Classifying Ultrasonic Mouse Vocalizations using a TensorFlow Network

Drew Kristensen
Dr Adam A Smith

Why classify mouse vocalizations?

What are the characteristics of an autistic mouse?

- It's difficult to create a mouse model of mental diseases

Decoding mouse ultrasonic vocalizations (USVs) might provide insight into their affective state

Potentially helpful in modeling mental disorders

Current methods rely on human operators to label where calls occur and don't occur - time intensive and we have better things to do
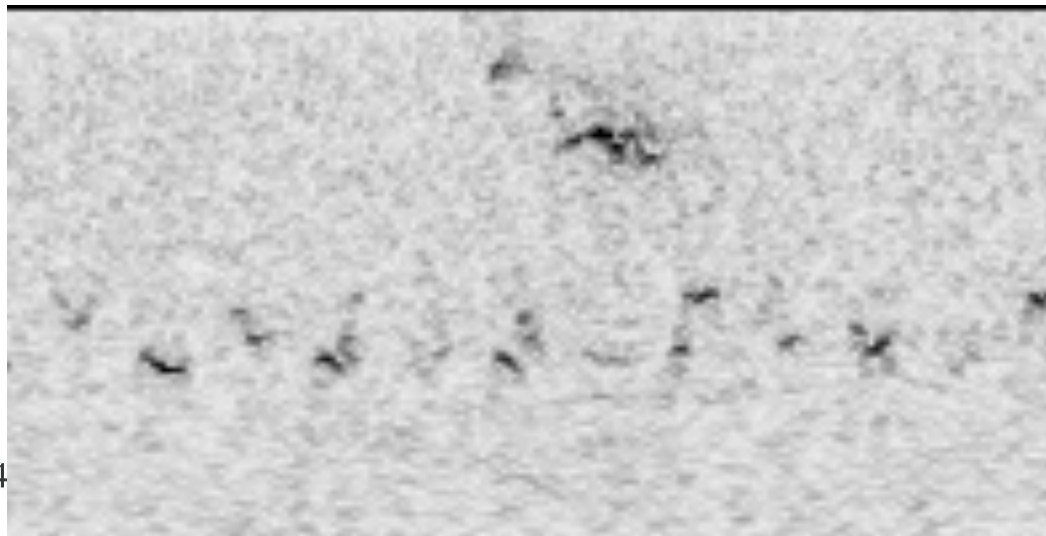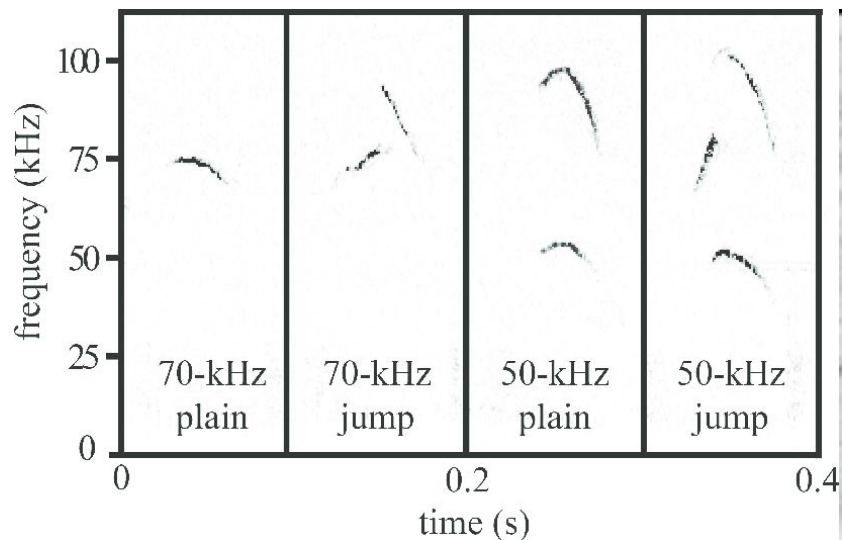
What is the goal?

Classify USVs from recordings using a neural network from the tensor flow framework

