Satakunnan ammattikorkeakoulu
Satakunta University of Applied Sciences

JUUSO LEHTONEN

Laser-induced breakdown spectroscopy spectra analysis with deep learning

DEGREE PROGRAMME IN ARTIFICIAL INTELLIGENCE
2022

Author Lehtonen Juuso	Type of Publication Bachelor's thesis	Date September 2022
	Number of pages 39	Language of publication: English
Title of publication Laser induced breakdown spectroscopy spectra analysis with deep learning		
Degree Programme Artificial Intelligence – Bachelor of Business Administration		

Abstract

As the prices of equipment suitable for laser-induced breakdown spectroscopy (LIBS) have come down, more research and applications are being implemented with LIBS. One of the improvements that could be made in LIBS is the removal of expertise needed for calibration for analysis of different elements.

As LIBS doesn't need extensive sample preparation and the results are available almost instantly, it could be useful for different kinds of purposes like analyzing the contents of batteries or online quality control of a production line.

Using a neural network to detect the contents of samples would make the analysis more available and easier to use, once the models are trained.

The aim of this work is to find out if neural networks are suitable for analyzing the LIBS spectra and what kind of network would work well in this task.

Keywords
Laser-induced breakdown spectroscopy, Deep learning, Neural network, Spectrometer, Laser

CONTENTS

LIST OF SYMBOLS AND TERMS

LIBS	Laser-induced breakdown spectroscopy
Laser	Light amplification by stimulated emission of radiation
NIST	National Institute of Standards and Technology
Nd:YAG	Neodymium-doped yttrium aluminum garnet, laser gain medium
Q-switch	Used to produce a pulsed output beam from laser
MSE	Mean squared error

1 INTRODUCTION

The first records of laser-induced breakdown spectroscopy (LIBS) are from the early 1960’s , a few years after the invention of lasers. (Cremers & Radziemski, 2013, p.7). As LIBS equipment has continued to advance, making it smaller and more portable, it has become a strong competitor in the field of spectroscopy. Major benefits of LIBS are that it doesn’t require sample preparation and having extremely fast measurement time allows online measuring. Combined with the increase of processing power in computers, advancements in deep learning and reduced equipment costs, LIBS has become more feasible to implement without extensive experience in the field.

In this work we will compare how different configurations of the LIBS equipment affect the predictions of the neural networks and which tested network architectures produce the most accurate predictions and if networks can be trained with theoretical data and transfer learning applied to those networks with the real data.

2 LASER-INDUCED BREAKDOWN SPECTROSCOPY

Laser-induced breakdown spectroscopy (LIBS) uses a pulsed laser to generate a plasma from the sample. Spectral features emitted by the excited species, mostly atoms, are used to obtain information about the samples. It has been used on solid, liquid, gas, slurry and aerosol samples. Some applications have been iron and steel sampling, searching for soil contaminants, analyzing Mars soil and rocks, dating artwork and sampling for toxic substances like anthrax. (Cremers & Radziemski, 2013, xi)

2.1 Laser

The word laser is an acronym that stands for "light amplification by stimulated emission of radiation." Lasers are kind of flashlights in that sense that energy, usually electricity, goes in and light comes out. Lasers can be extremely powerful, but more often than not they are much weaker than a typical flashlight. There are three differences between light from lasers and light from flashlights. First, laser beams are much narrower than flashlight beams. Second, lasers emit coherent light while light emitted by flashlights is incoherent. Third, all the light waves in a laser beam are aligned with each other, while the light waves from a flashlight are arranged randomly. Alignment increases the amplitude of the light wave, i.e. makes it more intense. (Hitz et al., 2012, p. 1)

