MACHINE LEARNING
Science and Technology

PAPER

CrossMark

Pulse shape discrimination and exploration of scintillation signals using convolutional neural networks

J Griffiths[2], S Kleinegesse[3], D Saunders, R Taylor and A Vacheret [ORCID]

1 Imperial College London, Department of Physics, London, SW7 2AZ, United Kingdom
2 Now at: University of Cambridge, Cavendish Laboratory, Cambridge, CB3 0HE, United Kingdom
3 Now at: University of Edinburgh, School of Informatics, Edinburgh, EH8 9AB, United Kingdom

E-mail: antonin.vacheret@imperial.ac.uk

Abstract

We demonstrate the use of a convolutional neural network to perform neutron-gamma pulse shape discrimination, where the only inputs to the network are the raw digitised silicon photomultiplier signals from a dual scintillator detector element made of ^{6}Li F:ZnS(Ag) scintillator and PVT plastic. A realistic labelled dataset was created to train the network by exposing the detector to an AmBe source, and a data-driven method utilising a separate photomultiplier tube was used to assign labels to the recorded signals. This approach is compared to the charge integration and continuous wavelet transform methods and a simpler artificial neural net. It is found to provide superior levels of discrimination, achieving an area under the curve of 0.996 ± 0.003. We find that the neural network is capable of extracting interpretable features directly from the raw data. In addition, by visualising the high-dimensional representations of the network with the t-SNE algorithm, we discover that not only is this method robust to minor mislabeling of the training dataset but that it is possible to identify an underlying substructure within the signals that goes beyond the original labelling. This technique could be utilised to explore and cluster complex, raw detector data in a novel way that may reveal more insights than standard analysis methods.
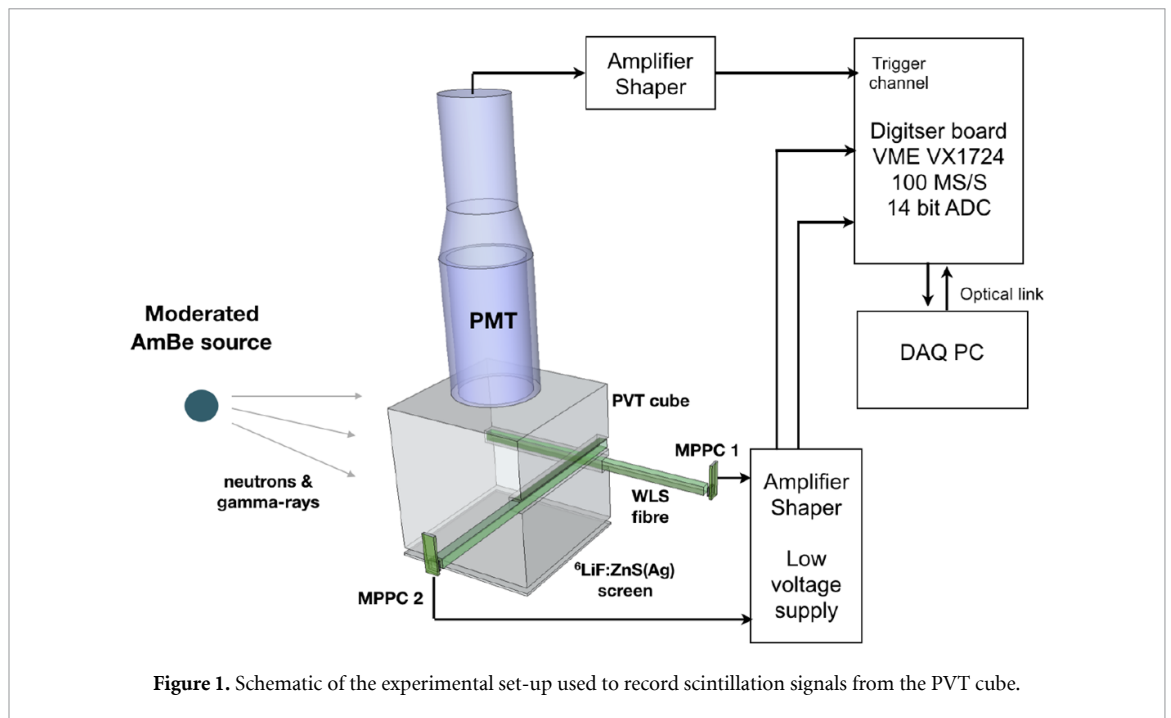
1. Introduction

Efficient pulse shape discrimination (PSD) techniques are required in many applications, from radiation measurements to particle physics experiments. PSD is a typical classification problem whereby digitised waveforms can be discriminated based on their time and energy features. Such features or characteristics are usually dictated by the underlying physical process that occurs in the detection medium. For example, in scintillation signals, a common characteristic used for PSD is the decay time of the detected scintillation light, which is dictated by the atomic or molecular structure of radiative states.

In this paper we focus on one specific application of PSD using ^{6}Li F:ZnS(Ag) phosphor screens similar to those used in the SoLid reactor neutrino experiment [1] and corresponding to energy deposited in the range of 100 keV to tens of MeV. The results obtained here are also relevant to other neutron detection applications where high discrimination is required. The neutron capture reaction on ^{6}Li produces highly ionising particles that excite the ZnS(Ag) scintillator:

$$n + {}^6\mathrm{Li} \rightarrow \alpha + {}^3_1\mathrm{H} \ (4.8\ \mathrm{MeV}) \tag{1}$$
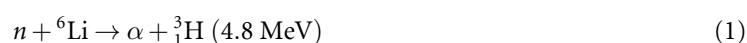
Thin sheets of ^{6}Li F:ZnS(Ag) provide high neutron detection efficiency due to the large ^{6}Li neutron capture cross section and the high scintillation yield of ZnS(Ag). Strong gamma-ray background rejection can be achieved as a result of the large difference between fast and slow ZnS(Ag) scintillation components for electron and nuclear interactions, respectively. As our focus here is the classification of those two type of signals, we make the distinction between these scintillation responses by defining as *electron scintillation* (ES) signals, the interactions from gamma-rays and other charged particles (electron, positrons, muons, etc) and *nuclear scintillation* (NS) signals as those produced by nuclei such as protons, alpha particles or heavier ions.