Mouse Vocalizations

Often emitted between 30 kHz and 100 kHz

Human hearing goes up to 20 kHz under perfect conditions (hence the 'ultrasonic')
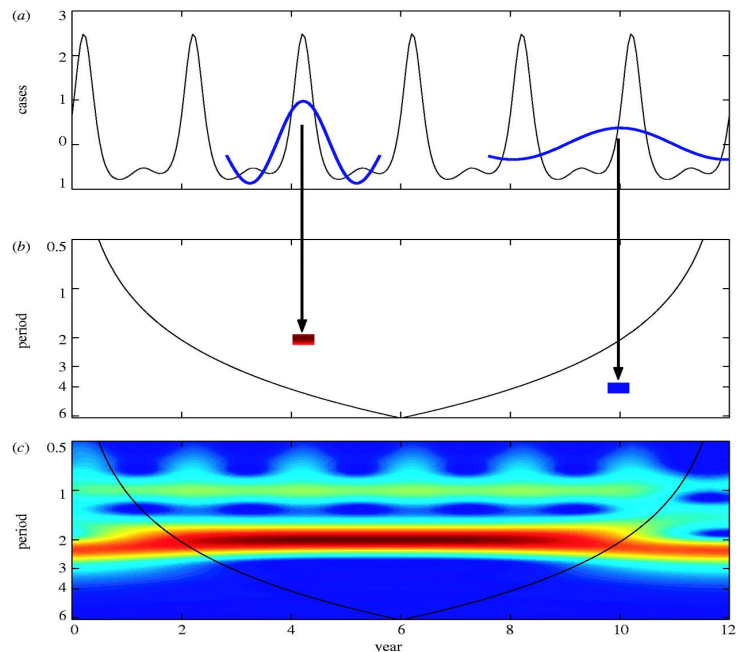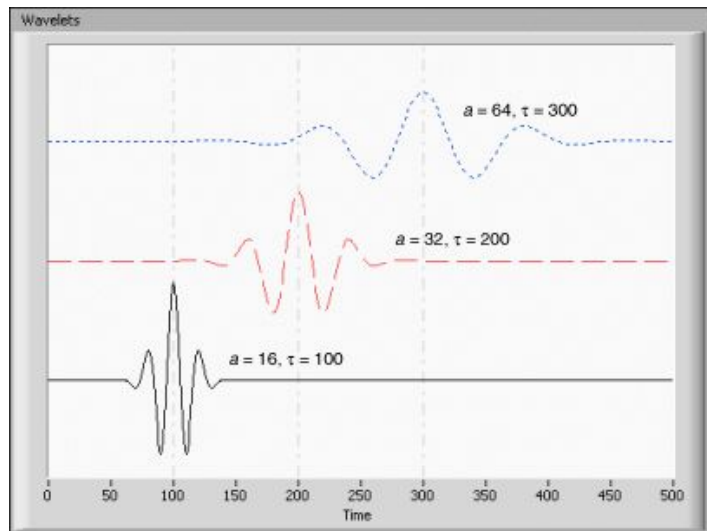
Scalograms

We hear in log-scale, so how can we make our data reflect that?

Scalograms are visual representations of the audio data

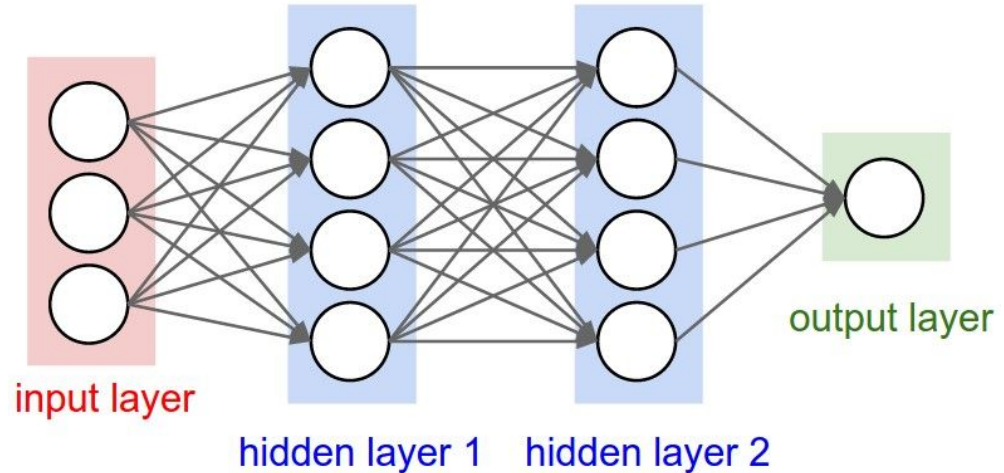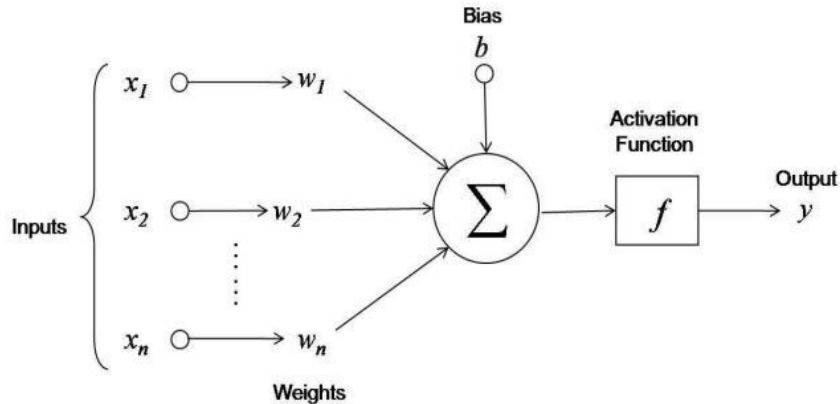Uses wavelet transforms instead of Fourier Series