Majority of LIBS lasers are flashlamp-pumped or diode-pumped Nd:YAG (neodymium-doped yttrium aluminum garnet) lasers. They work by pumping light into the gain medium, in this case the Nd:YAG rod, the photons excite the electrons in the neodymium to a higher state. If the pumping is sufficiently strong, a population inversion is established in which the upper level of the lasing transition is more populated than than the lower level. In this state, photons passing through the laser rod at the same frequency as the lasing transition will experience amplification by inducing decay of some of the ions from the upper to the lower state. This is called stimulated emission. If the rod is surrounded by a resonant cavity composed of

two mirrors in which some of this amplified light is directed back into the rod, significant amplification of light at the wavelength of the lasing transition can be achieved along the optical axis of the cavity resulting in a highly monochromatic and directional beam of light. (Cremers & Radziemski, 2013, p. 71). Figure 1 shows a common configuration for a Q-switched laser.

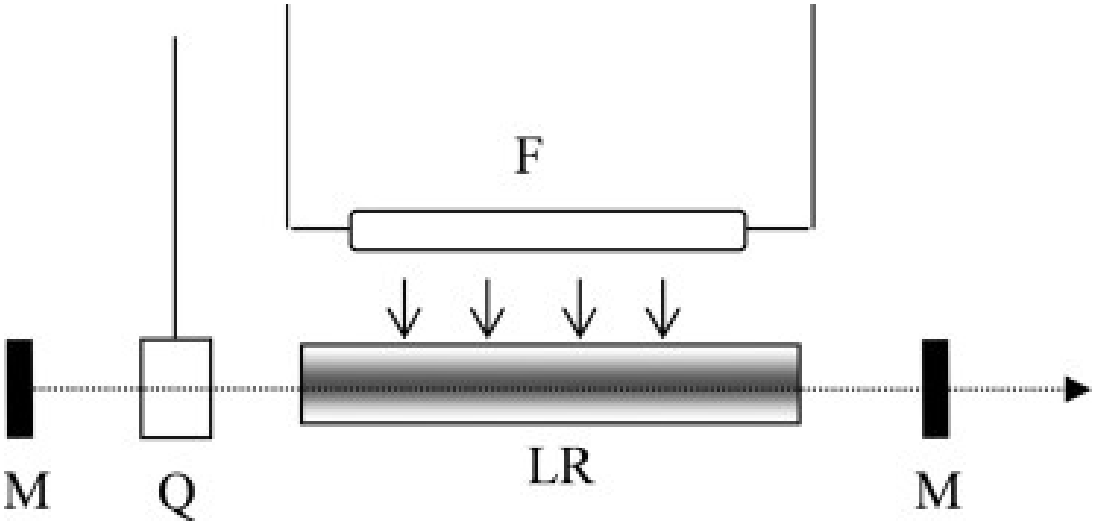

Figure 1: F, flashlamp; LR, laser rod; M, mirror; Q, Q-switch (Cremers & Radziemski, 2013, p. 71)

Different types of lasers use different gain medium. Laser gain media are typically made from partially transparent materials. Such materials could be crystalline or glass solids, organic dyes dissolved in liquid solvents, or various types of gases or gas mixtures. (Silfvast, 2004, p. 25) The gain medium is the main factor which defines the wavelength that the laser produces. The Nd:YAG laser produces a wavelength of 1064nm which is well suited for LIBS, since most of the emission from samples is observed under 1000nm wavelength and the higher wavelength laser doesn't interfere with that.

For LIBS, powerful laser pulses are needed to form the micro-plasma when focused to a small spot. These high powers are easily achieved using a pulsed and Q-switched laser having moderate pulse energies. In this case, an electro-optic Q-switch shutter is positioned in the cavity to prevent photons at the laser wavelength from making a complete path through the cavity and inducing stimulated emission. In this way, the

population inversion between the upper and lower levels of the lasing transition can become very large. When the Q-switch is activated by a suitably timed gate pulse (active Q-switching), the Q-switch becomes transparent, allowing photons to make many traverses of the laser cavity and resulting in a high-power pulse of short duration. A fraction of the pulse energy leaves the cavity through a partially transmitting mirror. For the Nd:YAG laser, the Q-switched pulse length is on the order of 5–10 nanoseconds. The pulse is of short duration because once lasing begins, the population inversion is rapidly depleted and lasing terminates. The Q-switch is intentionally closed shortly after the laser pulse to prevent the generation of additional pulses. (Cremers & Radziemski, 2013, p. 71)

The laser used in our LIBS equipment was a Quantel Falcon (5mj), pictured in figure 2, which is a diode-pumped solid-state Nd:YAG laser that emits light at 1570nm wavelength with a pulse duration less than 6 nanoseconds.

