of calibrated energies measured using a weak ^{22}Na point source are shown in figure 2c. The peaks produced by ^{176}Lu and ^{22}Na, after energy calibration, are fit with a Gaussian and match the known emission at 307 and 511 keV.

2.3. Monte-Carlo simulation

The measurement setup with the TOF-PET modules was emulated by Monte-Carlo simulation using GATE (Version 8.0) (Jan et al., 2004). All simulation parameters were set to be consistent with actual module specifications and measurements. A ^{176}Lu source was added and distributed evenly in all crystals with an activity of 277 Bq/cm^3 (McIntosh et al., 2011). The energy spectrum of simulated ^{176}Lu background coincidences was investigated and three energy peaks were found at 88, 202 and 307 keV, which are consistent with the energies of the three main prompt gamma photons emitted during ^{176}Lu decay. The timing and energy resolutions were set to be consistent with our measurements, which are 284 ps and 12.9%, respectively (Mikhaylova et al., 2019). For both measurement and simulation, the energy window was 250-750 keV and the timing window was 3 ns. For broad use of the classification model, the source used for generating simulated training coincidence events was 100-kBq ^{18}F evenly distributed within a $5\times5\times8$-cm^3-cuboid covering the FOV of the PET modules, so that the classification model established fits for identifying sources

at different locations. Subsequently, an ^{18}F point source, with an activity varying from 10 Bq to 2000 Bq was simulated and used to estimate the reconstructed spatial resolution and investigate how much the classification method can improve the spatial resolution at different source activities. Each acquisition was simulated to last 20 hours to acquire enough counts for the spatial resolution estimation as at least 10^5 prompt counts are required by the NEMA NU4-2008 standard.

2.4. Classification performance evaluation

Classification performance was evaluated by accuracy, sensitivity and specificity. If the source coincidence was positive and the background coincidence was negative, true positive (TP) and true negative (TN) means that the source and background coincidences were correctly classified. False positive (FP) and false negative (FN) indicates source and background coincidences are misclassified. With the classifer parameters optimized, the classification accuracies [(TP+TN) / (TP+TN+FP+FN)] of each classifier were estimated using six types of test datasets that are composed of

1. additional 10^5 simulated source events and measured background events mixed with a ratio of ~5.6:1 obtained under the same setting as the training events;

2. 10^5 measured background coincidence events;

3. 10^5 simulated source events using a 10 kBq ^{18}F point source positioned at the center;

4. 10^5 simulated source events using a 10 kBq ^{18}F point source positioned 2 cm off-center axially (parallel to detector faces);

5. 10^5 simulated source events using a 10 kBq ^{18}F point source positioned 2 cm off-center transaxially (perpendicular to detector faces);

6. 10^5 simulated background coincidence events.

The datasets were chosen to evaluate the established classification models for the coincidence identification under different conditions, i.e. coincidences from background and sources located differently and obtained from simulation or measurement. The classification was repeated five times for each type of test dataset. Independent test datasets were used for each repetition of the classification.

In order to estimate the influence of background on the spatial resolution and to demonstrate whether the proposed methods can improve it, the spatial resolution was measured as the mean of the full width at half maximums (FWHM) and the full width at tenth maximum (FWTM) averaged over two orthogonal directions (vertical and horizontal) of the image of the point source in the center reconstructed plane (parallel to detector module surface) using the methods in the NEMA NU 4-2008 standard ((NEMA), 2008). For spatial resolution estimation, 50 iterations and reconstructed image dimensions of $50 \times 100 \times 100$ voxels with a voxel size of 0.5 mm were used. The Weber image contrast of the reconstructed images of the low-activity Derenzo phantom with and without the background reduction using classification were compared. The image used for analysis is the sum of the reconstructed planes that are parallel to the detector module surface. The Weber image contrast is defined

as $(I-I_b)/I_b$, where I and I_b represent the mean pixel intensity within the activity-filled holes in the phantom and the remaining background area within a round area concentric with and containing the hole array area in the phantom (encompassed by the green circle in figure 6b). The holes were segmented by registering a mask of the actual hole-array onto the reconstructed image. The peak-to-valley ratio, defined as the ratio between mean peak amplitudes and mean valley amplitudes in regions of interest (ROIs) was evaluated to characterize the contrast of the reconstructed image (the sum of reconstructed planes) of the low-activity bar phantom.

3.   Results

3.1.   Feature comparison

The spectra for each feature derived from coincidence events used for training are compared to the corresponding spectra from measured coincidences using a ~1.5 kBq ^{18}F point source positioned at the center of the FOV and 10-minute acquisition in order to make sure that the classification model learned can be applied to identify the classes of measured coincidences, as shown in figure 3. The coincidences for training which produce the feature spectra shown in figure 3 encompass simulated coincidences from a ^{18}F source and measured coincidences from the ^{176}Lu background mixed at activities consistent with the realistic activities in the measurement. Figures 3a–3d compare the spectra of the deposited energies (in both detectors), average of energies deposited, difference of energies deposited in the pair of detectors, and difference of time stamps, respectively, between training and measured datasets. The training feature spectra are in agreement with the measured spectra, indicating the classification model learned using partially simulated training data can classify measured coincidence events.

The whole training set consists of 1.6×10^5 measured background coincidences (50-minute acquisition) and 8.6×10^5 simulated source coincidences that were obtained using the 100-kBq ^{18}F $5 \times 5 \times 8$-cm^3-cuboid source. The feature spectra for simulated source and measured background coincidences are compared in figure 4. The source and background spectra are significantly different, which supports the hypothesis that the two classes are differentiable.