Neural Network

All the rage in Machine Learning
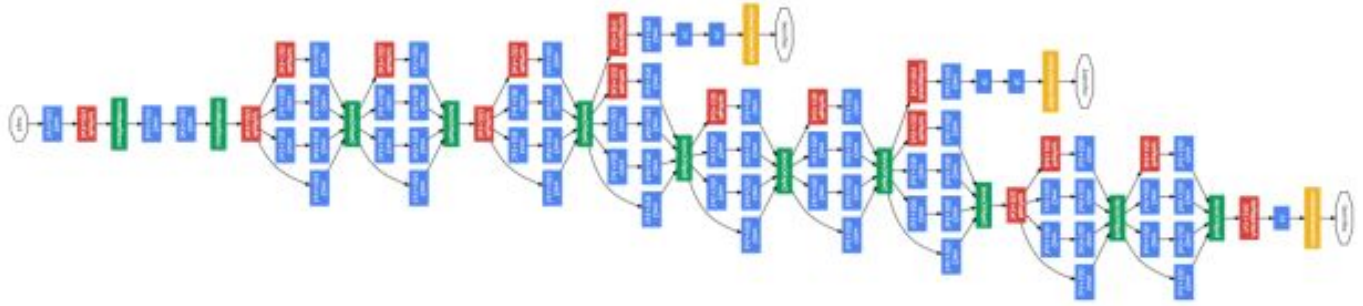
Useful for classification problems

An attempt to model the abilities of human pattern detection

TensorFlow

Google's open source framework for creating computational data flow graphs
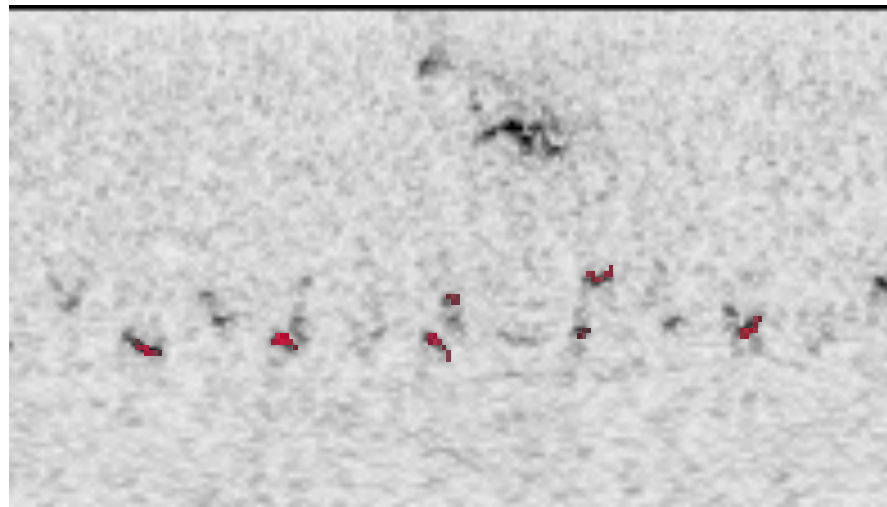
Extremely powerful

Google's LeNet (A really really complex data flow graph)

