**Speech Classification**

**Data Exploration:**

1. **Datasets**

* 4,000 records (.wav files) of "call 911" – Essence Dataset
* 60,000 records (.wav files) of 30 other words – taken from **Kaggle**:

"Tensorflow Speech Recognition Challenge"

* The sample rate of each record is 16kHz

1. **Balancing the data:**

For balancing the data, we choose randomly and uniformly 4000 records from the Kaggle dataset.

The result of this step is the balanced data:

* 4,000 records of "call 911"
* 4,000 records of other words

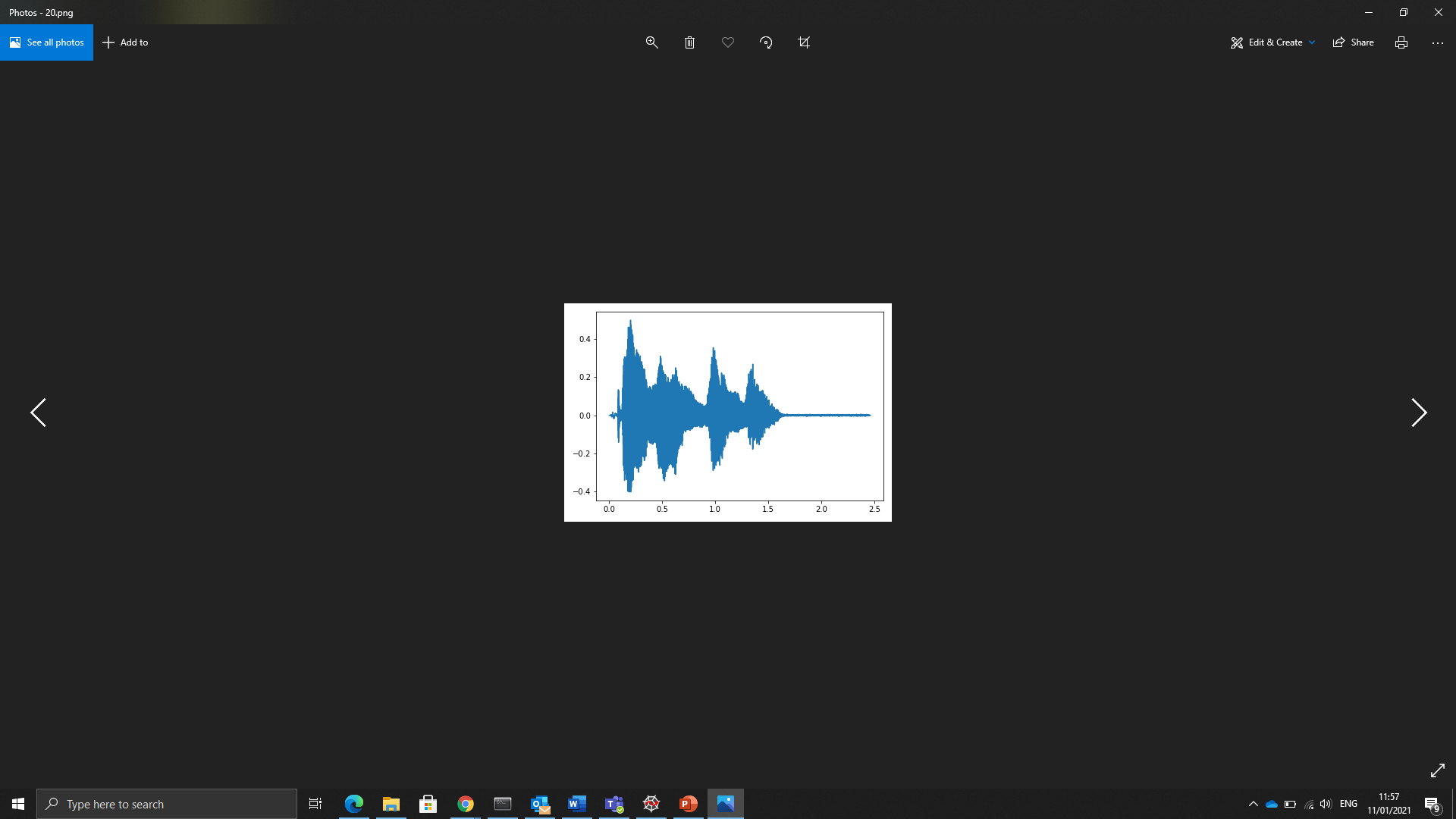
1. **Plot examples:**

X axis stands for time domain (sec)