3.2   Classifier optimization and classification accuracy

A grid search was implemented on each classifier to find the optimal classifier and parameter combination. After the grid search, five features remained based on PCA analysis and classifiers were set as follows (parameters not described here use default values):

1.   LR uses inverse of regularization strength of 1.0, stochastic average gradient (sag) descent solver, and reuses the solution of the previous call to fit as initialization.

2.   SVM uses penalty parameter of the error term of 2,550,000 and RBF kernel coefficient gamma of 0.2.

3.   RF uses 90 trees in the forest, maximum depth of the tree of 40, minimum number of samples 4 required to be at a leaf node, (0.12 × number of features)

features are considered when looking for the best split, and bootstrap samples are used when building trees.

4. XGBoost uses tree booster, maximum depth of a tree of 17, minimum loss reduction of 1.2 required to make a further partition on a leaf node of the tree, 84 boosted trees to fit, boosting learning rate of 0.87, subsample ratio of columns of 0.02 when constructing each tree, L2 regularization term on weights of 0.31 and L1 regularization term on weights of 0.49.

5. DNN uses Keras sequential model with RMSProp optimizer, normal initializer, binary cross-entropy loss function, 10 hidden layers, 100 epochs, dropout rate of 0.1, and a batch size of 2000.

The classification accuracies of each classifier were estimated using six types of test datasets explained in Section 2.4, and are given in Table 1. Table 1 shows that XGBoost performs the best of the classifiers investigated. It is able to select background coincidences at an accuracy of more than 99%. As test datasets II to VI do not include both positive coincidences (i.e. source coincidences) and negative coincidences (i.e. background coincidences), the sensitivity [TP/(TP+FN)] or the specificity [TN/(TN+FP)] cannot be determined, therefore only the classification accuracies of these five datasets are shown in the table. The percentage and the accuracy trend of different classifiers for measured test datasets are similar to simulated test datasets, indicating again the classification model works for both the measurement and the simulation. The difference in accuracy of simulated and measured test datasets (e.g. datasets II vs VI) results from the small discrepancies between the simulated and measured feature spectra. The accuracies for the point source positioned at different locations in the FOV are very similar, demonstrating that the classification models are independent of the source position.

3.3. Image quality improvement

In this section, the best classifier, XGBoost, with its optimal parameters are applied to imaging an ^{18}F point-like source, 2-D Derenzo phantom, and a planar axial hot bar phantom. Measured coincidences are classified in order to remove the background and enhance the reconstructed image.

Figure 5 compares the reconstructed images (the sum of reconstructed planes) of a measured 40 Bq ^{18}F point-like source without (figure 5a) and with (figure 5c) background reduction using the XGBoost classifier. As a comparison, the reconstructed images of a simulated 40 Bq ^{18}F ideal point source preprocessed without (figure 5b) and with (figure 5d) classification are also presented. For both simulated and measured point sources, the background is significantly reduced by classification. Then, reconstructed spatial resolution in terms of FWHM and FWTM were evaluated at different ^{18}F point source activities (10-2000 Bq) using simulation, as shown in figure 5e (FWHM) and 5f (FWTM). The spatial resolution is degraded because of the relatively high background at lowest source activities. The spatial resolution loss due to background events can be removed using classification leading to spatial resolution that is independent of activity.

The reconstructed images of the low-activity Derenzo phantom with and without the background reduction using classification are compared in figures 6a and 6b, and the corresponding Weber image contrasts for the area encompassed by green circle in figure 6b, respectively were estimated. Classification enhanced the image contrast of the Derenzo phantom by 89.4% (image contrast of 1.25 for XGBoost, compared to 0.66 for no classification applied).

Figure 7 shows the reconstructed image of a low-activity planar bar phantom with (figure 7a) and without (figure 7b) background reduction using classification. The pixel intensities averaged over columns of the bars with 4-mm center-to-center distance in the area enclosed by the green box and the bars with 2-mm center-to-center distance in the area enclosed by the blue box in figure 7b are plotted in figures 7c and 7d respectively. The valleys in the image to which classification was applied (figure 7a) are deeper than valleys in the image without classification (figure 7b), demonstrating the higher contrast achieved by removing background events for low-activity scans. The peak-to-valley ratio (ratio between mean peak height and mean valley height) was increased by 52.4% for the 4-mm bars and 14.6% for the 2-mm bars using the XGBoost classifier. The comparison of the effect of five different classifiers on the intrinsic background reduction for planar bar phantom imaging is shown in section S1 of the supplementary data.

4. Discussion

With the goal of reducing the intrinsic background of Lu-based PET scanners and ultimately improving the image quality in low-activity applications, a preprocessing approach based on classification algorithms was proposed. Each measured coincidence was classified as either a source or background event using a classification model trained using a combination of measured background coincidences and simulated source coincidences. Several commonly-used classifiers were compared to determine the best choice for this application. We found that XGBoost performed most effectively in classifying events and reducing the intrinsic background. This classifier was further validated using measurements on several phantoms where clear improvements in image resolution and contrast were observed. This strategy is particularly important for very low activity imaging, for example in tracking small numbers of radiolabeled cells, where the background cannot be reduced by simply subtracting an averaged background counts from the very limited number of counts recorded.