Figure 1. Schematic of the experimental set-up used to record scintillation signals from the PVT cube.

The ^{6}Li neutron capture reaction that produces a tritium and alpha therefore produces a distinctive NS signal in this detector medium.

Traditional PSD techniques that utilise time domain information such as charge-integration (CI) [2] are popular for their robustness and reasonable performance [3]. Other methods based on frequency information [4] can achieve superior performance by exploiting better the information available, but are more computationally expensive than simple integration. The performance of these approaches is dependent on a number of factors that include the choice of scintillator, experimental requirements and read out technique [3] (and references therein). They also tend to have limitations with real data that exhibit a number of additional features such as pile up or background interactions [5].

Neural networks have the potential to be superior classifiers provided they are trained with realistic data. In particular, convolutional neural networks (CNN), are well suited to raw data that have high local correlations such as waveform signals. They are very successful in computational vision tasks and can reach high performance with limited datasets. Once trained, obtaining a classification output from a CNN is also very fast.

In this work, we present the results of using a CNN to discriminate between ES and NS signals using the raw scintillation pulse from a single PVT cube with ^{6}Li F:ZnS(Ag). First, we describe the experimental set up that was used to collect a labelled dataset for the training of the network. We then present the details of the CNN architecture, an artificial neural net (ANN) as well as two other common PSD algorithms, in order to provide a set of benchmarks that are compared to the results of the CNN. Finally we investigate and discuss the features learned by the network, demonstrating that they are interpretable and can be utilised to identify a clear substructure within the ES and NS signals.

2. Experimental data set

In this work, supervised learning was used to train the CNN and therefore a labelled dataset of ES and NS signals was required. Instead of using a set of Monte-Carlo generated signals which may contain approximations or unrealistic features, a dataset consisting of real scintillation signals was collected. A schematic of the experimental setup used to collect this data is shown on figure 1. The detector element consists of a 250 um thick, neutron sensitive ^{6}Li F:ZnS(Ag) screen (Scintacor ND) coupled to a 5 cm × 5 cm × 5 cm polyvinyl-toluene (EJ-200) scintillating cube, which is only sensitive to ES signals. The detector element was exposed to an AmBe source that emits neutrons and gamma-rays over a wide range of energies (MeV to 10 MeVs for neutron and MeV gamma-rays), and therefore provides a realistic range of signals for the training set. A few per cent of the dataset includes signals from background interactions such as muons and natural radioactivity and the reasonable activity of the source also results in a fraction of pile up pulses. Scintillation light from both scintillators was collected by two Saint-Gobain single cladded, 3 mm × 3 mm square section BCF-91A wavelength shifting fibres placed in grooves on the side of the cube. The collected
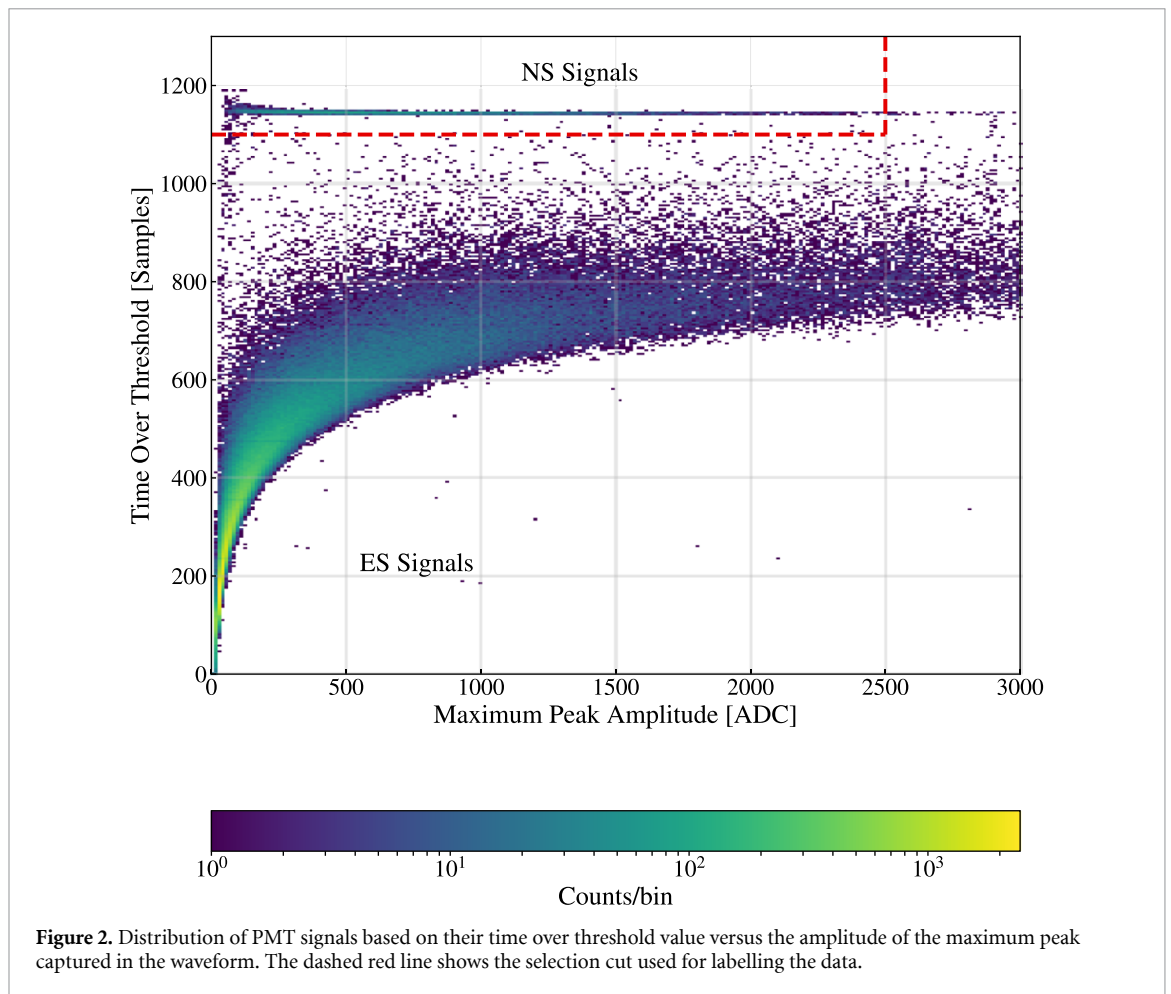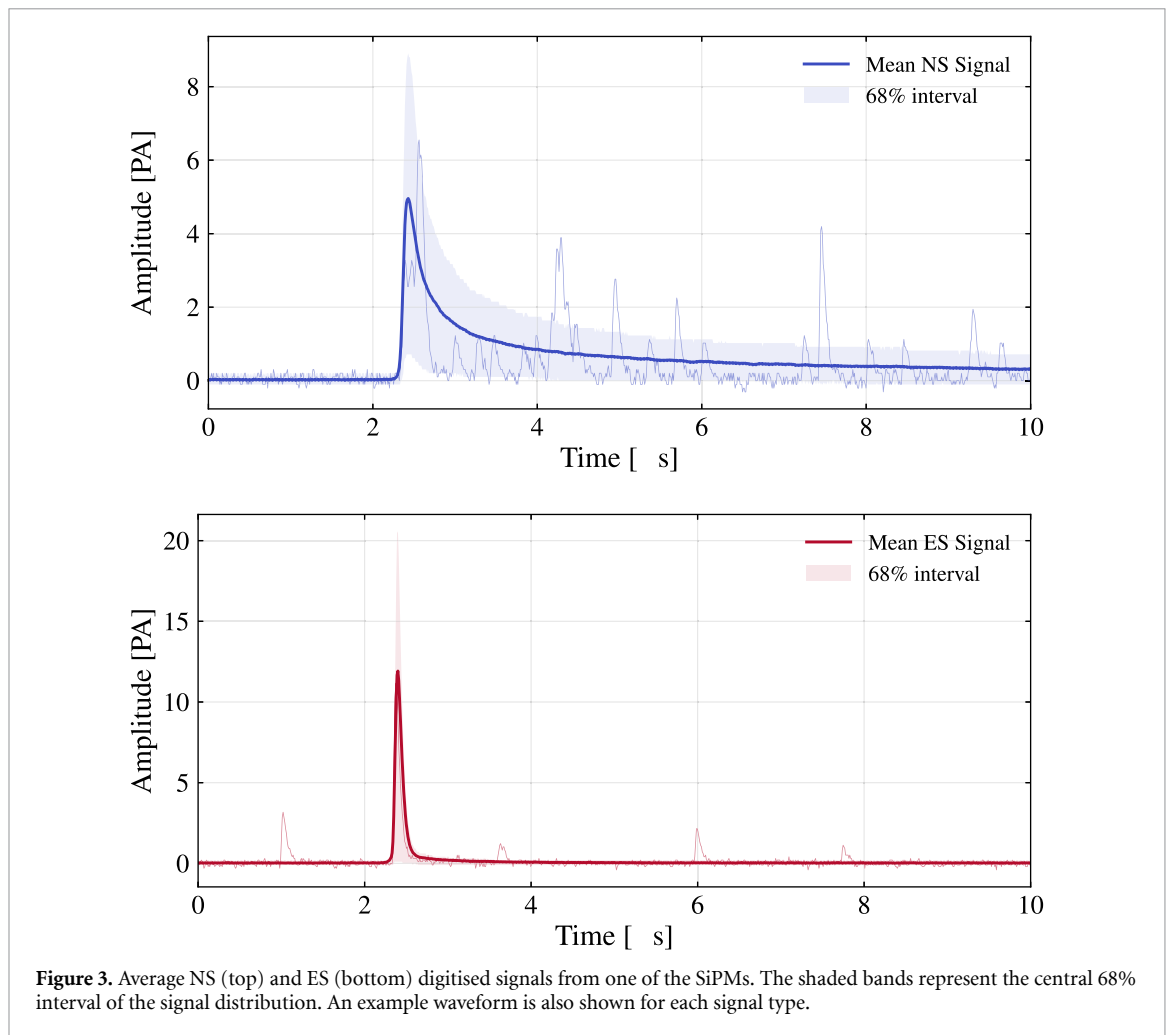
Figure 2. Distribution of PMT signals based on their time over threshold value versus the amplitude of the maximum peak captured in the waveform. The dashed red line shows the selection cut used for labelling the data.