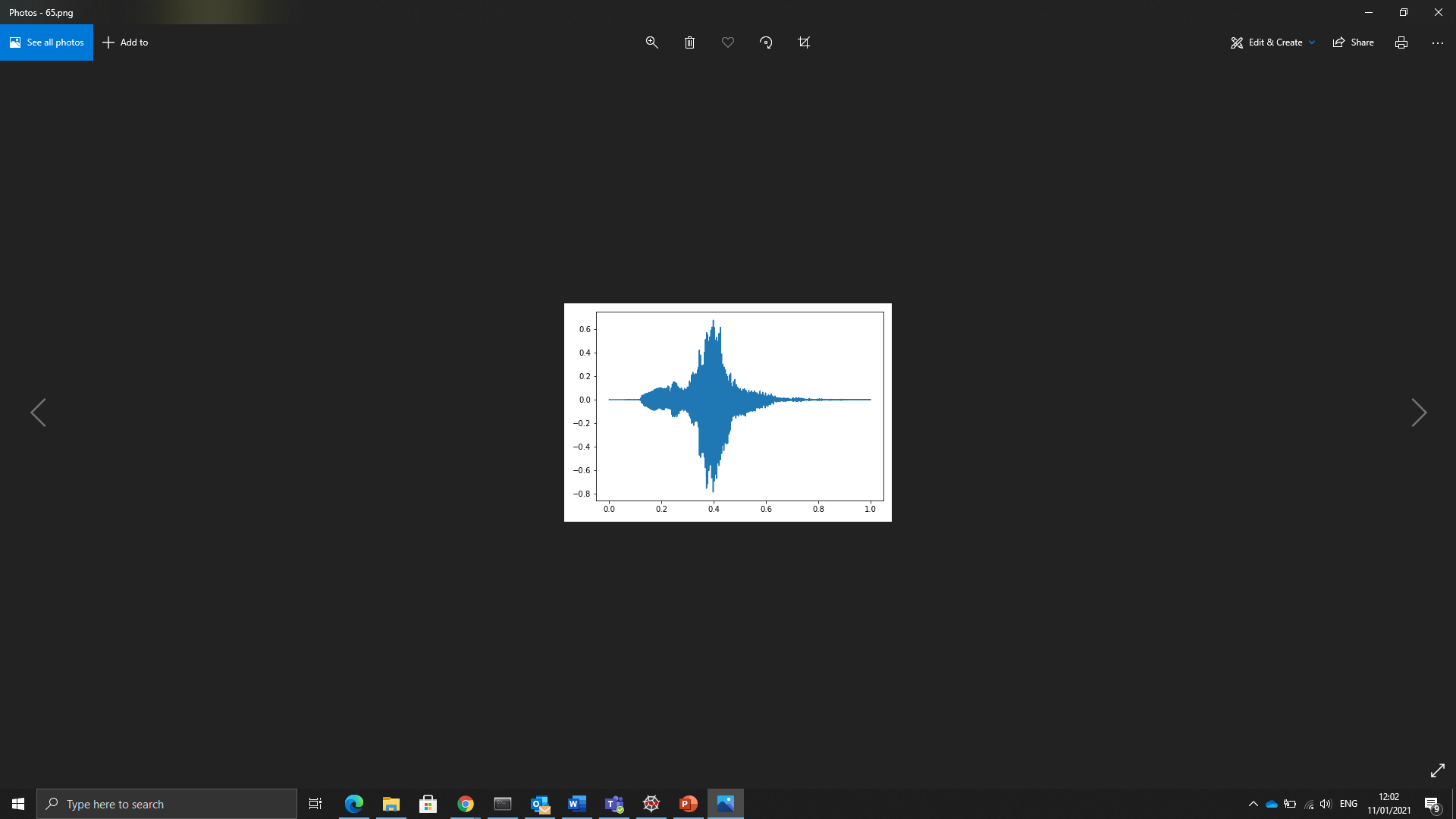
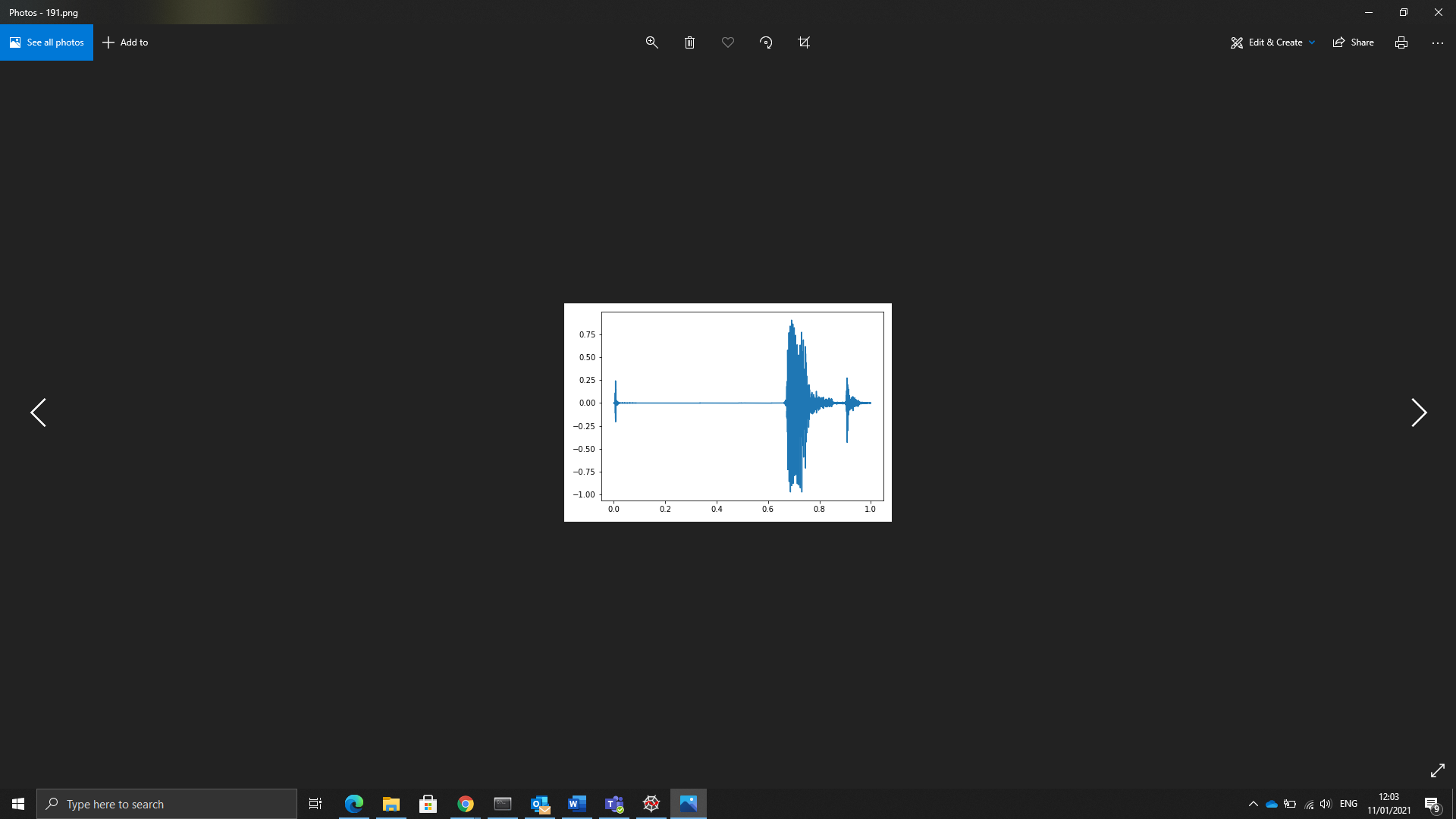