The proposed method should work not only for the pair of TOF-PET module detectors studied here, but also for complete PET scanners. To use this method, a classification model needs to be trained and validated for each scanner. Ideally, accurate simulations for the scanner also are available to aid in generating training data in which source and background events are known. Once the classification model is established, the coincidence identification process is fast, so the proposed method can be implemented either online or offline.

As each classifier has a number of tunable parameters, we cannot optimize them exhaustively considering the computation time that would be required. Only principal parameters were therefore optimized and default values are used for other parameters. The optimal classifier is not necessarily the same for different scanners and a classifier

optimization process is suggested for each individual scanner. Based on this work, XGBoost and DNN would be the top two classifiers recommended for first evaluation. XGBoost delivers high performance and accuracy as compared to other algorithms as it has in-built L1 (Lasso Regression) and L2 (Ridge Regression) regularization which prevents the classification model from overfitting, capability to handle missing values and effective tree pruning method to prevent the node splitting stop when a negative loss is encountered in the gradient boosting algorithm. DNN is well-known for outperforming nearly every other machine learning algorithm when the data volume grows. These are the potential reasons that XGBoost and DNN classifiers perform better in the task of this study.

This study investigated five commonly used classifiers with the purpose of validating the concept of using the features of coincidence events to differentiate source from background coincidences, and demonstrated that XGBoost and DNN classifiers worked well for the coincidence differentiation for the TOF-PET modules. Although there exist a variety of other powerful and well-established classifiers, it is hard to test all of them in this proof-of-concept study. For other PET scanners and detectors, the two classifiers performing best in this study are recommended for study along with trials of other classifiers.

Coincidences are divided into just two classes (source and background) in this study. In principle, further classifications are possible, for example classifying events in to true, random, scattered and background coincidences. In this case, the proposed method could also contribute to random and scatter corrections, where current correction methods may not be trivial to implement at very low activity levels. However, in this case, the features need to be selected carefully to allow the differentiation of random and scatter coincidences and avoid high rates of misclassification as they have smaller feature differences compared to the differences between true and background coincidences. Five features, both energies deposited in the paired detectors, which were taken as two individual features, average and differential of the two energies, and time difference of the coincidence events are extracted in this study, however a different set of features could be explored for different scanners, especially for scanners that incorporate additional information, for example depth of interaction (DOI).

The proposed method not only is suitable for very-low activity applications (e.g. in pediatric and possibly fetal imaging), but can also extend the imaging time span in kinetic studies, facilitating longer tracking of biological and metabolic processes. Also, in cell trafficking studies which are of interest in a variety of research and clinical fields, such as immune therapy, single or small numbers of cells may be loaded into the circulatory system with very low activities (<10 Bq/cell) and their spatio-temporal variation cannot currently be captured in the presence of background using Lu-based PET scanners. The proposed method can potentially facilitate the use of Lu-based PET scanners in such extreme low activity biological applications.

However, individual training procedures will likely be needed for each different PET scanner and different phantom/object in the FOV to guarantee the classification accuracy. For best results, and to extend the range for imaging extremely low activity sources, these methods should be combined with other methods to minimize the background relative to source

events, measure and model the background, and incorporate as much known information as possible into the reconstruction algorithm.

5. Conclusion

A classification methodology is proposed to identify and remove intrinsic background events in Lu-based PET detectors and scanners in order to facilitate ultra-low activity imaging that has a range of potential applications. By interrogating individual coincidence events, the origin (source or background) of each measured coincidence event is predicted using the classification model established by supervised learning of the training set composed of measured background coincidences and simulated source coincidences. A platform integrating five commonly-used classifiers was constructed, which allowed the selection of the optimal classifier and parameter combination to reach the highest classification accuracy. Based on this platform, XGBoost was demonstrated to be the most effective classifier in background reduction. The proposed methodology is not limited to the TOF-PET modules used in this study but can be generalized for any PET detectors and scanners using materials with intrinsic background radiation. This method has the potential to reduce the lower signal detection limits for PET scanners and drive the development of ultra-low activity imaging techniques for a range of interesting applications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

NEMA 2008 Performance Measurements of Small Animal Positron Emission Tomographs. In: NEMA Stardards Publication NU 4-2008, Rosslyn, VA

Alva-Sánchez H, Zepeda-Barrios A, Díaz-Martínez V, Murrieta-Rodríguez T, Martínez-Dávalos A and Rodríguez-Villafuerte M 2018 Understanding the intrinsic radioactivity energy spectrum from 176 Lu in LYSO/LSO scintillation crystals Scientific Reports 8 17310 [PubMed: 30470826]

Budinger TF 1983 Time-of-flight positron emission tomography: status relative to conventional PET Journal of Nuclear Medicine 24 73–8 [PubMed: 6336778]

Carlier T, Willowson KP, Fourkal E, Bailey DL, Doss M and Conti M 2015 (90)Y -PET imaging: Exploring limitations and accuracy under conditions of low counts and high random fraction Medical Physics 42 4295–309 [PubMed: 26133627]

Chen T and Guestrin C 2016 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, ACM Press pp 785–94

Cherry SR, Badawi RD, Karp JS, Moses WW, Price P and Jones T 2017 Total-body imaging: Transforming the role of positron emission tomography Science Translational Medicine 9

Cherry SR and Dahlbom M 2006 PET: Molecular Imaging and Its Biological Applications (New York: Springer New York) pp 1–124 Chollet F 2015 Keras. https://github.com/fchollet/keras