The Identification Process

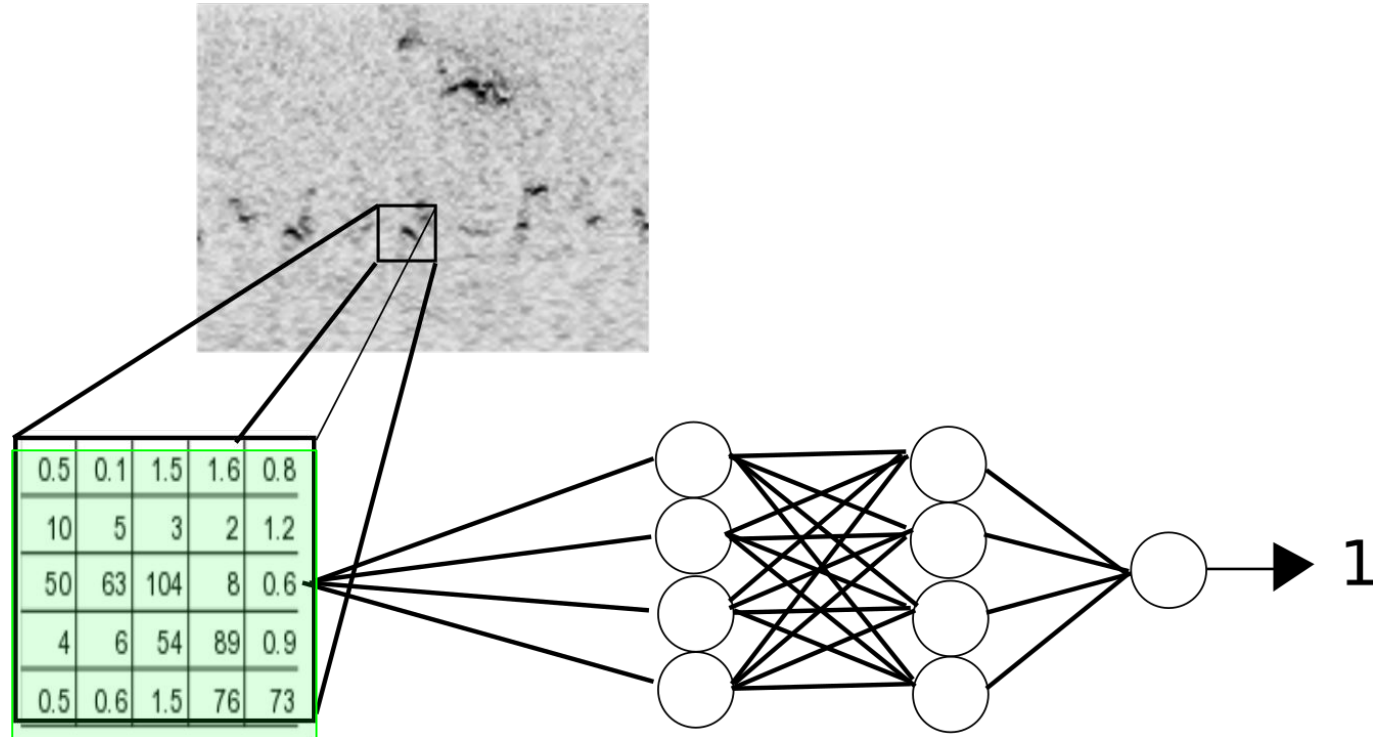Start with .wav files, sampled at 250 kHz

Find the wavelet coefficients between 30 kHz and 120 kHz
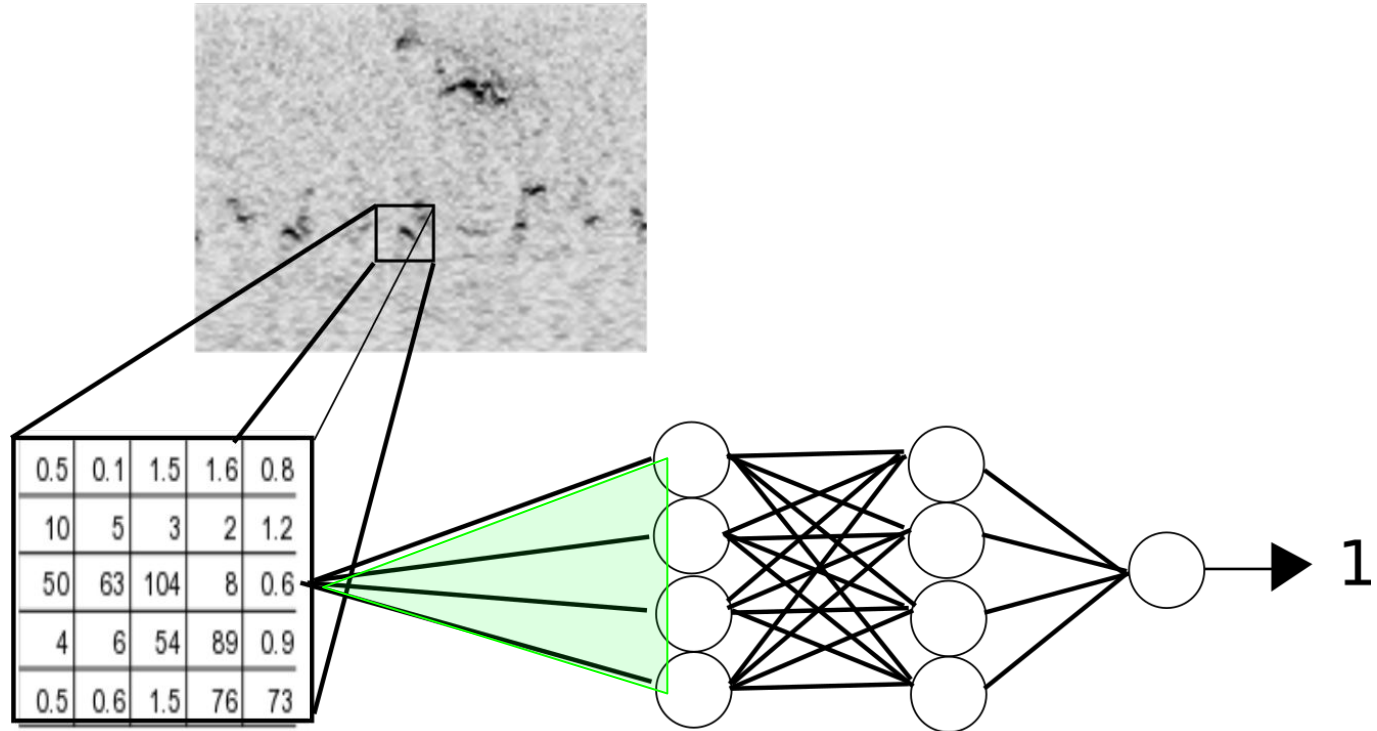
For training, annotate the scalogram