Figure 2: Quantel Falcon laser (Quantel, 2022)

2.2 Plasma

Plasma is the fourth state of matter, alongside solid, liquid and gas. In LIBS, the laser heats a small area of the sample to a high temperature, which causes some of the electrons to be free from atoms. Photons are emitted when the electrons accelerate or decelerate in collisions, this is known as the bremsstrahlung process. When the plasma starts cooling down, the free electrons start recombining with the atoms, which also creates photons, this is known as recombination radiation. (Noll, 2012, p. 167) This radiation is referred as the plasma continuum. The desired photons are created when electrons move from higher energy orbitals to lower ones. Each element emits a characteristic set of discrete wavelengths according to its electronic structure. For LIBS the desired observation period is when the continuum has mostly diminished but the atomic emission is still observable. Figure 3 shows the effect of the laser beam and the continuum on the spectra.

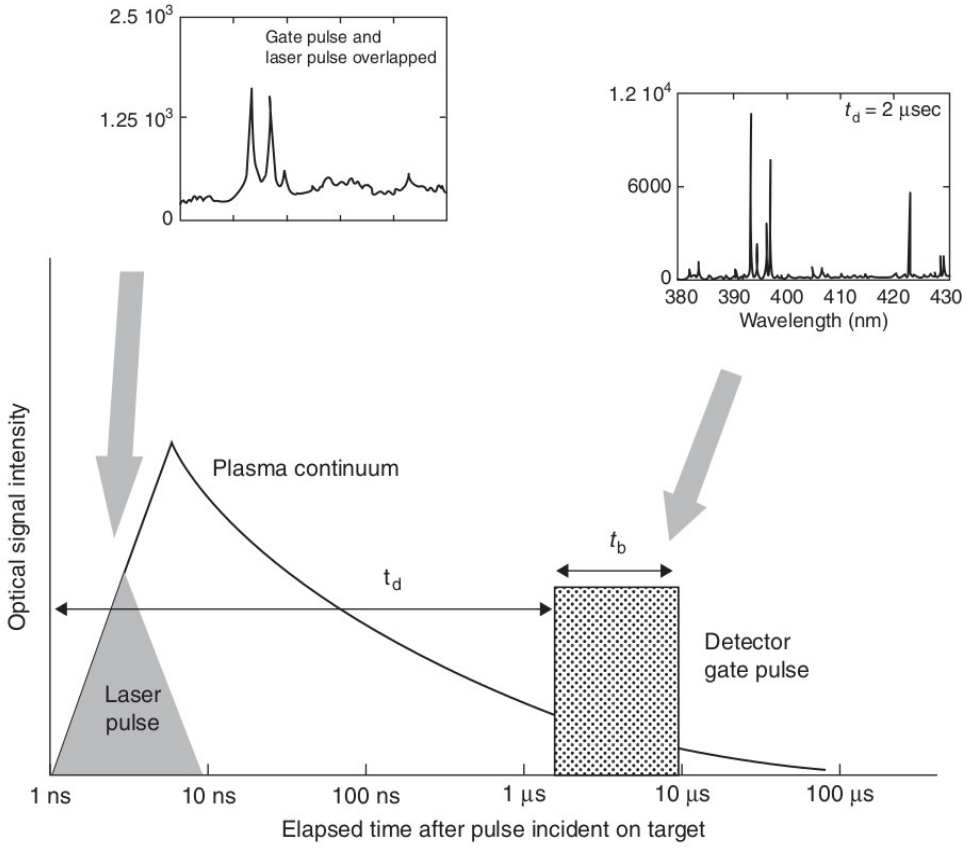

Figure 3: Graphs illustrate how the plasma continuum affects the spectrum (Cremers & Radziemski, 2013, p. 31)