Conti M and Bendriem B 2019 The new opportunities for high time resolution clinical TOF PET Clinical and Translational Imaging 7 139–47

Conti M, Eriksson L, Rothfuss H, Sjoeholm T, Townsend D, Rosenqvist G and Carlier T 2017 Characterization of 176Lu background in LSO-based PET scanners Physics in Medicine and Biology 62 3700 [PubMed: 28333041]

Du J, Bai X, Gola A, Acerbi F, Ferri A, Piemonte C, Yang Y and Cherry SR 2018 Performance of a high-resolution depth-encoding PET detector module using linearly-graded SiPM arrays Physics in Medicine and Biology 63 035035 [PubMed: 29324437]

Friedman JH 2001 Greedy function approximation: a gradient boosting machine Annals of Statistics 1189–232

Gambhir SS, Barrio JR, Phelps ME, Iyer M, Namavari M, Satyamurthy N, Wu L, Green LA, Bauer E and MacLaren DC 1999 Imaging adenoviral-directed reporter gene expression in living animals with positron emission tomography Proceedings of the National Academy of Sciences 96 2333–8

Gjelsteen AC, Ching BH, Meyermann MW, Prager DA, Murphy TF, Berkey BD and Mitchell LA 2008 CT, MRI, PET, PET/CT, and ultrasound in the evaluation of obstetric and gynecologic patients Surgical Clinics of North America 88 361–90 [PubMed: 18381118]

Jan S, Santin G, Strul D, Staelens S, Assie K, Autret D, Avner S, Barbier R, Bardies M and Bloomfield P 2004 GATE: a simulation toolkit for PET and SPECT Physics in Medicine and Biology 49 4543–61 [PubMed: 15552416]

McIntosh B, Stout DB and Goertzen AL 2011 Validation of a GATE model of 176Lu intrinsic radioactivity in LSO PET systems IEEE Transactions on Nuclear Science 58 682–6

Mikhaylova E, Brooks J, Zuro DM, Nouizi F, Kujawski M, Madabushi SS, Qi J, Zhang M, Chea J and Poku EK 2019 Prototype small-animal PET-CT imaging system for image-guided radiation therapy IEEE Access 7 143207–16 [PubMed: 32435548]

Ouyang Y, Kim TJ and Pratx G 2016 Evaluation of a BGO-based PET system for single-cell tracking performance by simulation and phantom studies Molecular Imaging 15 1536012116646489 [PubMed: 27175009]

Pawelke J, Enghardt W, Haberer T, Hasch B, Hinz R, Kramer M, Lauckner E and Sobiella M 1997 In-beam PET imaging for the control of heavy-ion tumour therapy IEEE Transactions on Nuclear Science 44 1492–8

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R and Dubourg V 2011 Scikit-learn: Machine learning in Python Journal of Machine Learning Research 12 2825–30

Schäfer JF, Gatidis S, Schmidt H, Gückel B, Bezrukov I, Pfannenberg CA, Reimold M, Ebinger M, Fuchs J and Claussen CD 2014 Simultaneous whole-body PET/MR imaging in comparison to PET/CT in pediatric oncology: initial results Radiology 273 220–31 [PubMed: 24877983]

Strydhorst J, Carlier T, Dieudonné A, Conti M and Buvat I 2016 A gate evaluation of the sources of error in quantitative (90)Y PET Medical Physics 43 5320

Thunemann M, Schörg BF, Feil S, Lin Y, Voelkl J, Golla M, Vachaviolos A, Kohlhofer U, Quintanilla-Martinez L and Olbrich M 2017 Cre/lox-assisted non-invasive in vivo tracking of specific cell populations by positron emission tomography Nature Communications 8 444

Yao R, Ma T and Shao Y 2008 Lutetium oxyorthosilicate (LSO) intrinsic activity correction and minimal detectable target activity study for SPECT imaging with a LSO-based animal PET scanner Physics in Medicine and Biology 53 4399–415 [PubMed: 18670052]

