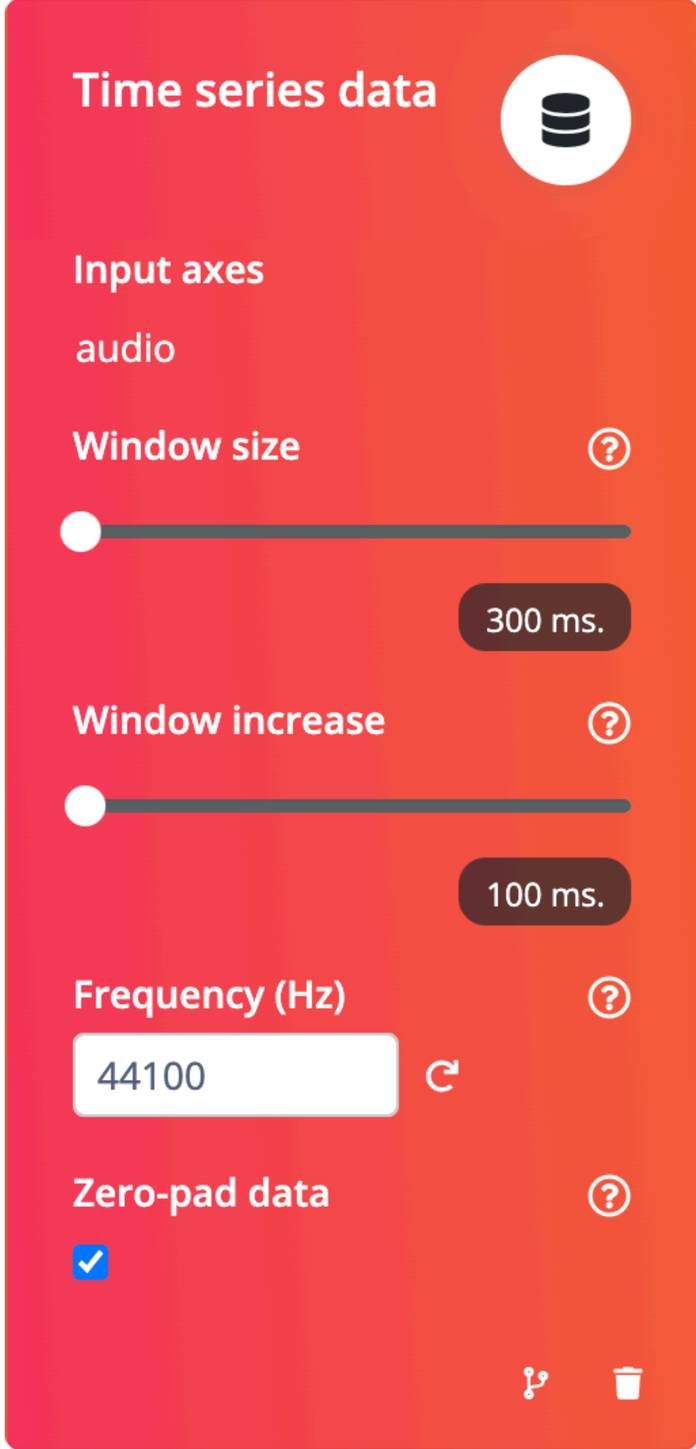
NotiSound

Time Series Data

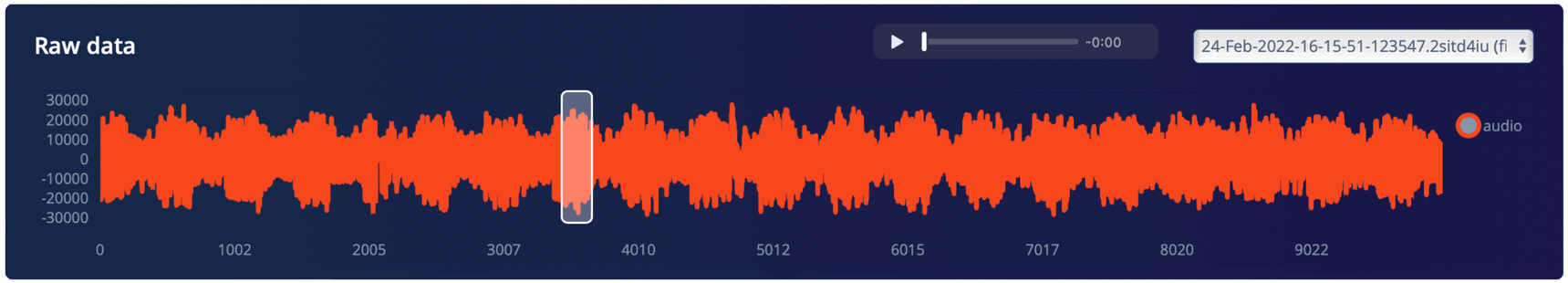
From Microphone

Edge impulse slices the raw input samples into windows that are used by the CNN during training.