light was registered by two Hamamatsu MPPC S12 572-050P silicon photomultipliers (SiPMs) attached to one end of each fibre. To enhance the reflection of light inside the cube and collection of that light by the sensor, the cube surface had a rough finish and was wrapped in Tyvek and optical grease was applied to both SiPM and PMT windows. Approximately one hundred photons per MeV is detected by each SiPM [6] with amplitude of signals for neutrons ranging from single photo-avalanche (PA) to tens of PA.

A Philips XP1911/UVPA photomultiplier tube (PMT) was placed on top of a hole cut into the Tyvek, and used to detect a large fraction of the blue scintillation light not collected by the fibres. The greater collection of emitted light in the PMT was intended to provide clear discrimination of NS and ES signal so that labels can be confidently assigned to the corresponding signals seen by the SiPMs. The data from both the PMT and SiPMs was digitised using a CAEN VX1724 digitiser card at 100 MS s^{-1} with a 14 bit range, and the larger PMT response was used to trigger the acquisition of the SiPMs signals. To ensure that the majority of the decay time of the ZnS scintillation component was acquired, the maximum length of each waveform aquisition was set to be 10 microseconds (or 1000 time samples) providing at the same time, good resolution on the few nanoseconds-wide ES signals.

Labels were then assigned to the NS and ES signals using a simple PSD method based on the time-over-threshold (TOT) and maximum peak amplitude (MPA) of the PMT signals, which provides very good results for discriminating events at a low threshold. This technique is also meant to replicate a realistic hardware trigger like the one used in SoLid [7]. In a real life system, this trigger is designed for high efficiency at the expense of contamination from various ES signals like the ones caused by muons. Offline PSD techniques discussed in the next sections provide the additional purity required for the application. Figure 2 shows the PSD parameter value as a function of signal amplitude for each of the PMT waveforms. As a result of the long shaping of the PMT pulse, NS events have a much longer decay time and are above threshold for the majority of the acquired time period. Signals with TOT $\geq$ 1100 samples and MPA $\leq$ 2500 ADC are labelled as NS and those outside of this range are labelled as ES. The identification of NS signals is very robust, and we estimated that 1% of this sample has contamination from other low amplitude events. Typical waveforms for neutrons (NS) and other signals such as gamma-rays (ES) are shown in figure 3.