The Identification Process

Take the coefficients around the point we want to classify and flatten them into a vector
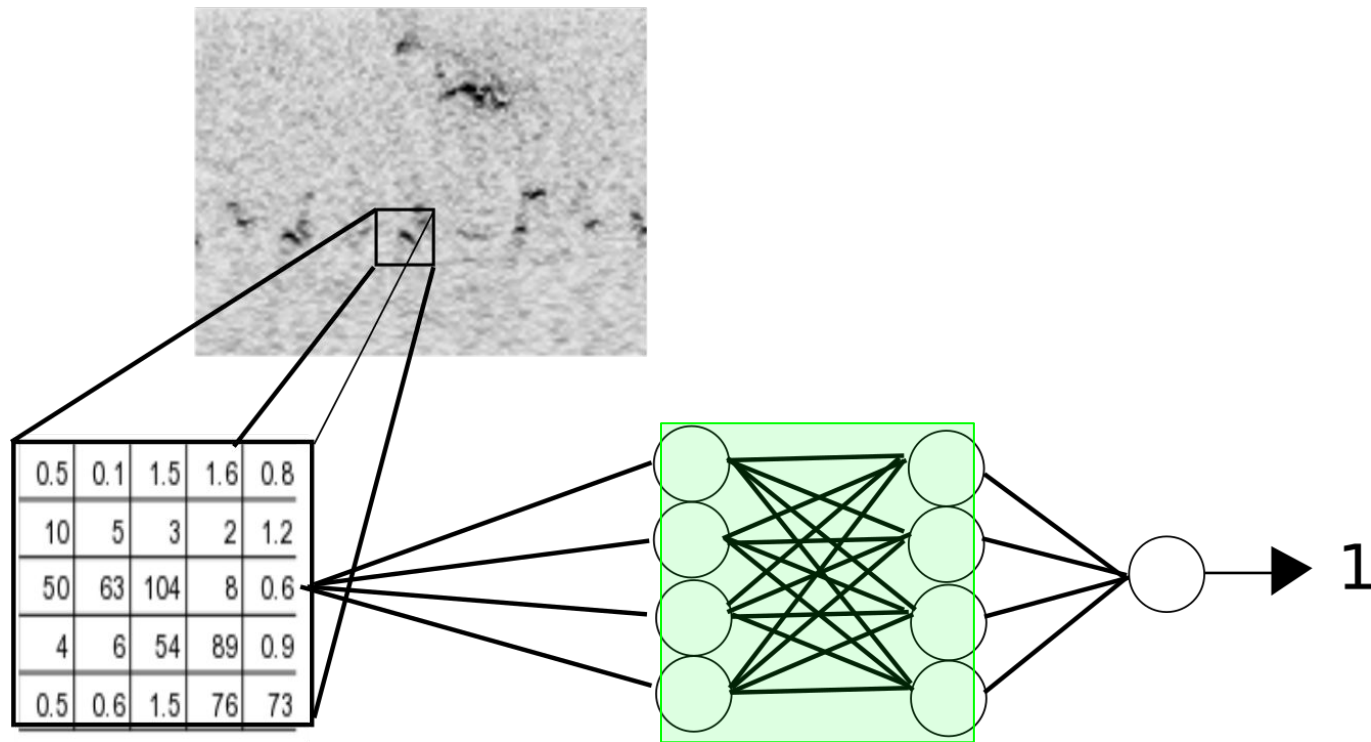
The Identification Process