- Window size: each window is 300 ms long
- Window increase: 100ms offset of each subsequent window from the previous



Mel-filterbank Energy Features



Raw features

```
-3510, -1170, 585, 2340, 7021, 9362, 11117, 8776, 5266, 2925, 1170, -4681, -11117, -11702,
```

Parameters

Mel-filterbank energy features

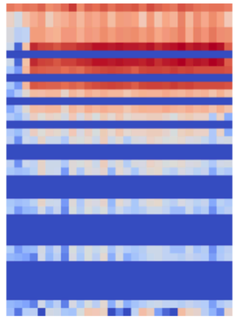
Frame length	0.02
Frame stride	0.01
Filter number	40
FFT length	256
Low frequency	300
High frequency	4000

Normalization

Noise floor (dB)	-52

DSP result

Spectrogram



Processed features

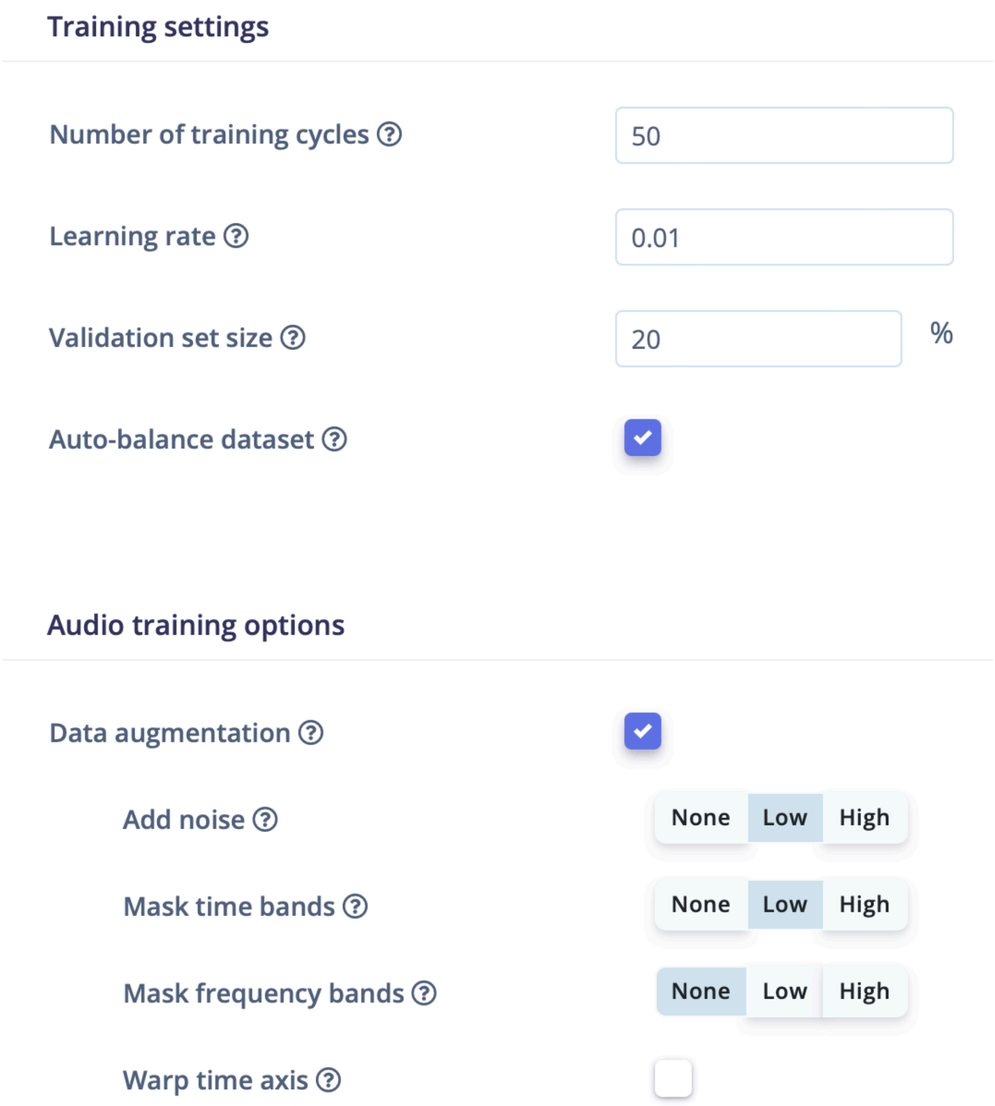
```
0.2031, 0.1797, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.1797, 0.2344, 0.0000, 0.0000…
```

On-device performance

PROCESSING TIME
69 ms.

PEAK RAM USAGE
16 KB

Convolutional Neural Network Configuration



- The model was trained over 50 epochs, with a learning rate of 0.01 (how fast the NN learns)
- Validation Set - The percentage of data samples used for validation: 20%

- Random noise was added to the training data
- Random time bands were masked

Convolutional Neural Network Architecture

1D convolution

Keras Model:

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 reshape_3 (Reshape)         (None, 29, 40)            0

 conv1d_6 (Conv1D)           (None, 27, 8)             968

 max_pooling1d_6 (MaxPooling  (None, 9, 8)             0
 1D)

 dropout_6 (Dropout)         (None, 9, 8)              0

 conv1d_7 (Conv1D)           (None, 7, 16)             400

 max_pooling1d_7 (MaxPooling  (None, 3, 16)            0
 1D)

 dropout_7 (Dropout)         (None, 3, 16)             0

 flatten_3 (Flatten)         (None, 48)                0

 dense_3 (Dense)             (None, 3)                 147

=================================================================
```

Performance

- *Accuracy: 97.8%* of windows of audio that were correctly classified
- *Loss: 0.09* is the cross entropy loss of the model
- *Confusion matrix*: shows the balance of correctly versus incorrectly classified windows.
- F1 scores: the harmonic mean of precision and recall is around 0.97-0.98 for all the three classes

Last training performance (validation set)

ACCURACY
97.8%

LOSS
0.09

Confusion matrix (validation set)

	DONG	FIREALARM	NOISE
DONG	94.2%	3.5%	2.3%
FIREALARM	0.1%	97.3%	2.6%
NOISE	0.0%	0.2%	99.7%
F1 SCORE	0.97	0.98	0.98