Figure 3. Average NS (top) and ES (bottom) digitised signals from one of the SiPMs. The shaded bands represent the central 68% interval of the signal distribution. An example waveform is also shown for each signal type.

3. Pulse shape discrimination

PSD techniques utilise the dominant shape and amplitude features of digitised signals by selecting a subset of the total information and compressing this into a reduced quantity. An alternative approach that can take advantage of all the information within a signal is to directly use the digitised signals as an input to a suitable machine learning algorithm, which is trained to assign labels to individual signals. This has been shown to provide superior performance in specific applications [8][9]. An ideal machine learning algorithm for PSD are CNN [10] as they have the ability to efficiently extract and combine locally correlated features directly from raw data. As a result they have become the established technique for complex image recognition problems and in recent work have been applied in high-energy physics data analysis where they have been shown to provide comparable or improved performance in tasks such as background rejection [11–13], event reconstruction [14], and event classification [15–19] when trained on the raw detector outputs. CNNs have a number of distinct advantages when working with digitised signals from experimental instruments compared to ANNs, which have previously been demonstrated for PSD [19, 20]; multiple sensor readouts can be processed simultaneously whilst preserving the timing information and correlations between them. This is achieved by assigning each readout to a separate dimension of the input, similar to the RGB channels of a colour image. Additionally, triggered signals from data acquisition systems are generally of varying length, dependent on the energy deposited in an interaction. Such varying length inputs can naturally been handled by fully-convolutional architectures [21], requiring less pre-processing of the digitised signals. In this section we introduce the optimised CNN architecture used for classifying ES and NS signals, as well as two commonly used PSD techniques: the CI and continuous wavelet transform (CWT) methods whose results are used as a benchmark to compare against.

3.1. Charge integration
The CI method [2] is a robust and easily interpretable method of PSD. It uses the ratio of pulse area contained in the tail of a pulse to that contained within a short, initial time period. For a signal $f(t)$, a short integration window, $Q_{short} = \int_0^{t_{short}} f(t)dt$, and long integration window $Q_{long} = \int_0^{t_{long}} f(t)dt$ are defined and the quantity

$$\mathrm{CI} = \frac{Q_\mathrm{long} - Q_\mathrm{short}}{Q_\mathrm{long}} \tag{2}$$

can be used to discriminate between pulse types. The optimal values of the parameters t_short and t_long can be determined by maximising the area under the curve (AUC) metric of the receiver operating characteristic (ROC) curve. They were found to be 245 and 497 ns, respectively.

3.2. Continuous wavelet transform

The CWT [4] provides a powerful method of PSD that is capable of utilising information from both the time and frequency domains of a signal. The CWT of a signal $f(t)$ at a scale a and shift b is defined as

$$W_\mathrm{f}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right) \mathrm{dt}, \tag{3}$$

which can be interpreted as the convolution of the signal with a series of scaled and shifted wavelets $\psi(t)$. The energy density of the wavelet transform at a scale a is defined as

$$P(a) = \frac{1}{1+n_b} \sum_{j=0}^{n_b} |W_f(a,b_j)|^2, \tag{4}$$

which depends heavily on the shape of the signal. Two different scales, a_1 and a_2, can therefore be chosen to discriminate between different signal shapes using the variables $f_1 = P(a_1)$ and $f_2 = P(a_1)/P(a_2)$. A hyperplane in the (f_1, f_2) space can then be used to select different signal shapes. Using the Ricker wavelet, the optimal values of a_1 and a_2 that maximise the AUC were found to be 1 and 900, respectively.
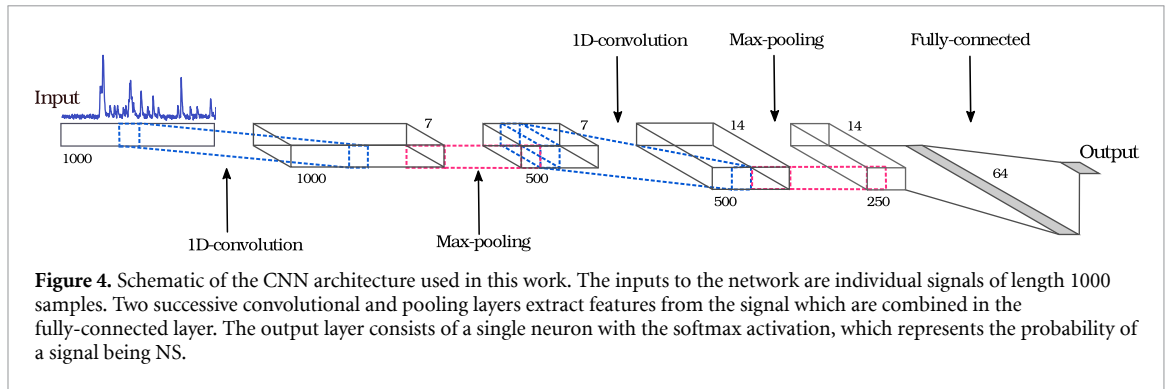
3.3. Convolutional neural network