Plasmas are characterized by a variety of parameters, such as the degree of ionization, the plasma temperature, and the electron density. A weakly ionized plasma is one in which the ratio of electrons to other species, i.e. atoms, ions and molecules, is less than 10%. At the other extreme, highly ionized plasmas may have atoms stripped of many of their electrons, resulting in very high electron to atom/ion ratios. LIBS plasmas typically fall in the category of weakly ionized plasmas. (Cremers & Radziemski, 2013, p. 29)

2.3 Atomic emission spectrometry

Atomic emission spectroscopy is a method of chemical analysis that uses the intensity of light emitted from a flame, plasma, arc, or spark at a particular

wavelength to determine the quantity of an element in a sample. In LIBS a laser is used to create the plasma. The basis of any LIBS measurement is the plasma spectrum, shown in figure 4, that contains information about the elements in the target sample. This information is in the form of emission lines located at specific wavelengths, the intensity of the lines, their relative intensities, and sometimes other data such as the line widths or temporal behavior. (Cremers & Radziemski, 2013, p. 151)

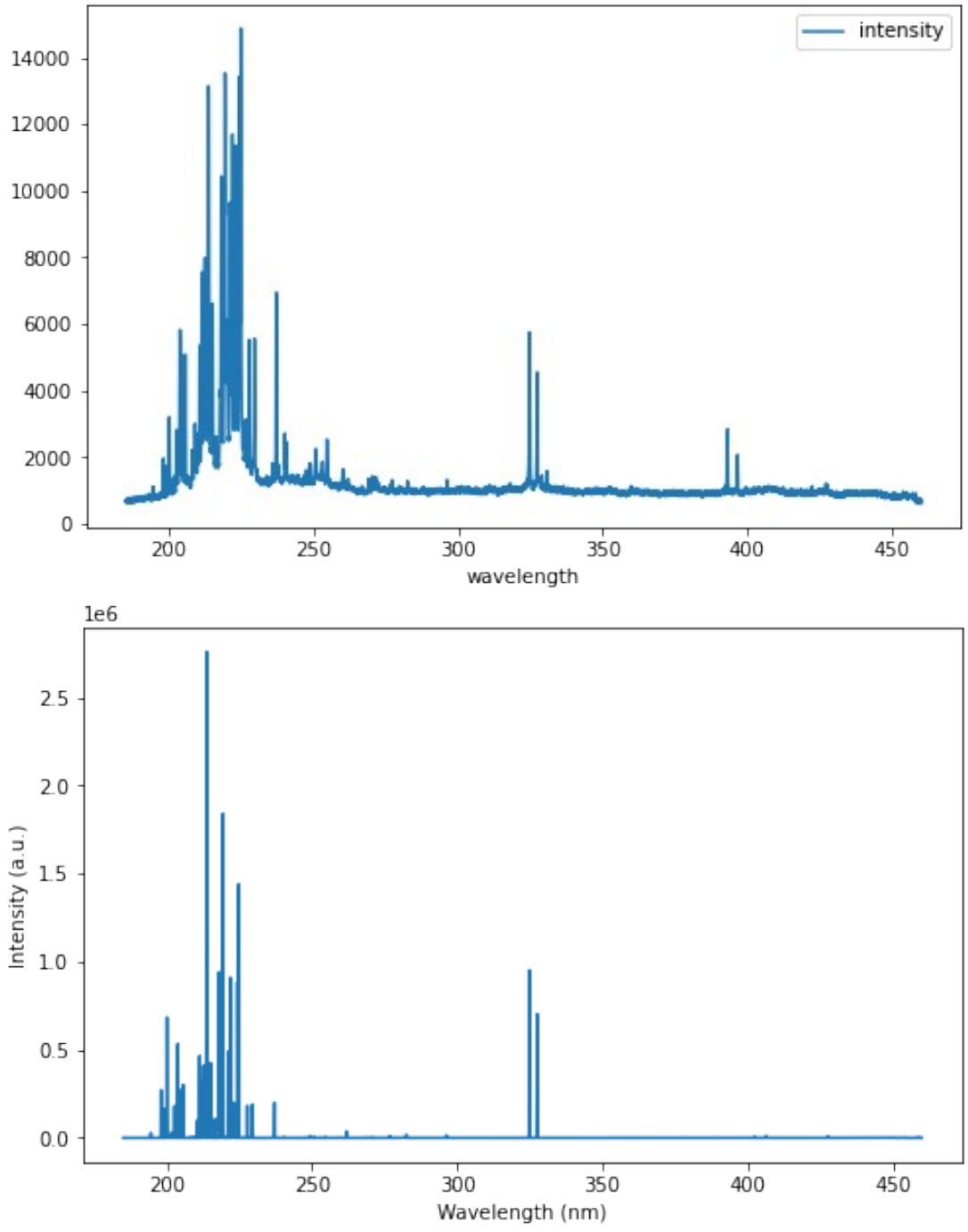

Figure 4: Measured spectra from a copper sample and below a theoretical copper sample

Ideally the relative line intensities of elements are directly proportional to their concentration in the sample. Difficulties in interpreting are caused by the matrix effect, which arises from several elements being present in the sample. The overlapping lines between elements make it challenging to determine how much of each element is present in the sample. Example of this is shown in figure 5.

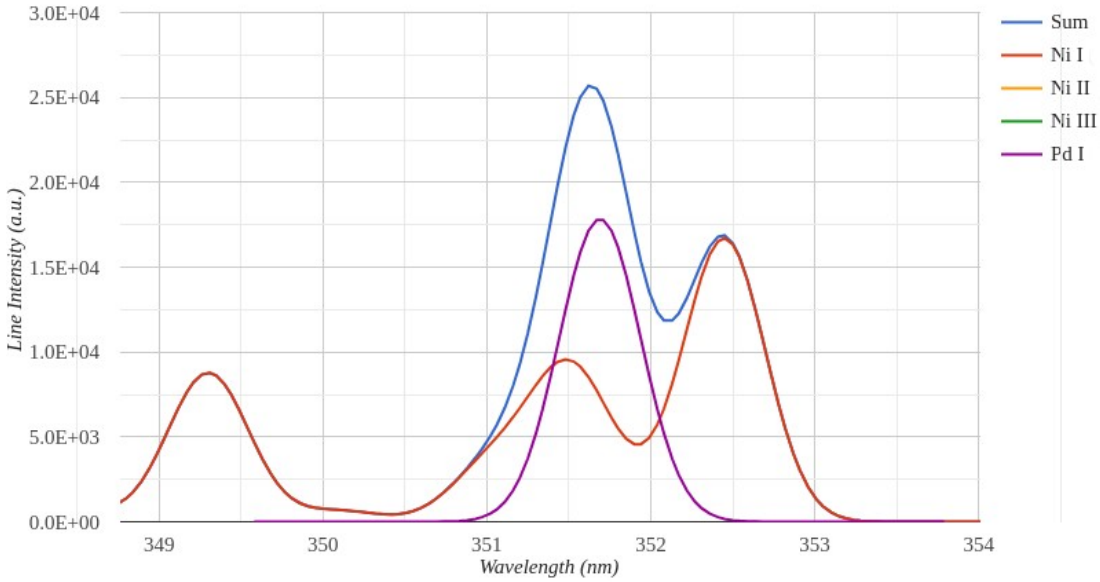

Figure 5: Nickel and palladium lines overlapping (NIST, 2022)

The lines are measured with a spectrometer. Important technical specifications in a spectrometer for LIBS include wavelength range and resolution. Wavelength range should be chosen depending on which elements are to be studied. Most intense emission lines are observed between 150-1000nm. A smaller resolution helps distinguish between peaks, which often are close to each other.