Yoshida E, Tashima H, Nishikido F, Murayama H and Yamaya T 2014 Reduction method for intrinsic random coincidence events from 176Lu in low activity PET imaging Radiological Physics and Technology 7 235–45 [PubMed: 24496884]
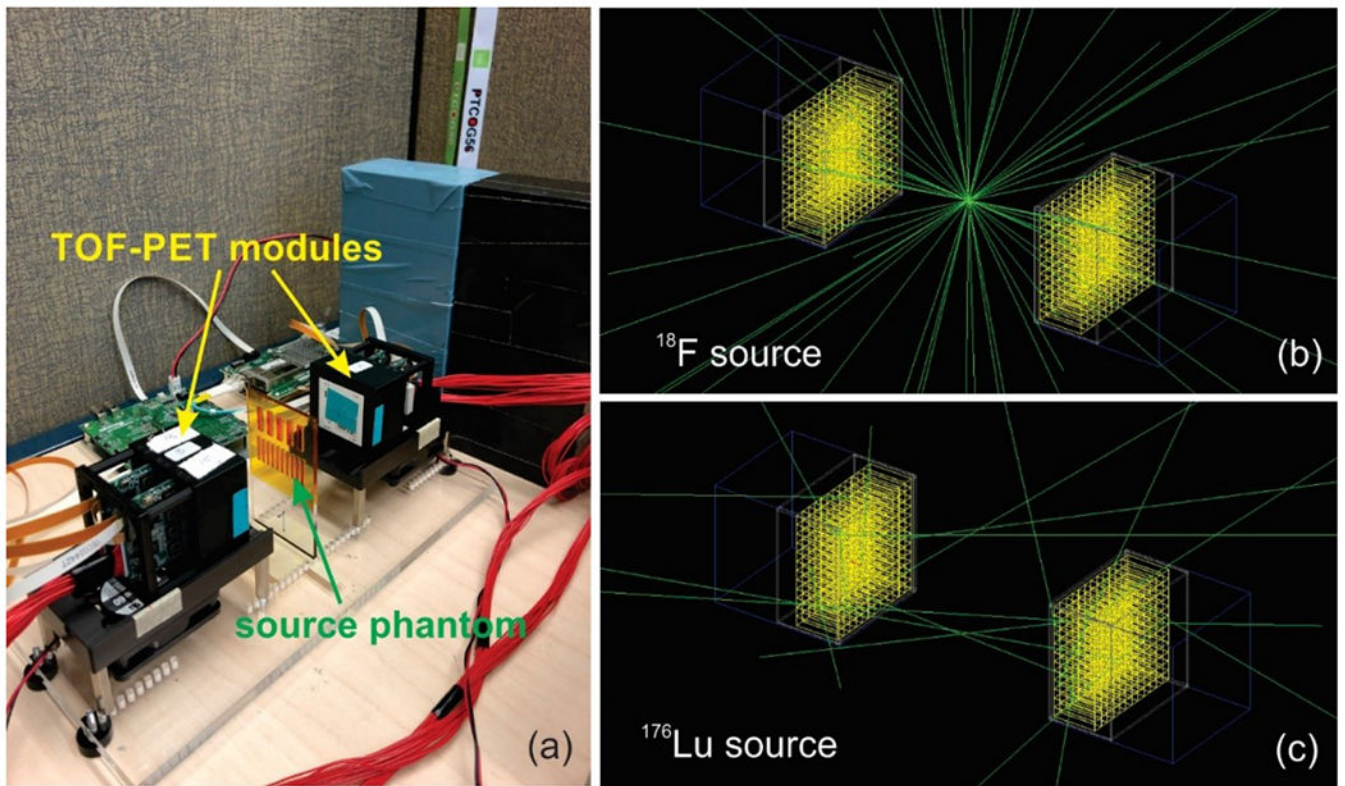
Figure 1.
(a) TOF-PET modules and planar bar phantom; (b) GATE simulation of ^{18}F point source detection with TOF-PET modules; (c) GATE simulation of ^{176}Lu intrinsic radiation detection with TOF-PET modules.
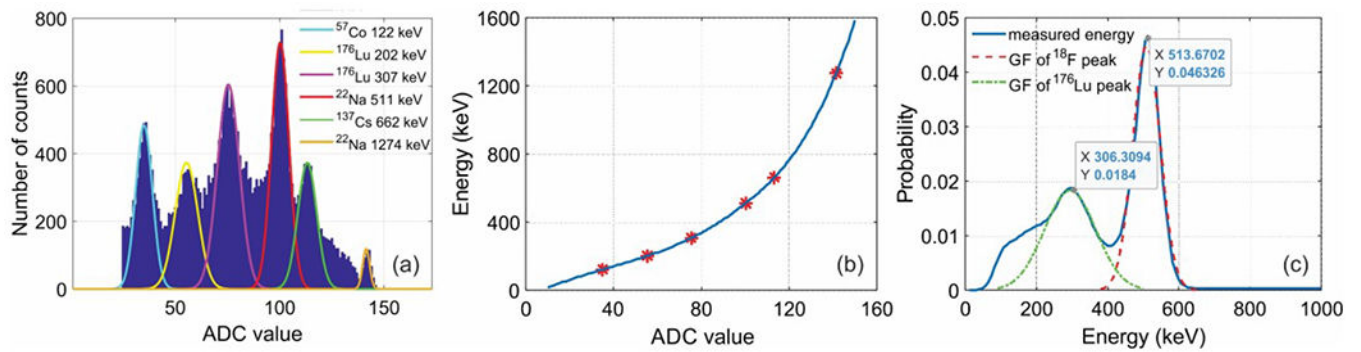
Figure 2.
Energy calibration procedures. (a) Energy peak localization by Gaussian fitting to 6 individual peaks generated by ^{22}Na, ^{57}Co, ^{137}Cs and ^{176}Lu sources in the ADC histogram; (b) Establishing an energy-ADC relationship for energy calibration. The blue curve is the cubic spline extrapolation of the energy-ADC points obtained in figure (a); (c) Validation of calibrated energies