Feed the input
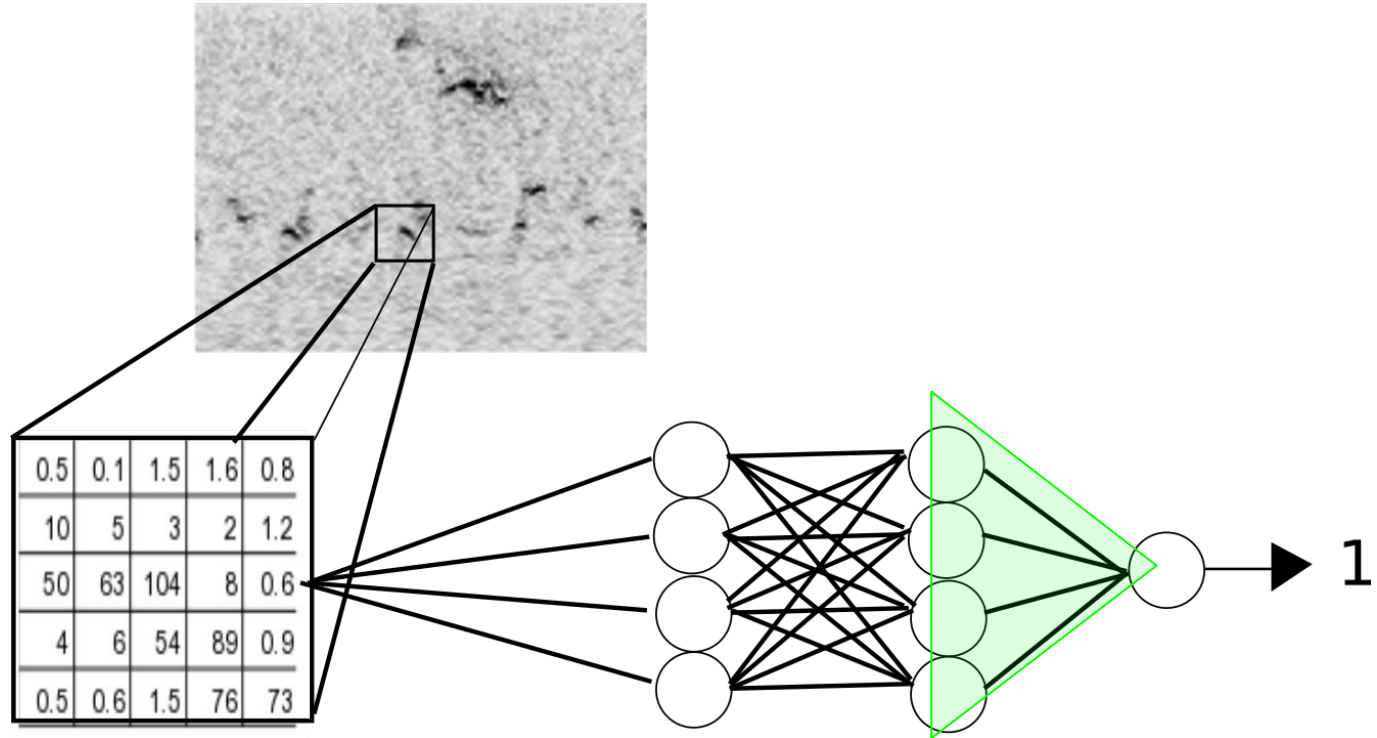vector into the first
layer of the neural
network