CNNs are an extension to feed-forward neural networks that are capable of extracting locally correlated features from a multi-dimensional input. This is achieved through the use of *convolutional layers* and *pooling layers*. For a one-dimensional input of length k, a convolutional layer is specified by a set of α filters, each with a width m and a set of learnable weights w_i ($i = 1...m$). This layer transforms an input f with c separate channels each of length k through the operation

$$F_i^\alpha = g\left(\sum_{\beta=1}^{c}\sum_{m=1}^{k} h_i^{\alpha\beta} f_{i+m}^\beta\right), \tag{5}$$

where f^β is the β channel of the input, $h_i^{\alpha\beta}$ are the learnable weights of the filter of the α output channel applied to the β channel of the input, and g is a non-linear activation function. The resulting outputs F_i^α are the *feature maps*, which can be interpreted as a non-linear representation of the input that consists of features extracted by the filters. The filters are learnt during a supervised learning process that minimises a loss function, which for ES/NS discrimination is the standard binary logistic loss. The pooling layers perform dimensionality reduction of the feature maps, and therefore reduce the total number of learnable parameters of the network. *Max-pooling*, defined by taking the maximum value in a small region of the input, is used in this work as it provides a degree of translation invariance. Common CNN architectures use successive convolutional and pooling layers to produce multiple abstract representations of the input, which can then be combined to perform classification. Further explanation of CNNs can be found in [22].
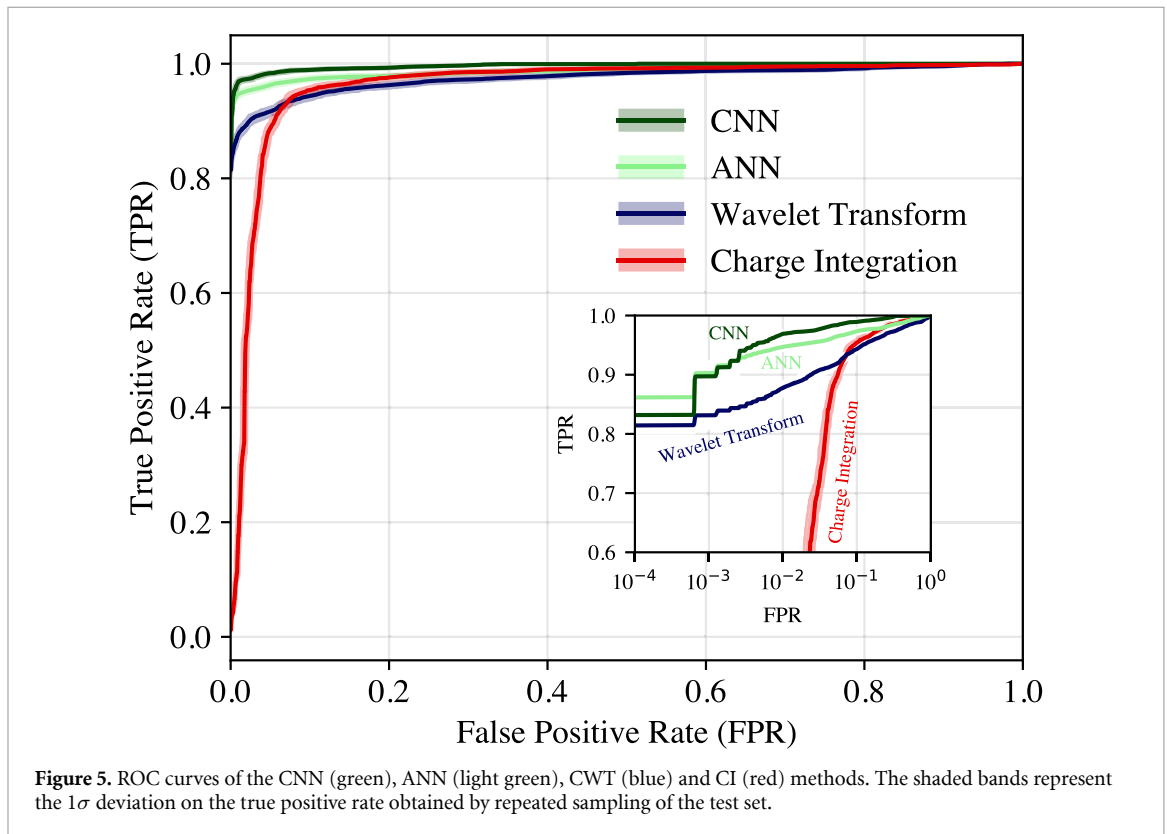
In this work, a CNN architecture was developed that classifies the raw, digitised signals obtained from the experimental set-up introduced in section 2 as either ES or NS signals. For a fair comparison with conventional PSD techniques only one of the SiPM signals was used. A schematic of the complete architecture is given in figure 4, with the specific parameters of the convolutional and pooling layers given in table 1. The CNN transforms each signal through two successive convolution and pooling layers (window size of 2), where the non-linear ReLU activation function was used in both convolutional layers. The resulting feature maps are fed into a fully-connected layer with 64 neurons, that each have the ReLU activation. The output of the CNN consists of a single neuron with the softmax activation, and can therefore be be interpreted as the probability of the signal being NS. To classify signals, a threshold must be determined above (below) which all signals will be classified as NS (ES). Keras [23] was used with the Tensorflow [24] back-end to train the CNN. The weights of the network were optimised by minimising the cross-entropy loss on a training sample of 12 000 signals with an equal proportion of ES and NS signals. The network was trained for a total of 50 epochs using the Adam optimiser [25] with a batch size of 256 samples and initial learning rate of 0.001.

A simple ANN based on one fully connected layer of 64 neurons was also trained to evaluate the difference of performance given by the presence of the convolutional layers. Other ANN model with different

Figure 4. Schematic of the CNN architecture used in this work. The inputs to the network are individual signals of length 1000 samples. Two successive convolutional and pooling layers extract features from the signal which are combined in the fully-connected layer. The output layer consists of a single neuron with the softmax activation, which represents the probability of a signal being NS.

Table 1. Parameters of the convolutional and pooling layers of the CNN architecture.

Layer	# Channels	Filter Size
Conv	7	10
Max-pool	–	2
Conv	14	10
Max-pool	–	2

Figure 5. ROC curves of the CNN (green), ANN (light green), CWT (blue) and CI (red) methods. The shaded bands represent the 1σ deviation on the true positive rate obtained by repeated sampling of the test set.