The spectrometer used in our experiments is a StarLine AvaSpec-ULS4096CL-EVO, shown in figure 6, from Avantes with wavelength range 185nm-460nm and a resolution of 0.067nm. A fiber optic cable connected to a lens is used to collect the light from the plasma. It outputs values between 0 and 65535 on 4096 channels. The spectrometer uses Hamamatsu S13496 CMOS linear image sensor. The photodiodes of the CMOS sensor accumulate electrical charge when exposed to light, and those charges are then converted to voltage, amplified and transmitted as electrical signals. Figure 7 shows the path light travels inside the spectrometer.

Figure 6: StarLine AvaSpec-ULS4096CL-EVO (Avantes, 2022)

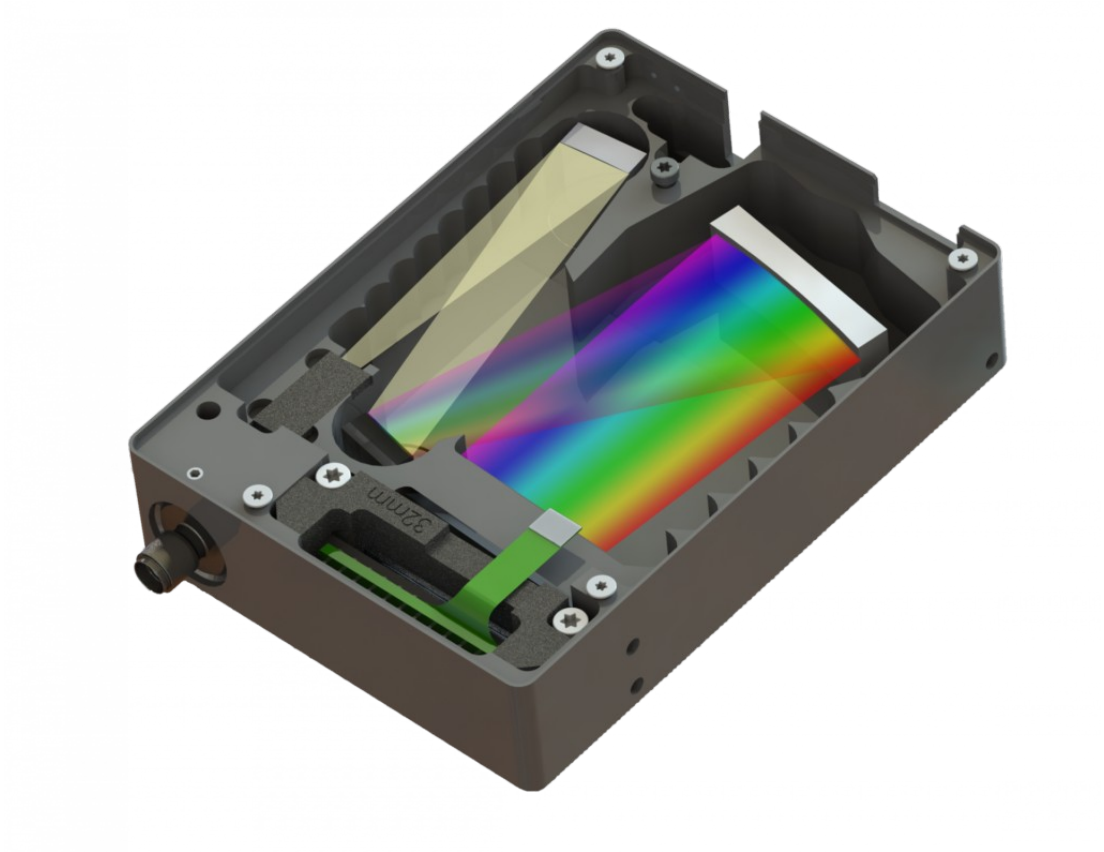

Figure 7: Light enters the optical bench through a standard SMA-905 connector and is collimated by a spherical mirror. A plain grating diffracts the collimated light after which a second spherical mirror focuses the resulting diffracted light. An image of the spectrum is projected onto a 1-dimensional linear detector array. (Avantes, 2022)