The Identification Process

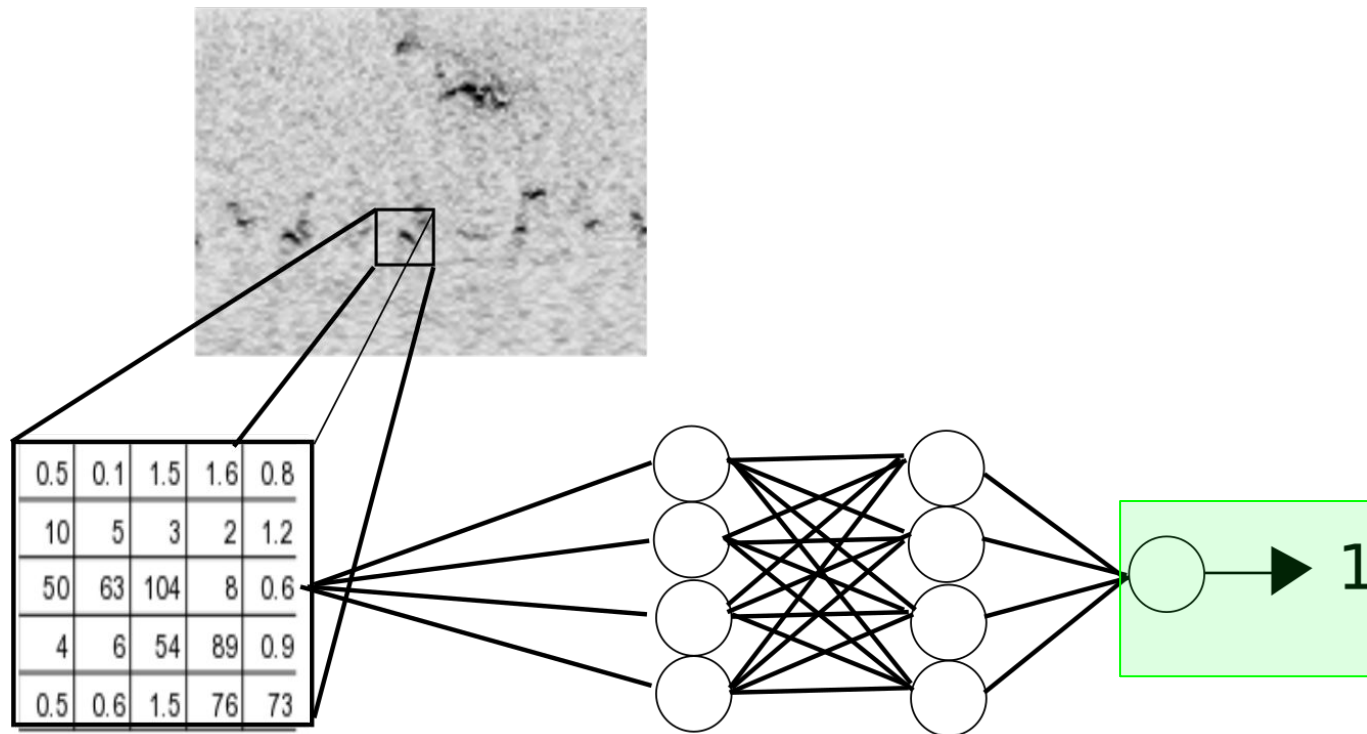The first layer outputs their weighted sums to the second layer

The Classification Process

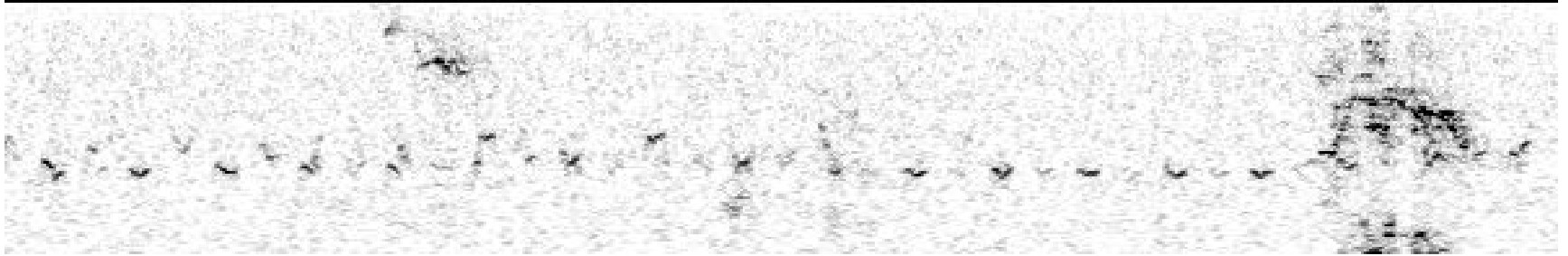The second layer outputs their values to the final neuron - the output layer