by checking the position of peaks produced by ^{176}Lu and ^{22}Na sources. Gaussian fits (GF) are applied to the calibrated peaks produced by ^{176}Lu (green curve) and ^{22}Na (red curve) sources.
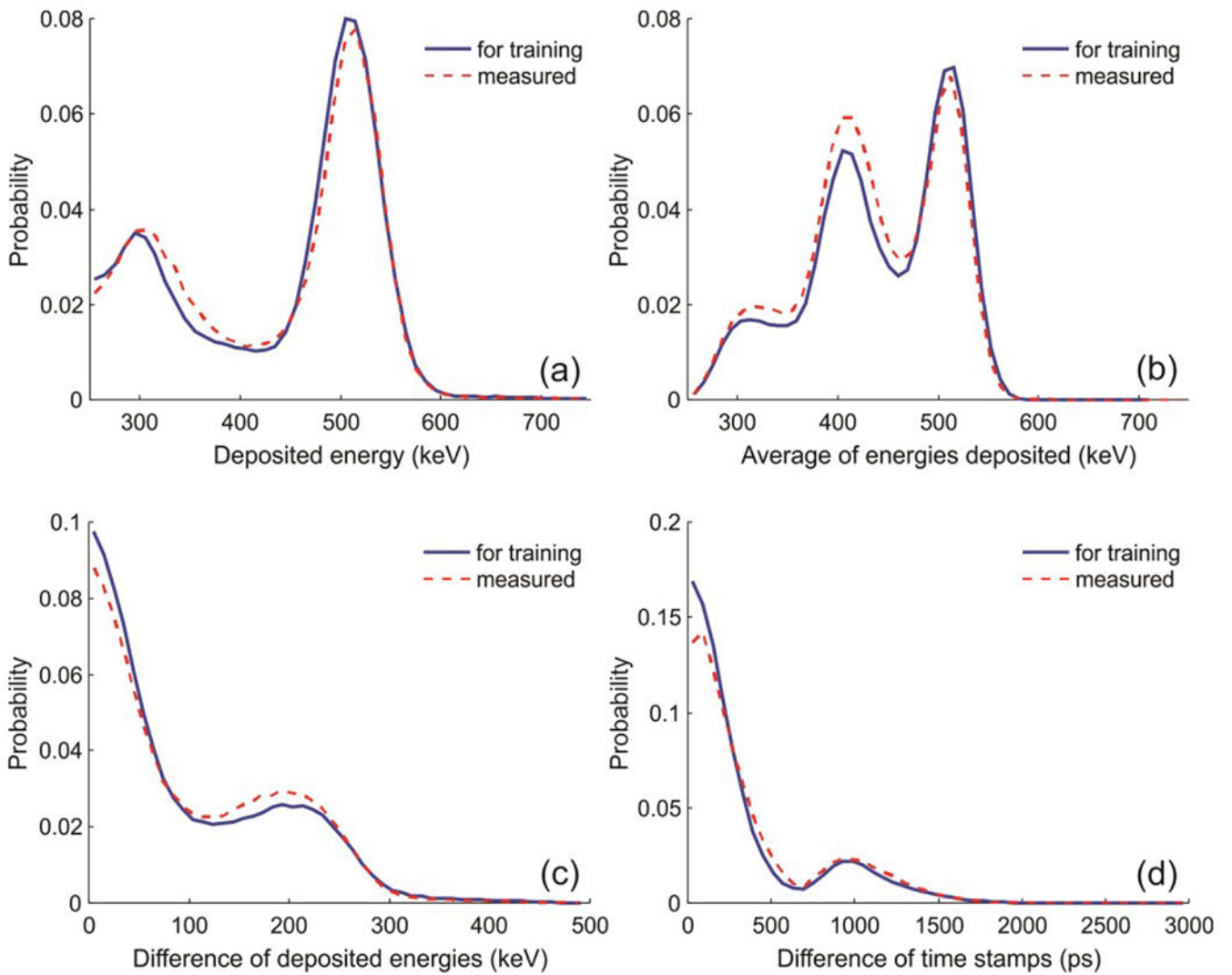
Figure 3.
Comparison between the spectra of features used for training the classification model and measured features. Figures (a) to (d) are the spectra of deposited energies, average of energies deposited, difference of energies deposited in the pair of detectors, and time difference of all coincidences, respectively
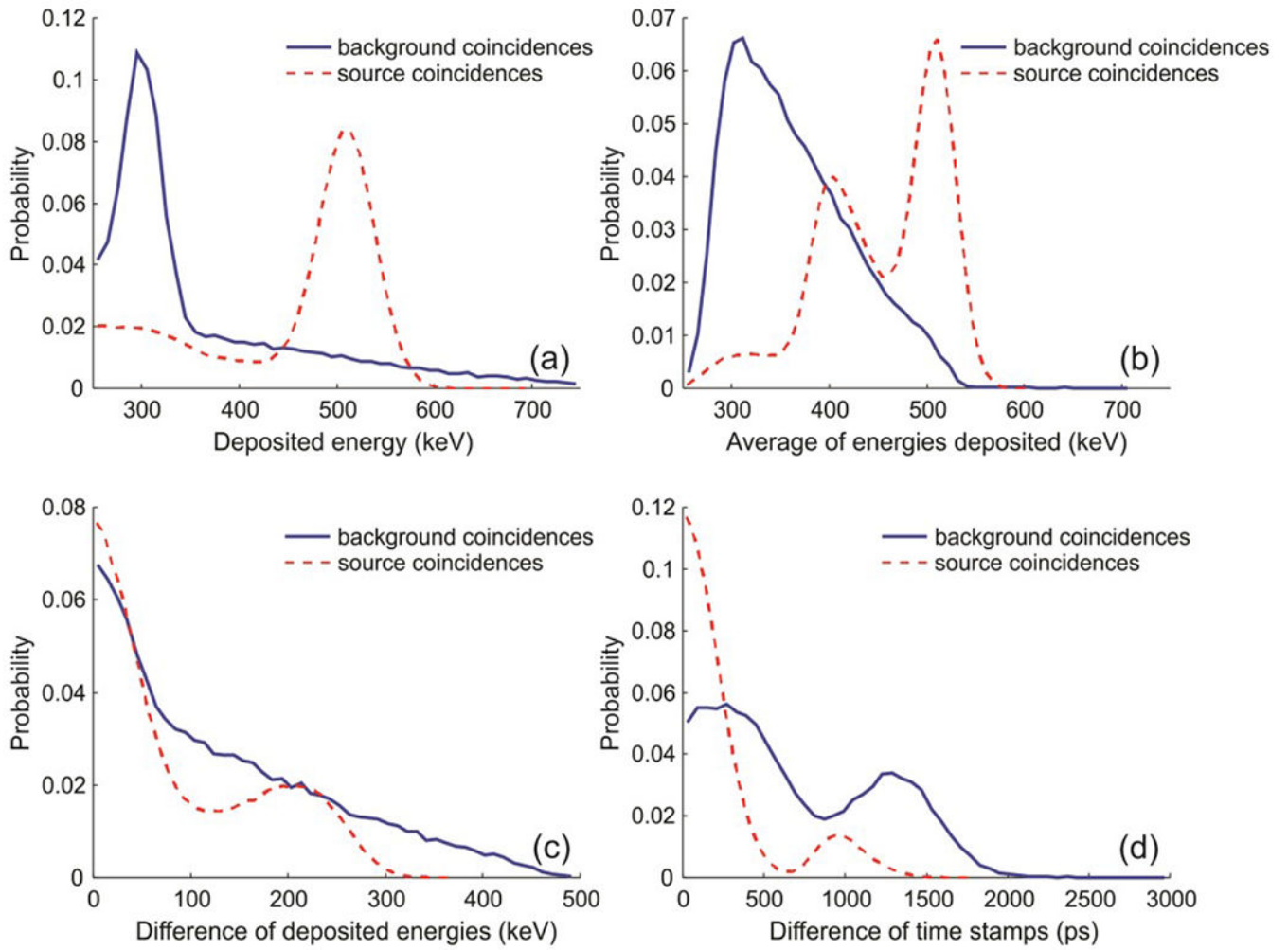
Figure 4.
Comparison of feature spectra between background and source coincidences. Figures a-d represent the spectra of deposited energies, average of energies, difference of energies deposited in the pair of detectors, and difference of time stamps, respectively.
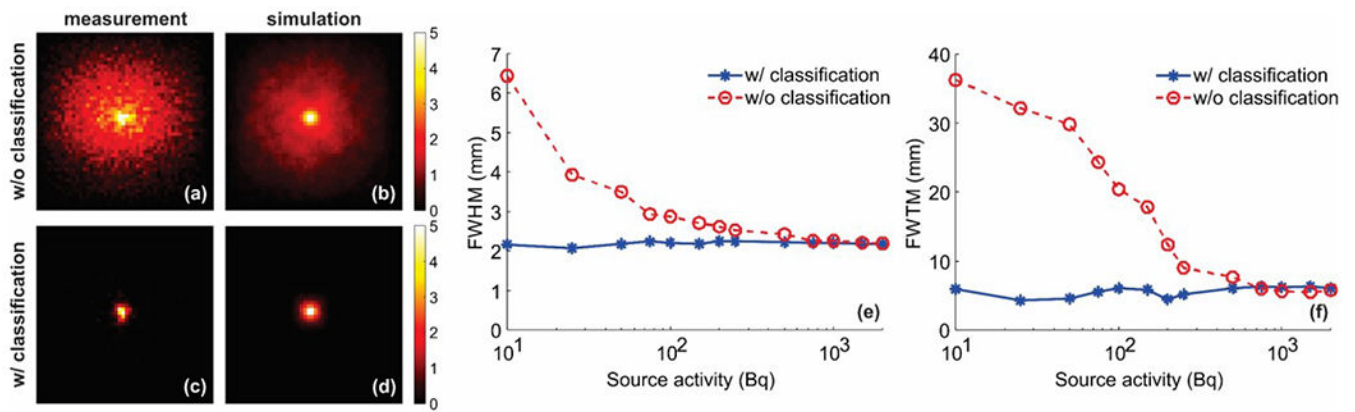
Figure 5.