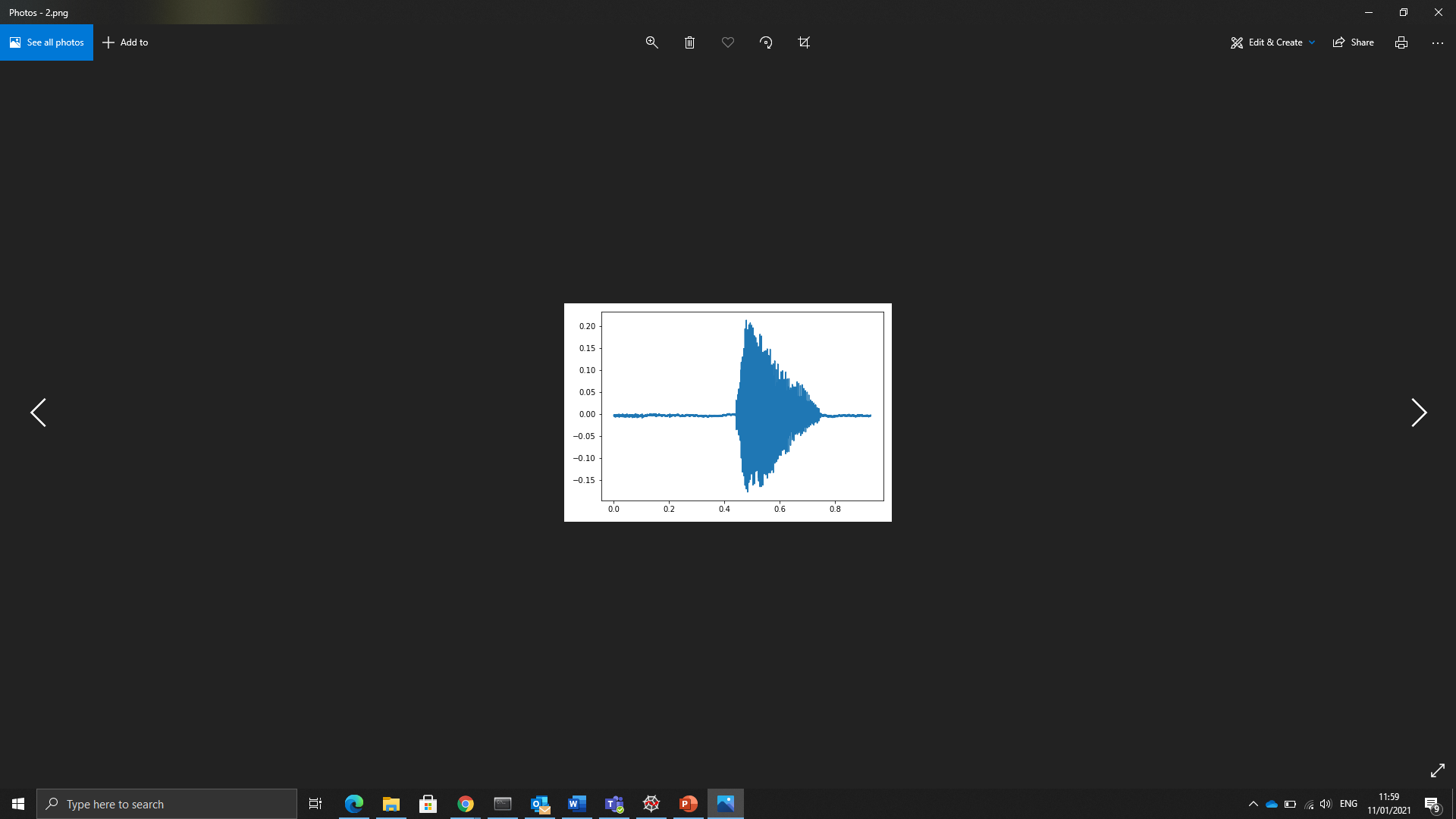
Y axis stands for the signal's amplitude