number of neurons where tested to ensure that this model had sufficient capacity for a fair comparison. The Adam optimiser and a similar number of epochs was also used to train the ANN.
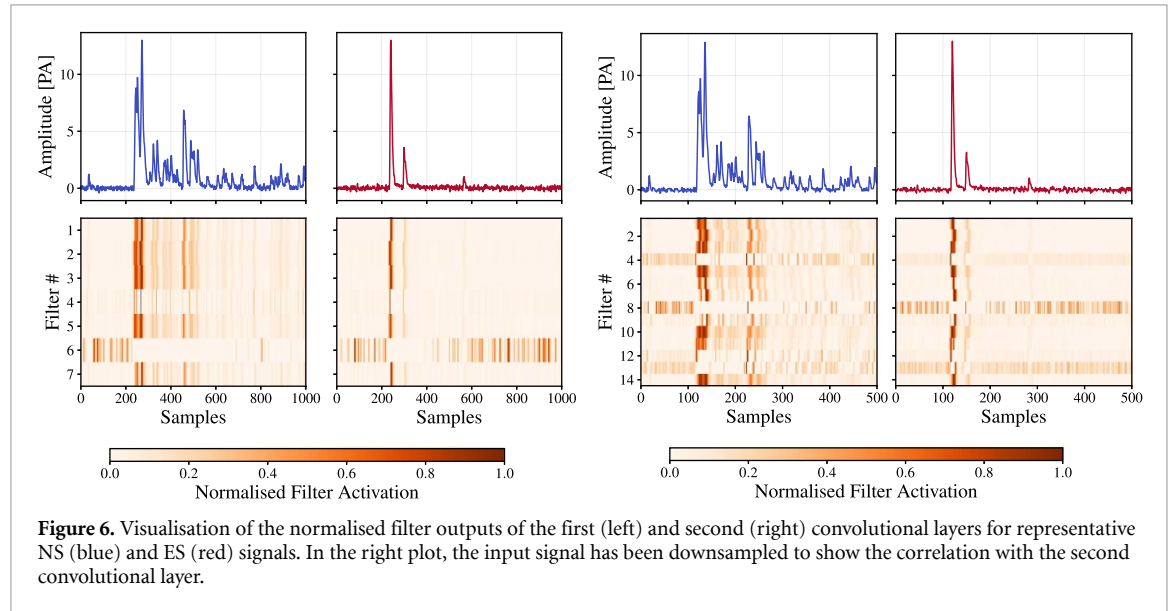
4. Results

The performance of the CNN, the ANN, as well as the CI and CWT methods was evaluated on a separate test sample of 3000 signals. Figure 5 compares the ROC curves of the CNN to the other methods on this sample, with the AUC metrics given in table 2. Errors on the ROC curves and AUC metric are estimated by repeatedly evaluating the performance on random samples of the test set. The order of performance obtained follows the expected use of input features by the different algorithms. Both the CNN and the ANN provides a better

Table 2. Area under the curve (AUC) values for each of the PSD methods considered in this work, with 1σ uncertainties given in brackets.

Method	AUC	$(\pm 1\sigma)$
CNN	0.996	(± 0.003)
ANN	0.987	(± 0.003)
CWT	0.974	(± 0.003)
CI	0.964	(± 0.004)

Figure 6. Visualisation of the normalised filter outputs of the first (left) and second (right) convolutional layers for representative NS (blue) and ES (red) signals. In the right plot, the input signal has been downsampled to show the correlation with the second convolutional layer.

AUC than other methods and the CNN is able to achieve a significantly higher level of ES/NS discrimination compared to all other methods, with an AUC of 0.996 ± 0.003.

5. Feature interpretation

The superior performance of the CNN is due to its ability to efficiently extract and combine the features relevant for discrimination, directly from the raw signal. To understand these features, the normalised feature maps of the two convolutional layers are visualised in figure 6 for representative NS and ES signals. It can be seen that in the first convolutional layer, most of the filters activate at the peaks of the signals and it can be see that filter 4 activates only at the maximum of each peak. This allows the CNN to encode both the number of peaks in a signal along with their relative amplitudes. Filter 6 appears to activate when the signals are below some threshold, which allows the CNN to extract quantities similar to ToT. It can be seen that in the second convolutional layer the feature maps become more complex, with multiple features of the first convolutional layer combined. For example, filter 4 appears to activate both on low amplitude regions of the signal as well as at the steeply rising edges of high amplitude pulses.