(a) & (c) Reconstructed images of a measured 40 Bq ^{18}F point-like source preprocessed without and with XGBoost classifier, respectively. (b) & (d) Reconstructed images of a simulated 40 Bq ^{18}F point source preprocessed without and with XGBoost classifier, respectively. (e) & (f) FWHM and FWTM as a function of ^{18}F point source activity (10-2000 Bq) with classification applied (blue stars) and not applied (red circles).
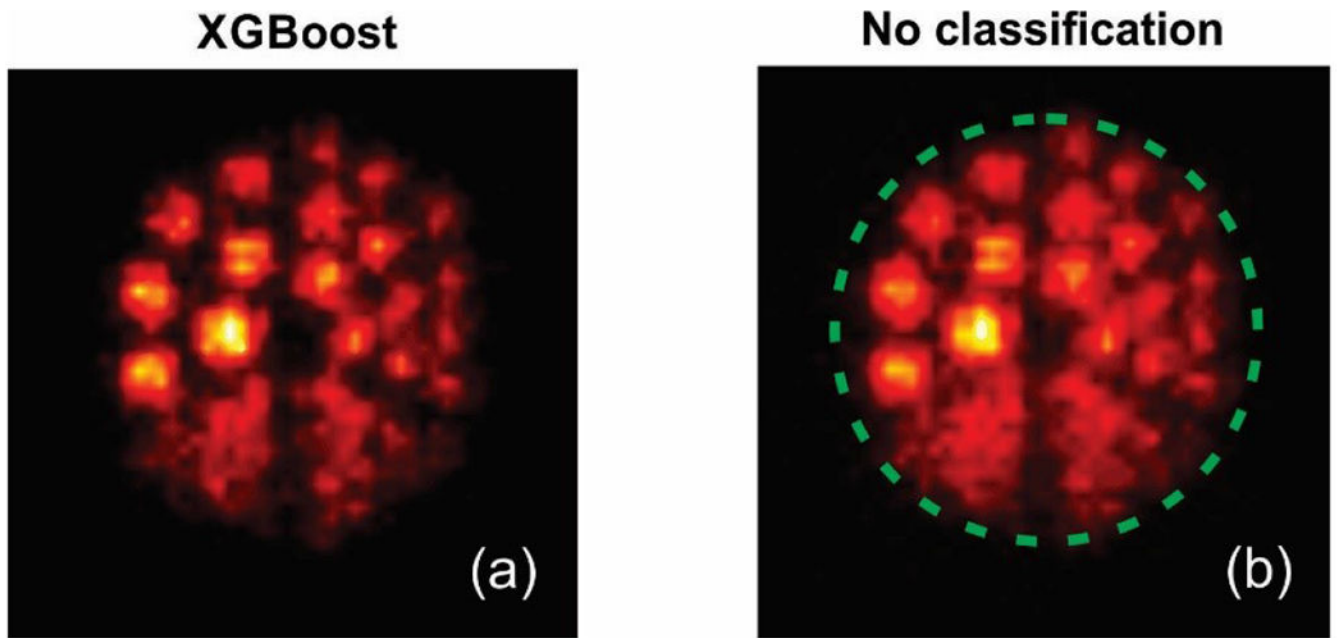
Figure 6.
Reconstructed images of the low-activity planar Derenzo phantom with (a) and without (b) classification applied. The green circle in (b) encloses the hole-array area.
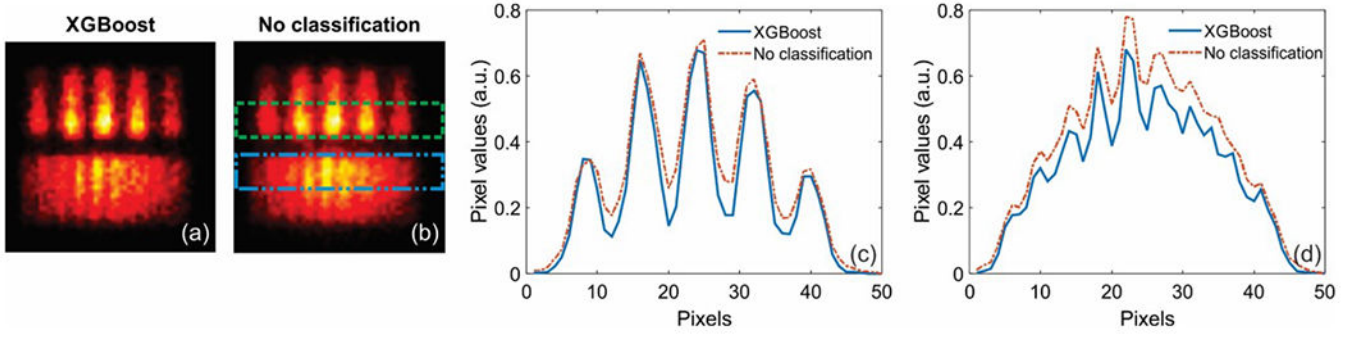
Figure 7.
Reconstructed images of a low-activity planar bar phantom with (a) and without (b) classification applied. (c) Pixel value plot averaged over the columns of the bars with 4-mm center-to-center distance in the green box in (b) with and without XGBoost classification applied. (d) Pixel value plot averaged over the columns of the bars with 2-mm center-to-center distance in the blue box in (b) with and without XGBoost classification applied.