Chart, histogram

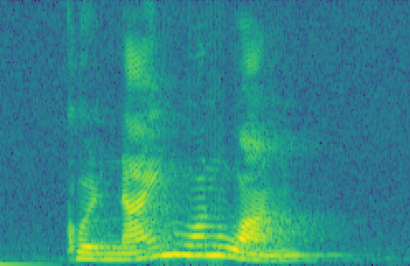
Description automatically generatedChart, histogram

Description automatically generated**Call 911 records:**

**Other words records:**

****

1. **Spectrogram images:**

**Call 911 records:**

**Other words records:**

**A picture containing text, green

Description automatically generated**

**Pipeline:**

**Diagram:**

Down Sampling

LPF

Signal {x[n]}

Windowing

Wavelets CNN

Spectrogram

1D CNN

2D CNN

Binary output

Binary output

Binary output

1. **Preprocessing**

* **Windowing:** we cut a window of 2 sec out from the signal. The length of the partial signal is 2\*sample\_rate.
* **LPF:** Each of the partial signals is transmitted in the Low-Pass filter and at the filter output a smoothed noise-reduced signal is obtained.
* **Down sampling:** Decimation down sampling

1. **Classification approaches:**

* **1 dimensional CNN:** the classifier input is 1 dimensional signal -
* **2 dimensional CNN:** the classifier input is 2 dimensional spectrogram image
* **Wavelets CNN:** WIP