Further understanding of how the CNN interprets the signals can be obtained by visualising the high dimensional representation of each signal that is encoded within the 64-dimensional space of the fully-connected layer. The t-SNE algorithm [26] is used to create a two-dimensional embedding of this space, which preserves a level of the local and global structure of the original space. This allows us to visualise in a qualitative way both the features encoded by the CNN, and to determine clusters of signals that have similar representations. The results of this embedding on the training set are shown in figure 7, with each signal in this space coloured by the magnitude of the CNN output. It can be seen that the t-SNE shows the signals spread along an S-shape, forming two distinct ES and NS clusters. Within these clusters, four regions (A, B, C, D) of distinctly different signals have been highlighted, with representative signals in these regions shown in figure 8. By examining these regions we find that there is is a continuous substructure of signals with varying energies, that gradually transforms from ES to NS signals, and that the CNN is using both the shape and energy information to discriminate between signals. The upper cluster consists entirely of NS signals, with those close to region A having a large amplitude pulse and a slowly decaying tail with many peaks, characteristic of high energy NS signals. Moving through this cluster we find waveforms with a continuous decrease in the amplitude and length of the NS pulse.
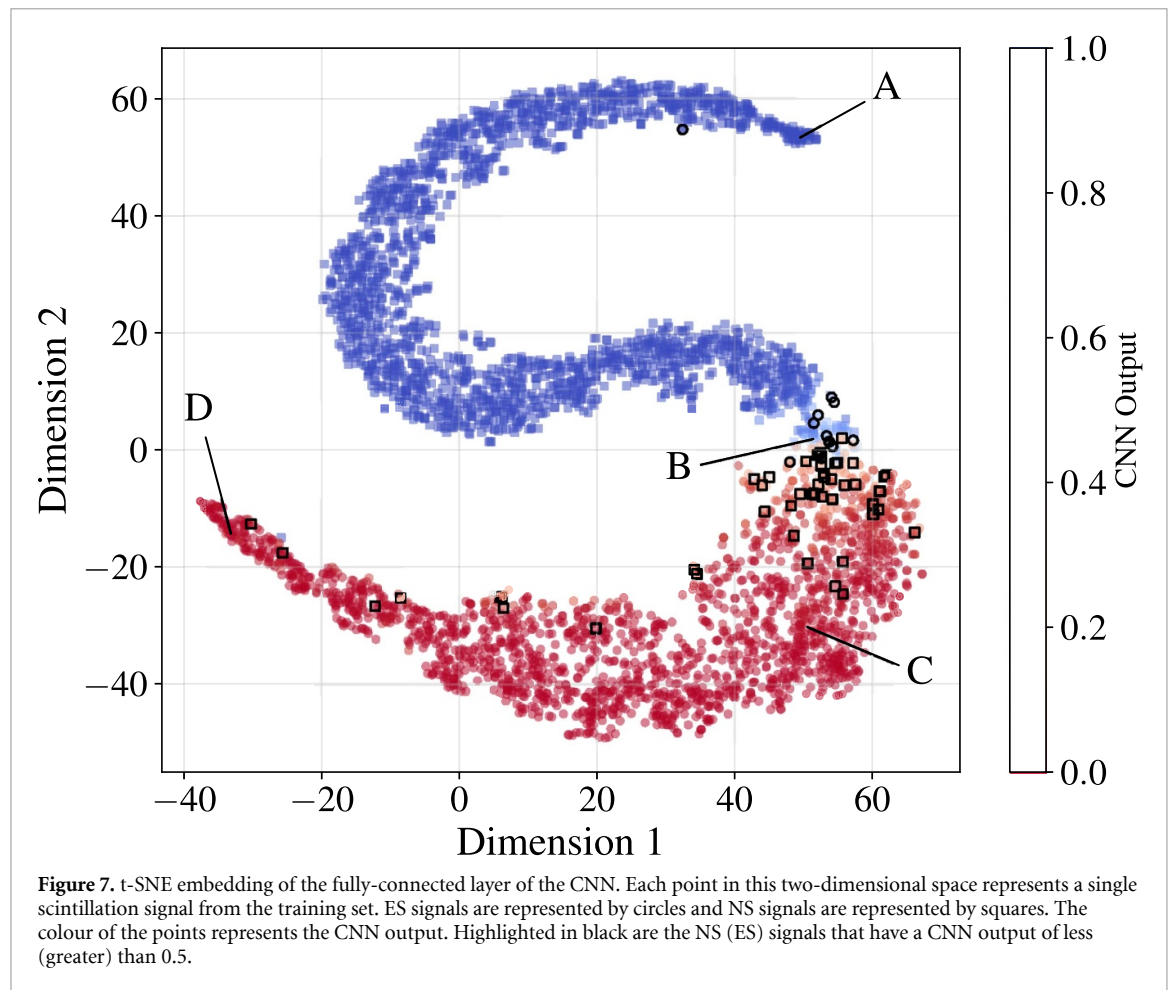
Figure 7. t-SNE embedding of the fully-connected layer of the CNN. Each point in this two-dimensional space represents a single scintillation signal from the training set. ES signals are represented by circles and NS signals are represented by squares. The colour of the points represents the CNN output. Highlighted in black are the NS (ES) signals that have a CNN output of less (greater) than 0.5.

A similar structure is seen for the lower cluster of ES signals. The lower energy ES signals with a single prompt peak are found close to C, and in region D are high amplitude saturated signals, most likely caused by extremely energetic atmospheric muons. Region B contains the pulses that the CNN is not able to confidently classify as either ES or NS. These appear to be a collection of low energy signals that have several prompt peaks, most likely to be the pile-up of single photon avalanches.
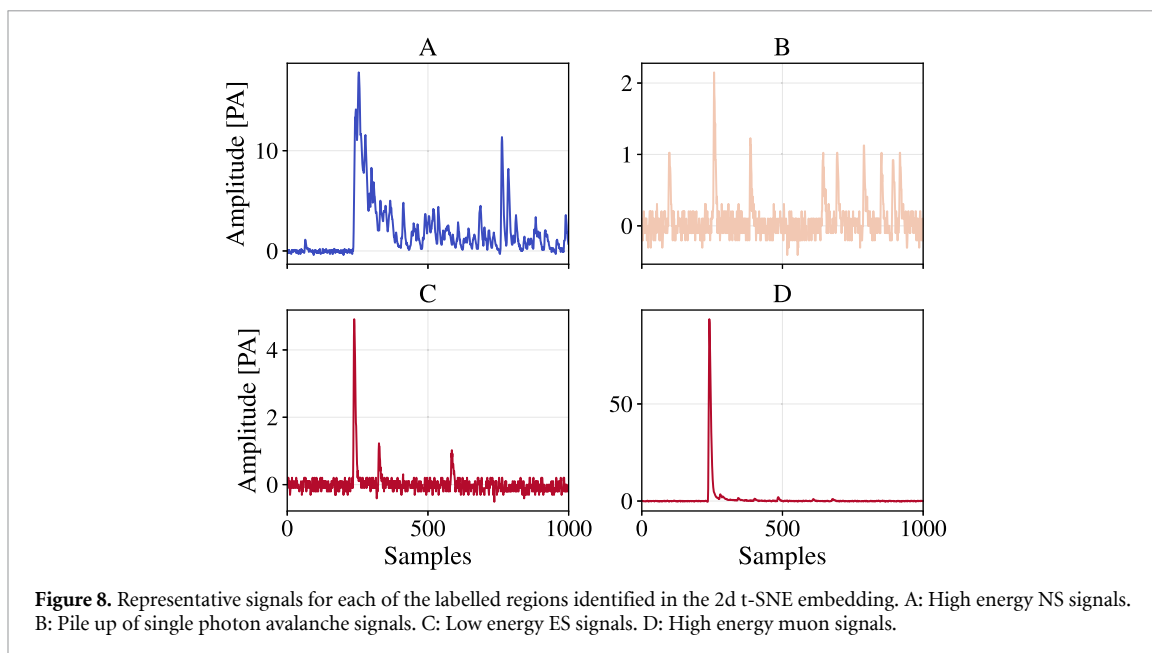
The signals highlighted in black are the NS (ES) signals that have a CNN output value of less (greater) than 0.5, and would therefore likely be misclassified when threshold cut is placed on the CNN output. The majority of these events appear in the cluster of ES signals and upon further inspection of these signals, they are clearly ES or pile-up signals that were incorrectly labelled by the simple selection shown in figure 3. The fraction of misclassified signals is $\mathcal{O}(1\%)$ and is therefore in agreement with the original estimation of the contamination of the NS sample. This demonstrates that the CNN is robust to, at least, a small proportion of mislabeled training data and can correct the originally mislabelled signals.