Table 1.

Accuracies of different classification methods for different test datasets

Classifier	Accuracy (%) (test dataset I)	Sensitivity (%) (test dataset I)	Specificity (%) (test dataset I)	Accuracy (%) (test dataset II)
LR	87.93±0.021	95.54±0.038	45.74±0.054	45.45±0.50
SVM	91.85±0.014	95.66±0.053	70.68±0.25	70.65±0.84
RF	92.94±0.035	96.09±0.048	75.46±0.30	81.91±0.60
XGBoost	99.67±0.11	99.76±0.069	99.38±0.17	99.66±0.034
DNN	91.70±0.51	93.57±1.62	81.32±6.46	81.61±0.76
Classifier	Accuracy (%) (test dataset III)	Accuracy (%) (test dataset IV)	Accuracy (%) (test dataset V)	Accuracy (%) (test dataset VI)
LR	86.76±0.069	88.24±0.048	88.78±0.015	52.87±0.13
SVM	88.08±0.086	79.88±0.13	89.31±0.021	76.26±0.035
RF	92.13±0.080	91.32±0.065	94.15±0.060	73.05±0.11
XGBoost	97.66±0.049	97.51±0.017	98.89±0.23	86.48±0.035
DNN	94.73±0.083	95.11±0.053	95.70±0.057	72.41±0.19