The Identification Process

Since the input was centered around a call, the network outputs a value of 1
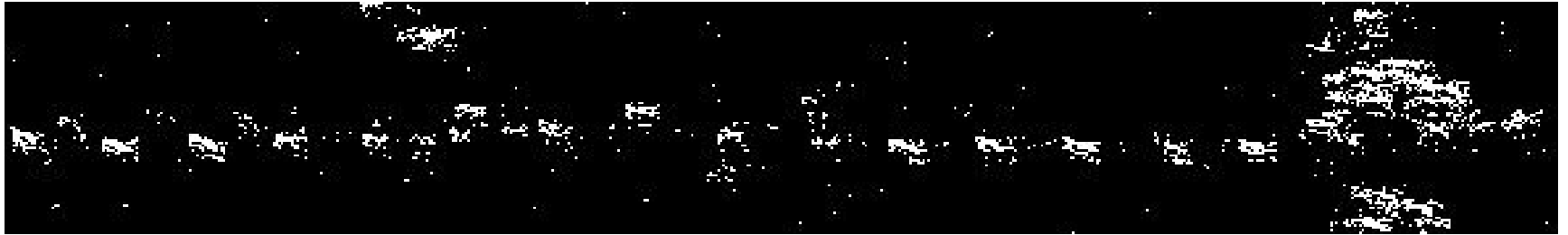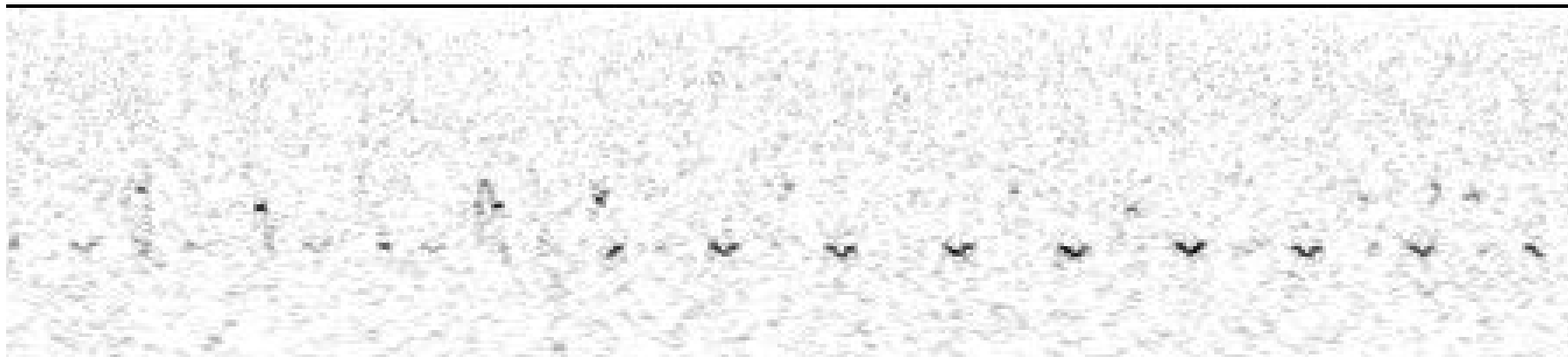
Current Results

Input Scalogram
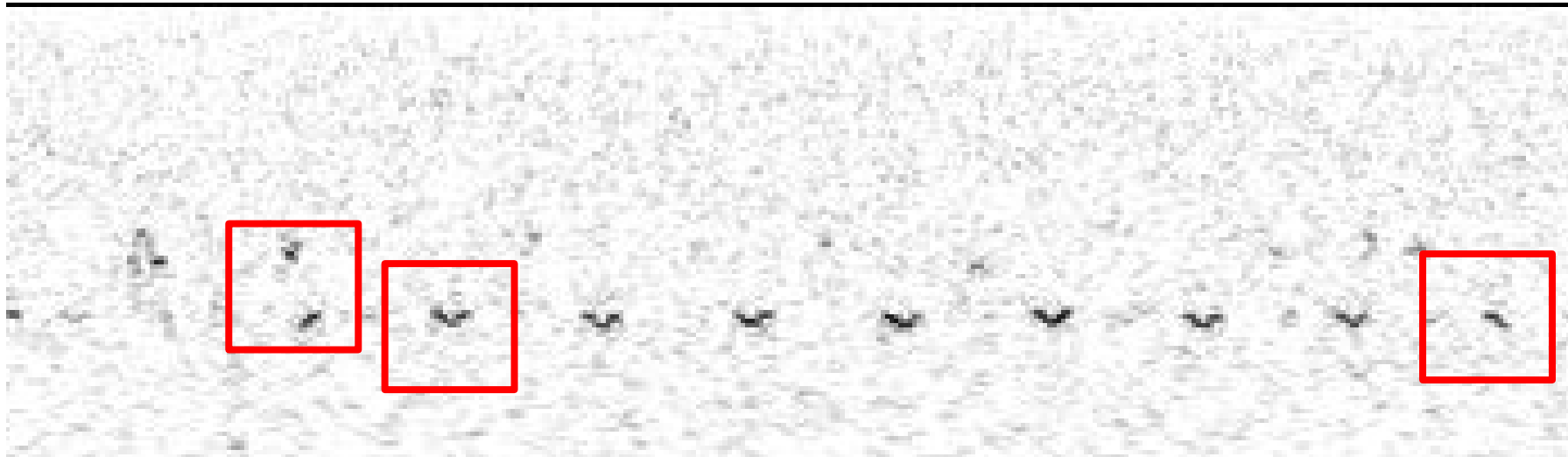

Elements with output of at least 1

Challenges

Need for human labeled data - not always easy to annotate*

What is a call and what isn't?

Future Directions

Classify different types of calls - not just their presence

Questions?