Furthermore, it is possible to use the discovered substructure shown in figure 8 to further sub-classify and divide pulses beyond that of the initial labeling of ES and NS, and could be used for example to remove pile-up or introduce a more suitable set of labels for the signals. To summarise, our investigation suggests that first, it is not necessary to have a perfect set of labels, the CNN is robust to a small proportion of misclassification and second, supervised learning is effective at recognising a wide range of sub-structure in the data. Through the exploration and clustering of data in this low-dimensional embedding, it is possible to further explore and understand a raw detector dataset without requiring unsupervised learning techniques.

6. Conclusions

In this work we have demonstrated a convolutional neural network architecture that provides superior performance in the classification of digitised signals compared to an ANN and more traditional PSD methods, achieving an AUC of 0.996 ± 0.003 on the set of high purity signals obtained from a single ^{6}Li PVT detector element. The 1-dimensional CNN architecture presented here is also more flexible than a simple ANN, fast to train on a CPU and therefore, competes very well with current methods in all practical aspects

Figure 8. Representative signals for each of the labelled regions identified in the 2d t-SNE embedding. A: High energy NS signals. B: Pile up of single photon avalanche signals. C: Low energy ES signals. D: High energy muon signals.

of implementation of a PSD algorithm. The use of raw digital inputs means the approach can be generalised to other scintillators and possibly, in the future, deployed directly in front-end electronics. By investigating the representations learned by the CNN we have shown that it is possible to interpret the features extracted by the convolutional layers and therefore gain an understanding of how the CNN discriminates between different signal types. Further to this, we have shown that by visualising the high-dimensional representations it is possible to identify substructure within the signal types, even though the CNN was trained to perform a binary classification task. This approach could be utilised in many other areas of physics data analysis, such as to discover clusters of events in raw detector data without relying on hand-crafted variables.

Acknowledgments

ORCID iD

A Vacheret https://orcid.org/0000-0001-7792-4349

References

[1] Abreu Y *et al* 2017 A novel segmented-scintillator antineutrino detector *JINST* **12** P04024
[2] Legler E, Attwenger W, May F and Quittner G 1965 Pulse Shape Discrimination System for ^{6}LiF(ZnS) Scintillation Counters *Rev. Scientific Instrum.* **36** 1167
[3] Balmer M J, Gamage K A and Taylor G C 2015 Comparative analysis of pulse shape discrimination methods in a 6li loaded plastic scintillator *Nucl. Instrum. Methods Phys. Res.* A **788** 146–53
[4] Yousefi S, Lucchese L and Aspinall M 2009 Digital discrimination of neutrons and gamma-rays in liquid scintillators using wavelets *Nucl. Instrum. Meth.* **A598** 551–5
[5] Nakhostin M 2020 A technique for the reduction of pulse pile-up effect in pulse-shape discrimination of organic scintillation detectors *Nucl. Eng. Technol.* **52** 360–5
[6] Abreu Y *et al* 2018 Performance of a full scale prototype detector at the BR2 reactor for the SoLid experiment *JINST* **13** P05005 arXiv: 1802.02884
[7] Abreu Y *et al* 2019 Commissioning and operation of the readout system for the SoLid neutrino detector, *JINST* **14** P11003 arXiv:1812.05425
[8] Flores J L, Martel I, Jiménez R, Galán J and Salmerón P 2016 Application of neural networks to digital pulse shape analysis for an array of silicon strip detectors, *Nucl. Instrum. Meth.* **A830** 287–93
[9] Sanderson T S, Scott C D, Flaska M, Polack J K and Pozzi S A 2012 Machine learning for digital pulse shape discrimination *2012 IEEE Nuclear Symp. and Medical Conf. Record (NSS/MIC)* pp 199–202
[10] LeCun Y, Bengio Y, Hinton G and Learning D 2015 *Nature* **521** 436

[11] Komiske P T, Metodiev E M and Schwartz M D 2017 Deep learning in color: towards automated quark/gluon jet discrimination *JHEP* **2017** 110

[12] Renner J *et al* 2017 Background rejection in next using deep neural networks *JINST* **12** T01004

[13] Qiao H, Lu C, Chen X, Han K, Ji X, Wang S 2018 Signal-background discrimination with convolutional neural networks in the PandaX-III experiment arXiv:1802.03489

[14] Delaquis S, *et al* 2018 Deep neural networks for energy and position reconstruction in EXO-200 arXiv:1804.09641

[15] Racah E *et al* 2016 Revealing Fundamental Physics from the Daya Bay Neutrino Experiment using Deep Neural Networks arXiv:1601.07621

[16] Aurisano A *et al* 2016 A Convolutional Neural Network Neutrino Event Classifier *JINST* **11** P09001

[17] Acciarri R *et al* 2017 Convolutional neural networks applied to neutrino events in a liquid argon time projection chamber *JINST* **12** P03011

[18] Khosa C K, Mars L, Richards J and Sanz V 2020 Convolutional neural networks for direct detection of dark matter *J. Phys. G: Nucl. Part. Phys.* **47** 9

[19] Holl H L M B P *et al* 2019 Deep learning based pulse shape discrimination for germanium detectors *Eur. Phys. J.* C **79** 450 arXiv:1903.01462v2

[20] Fu C, Di Fulvio A, Clarke S, Wentzloff D, Pozzi S and Kim H 2018 Artificial neural network algorithms for pulse shape discrimination and recovery of piled-up pulses in organic scintillators *Ann. Nucl. Energy* **120** 410–21

[21] Long J, Shelhamer E, Darrell T 2014 Fully convolutional networks for semantic segmentation arXiv:1411.4038

[22] Rawat W and Wang Z 2017 Deep convolutional neural networks for image classification: A comprehensive review *Neural Comput.* **29** 2352–449

[23] Chollet F, *et al* 2015 Keras (https://keras.io)

[24] Abadi M, *et al* 2016 TensorFlow: A system for large-scale machine learning arXiv:1605.08695

[25] Kingma D P and Ba J 2014 Adam: A method for stochastic optimization, CoRR abs/1412.6980 arXiv:1412.6980

[26] van der Maaten L and Hinton G 2008 Visualizing data using t-SNE *J. Machine Learning Res.* **9** 2579–605