3 DEEP LEARNING

3.1 What is deep learning?

Throughout its history deep learning has been known by many names. Broadly speaking, there have been three waves of development of deep learning: deep learning known as cybernetics in the 1940s–1960s, deep learning known as connectionism in the 1980s–1990s, and the current resurgence under the name deep learning beginning in 2006. Some of the earliest learning algorithms we recognize today were intended to be computational models of biological learning, i.e. models of how learning happens or could happen in the brain. As a result, one of the names that deep learning has gone by is artificial neural networks (ANNs). (Goodfellow et al, 2016, p. 13). Figure 8 shows how deep learning and AI connect.

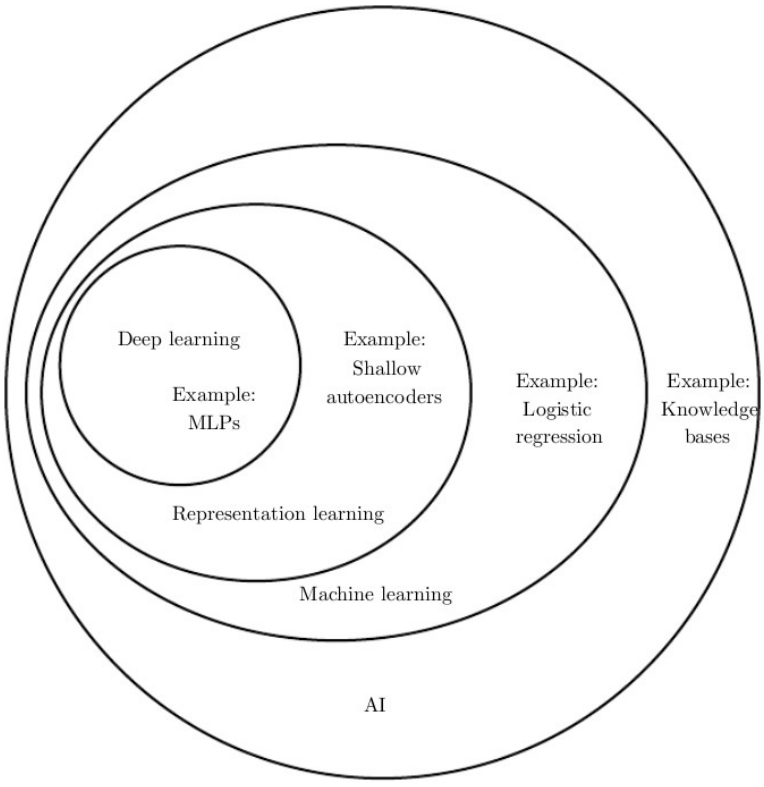

Figure 8: A Venn diagram showing how deep learning is a kind of representation learning, which is in turn a kind of machine learning, which is used for many but not all approaches to AI. Each section of the Venn diagram includes an example of an AI technology. (Goodfellow et al, 2016, p. 9)

Extremely simplified way of looking at neural networks is that numbers go in and numbers come out. They consist of input layer, output layer and in between are hidden layers. Networks with only one hidden layer are called shallow neural networks whereas with multiple hidden layers they are called deep neural networks. There are different kinds of layers that can be used. One of the most common layers are dense layers which consist of nodes also known as neurons. Dense layer nodes have their own weights and bias. When they receive an input, that value is multiplied by the weight and the bias is added. In dense neural network each node is connected to every node on the previous and the next layer. A simple neural network is shown in figure 9.

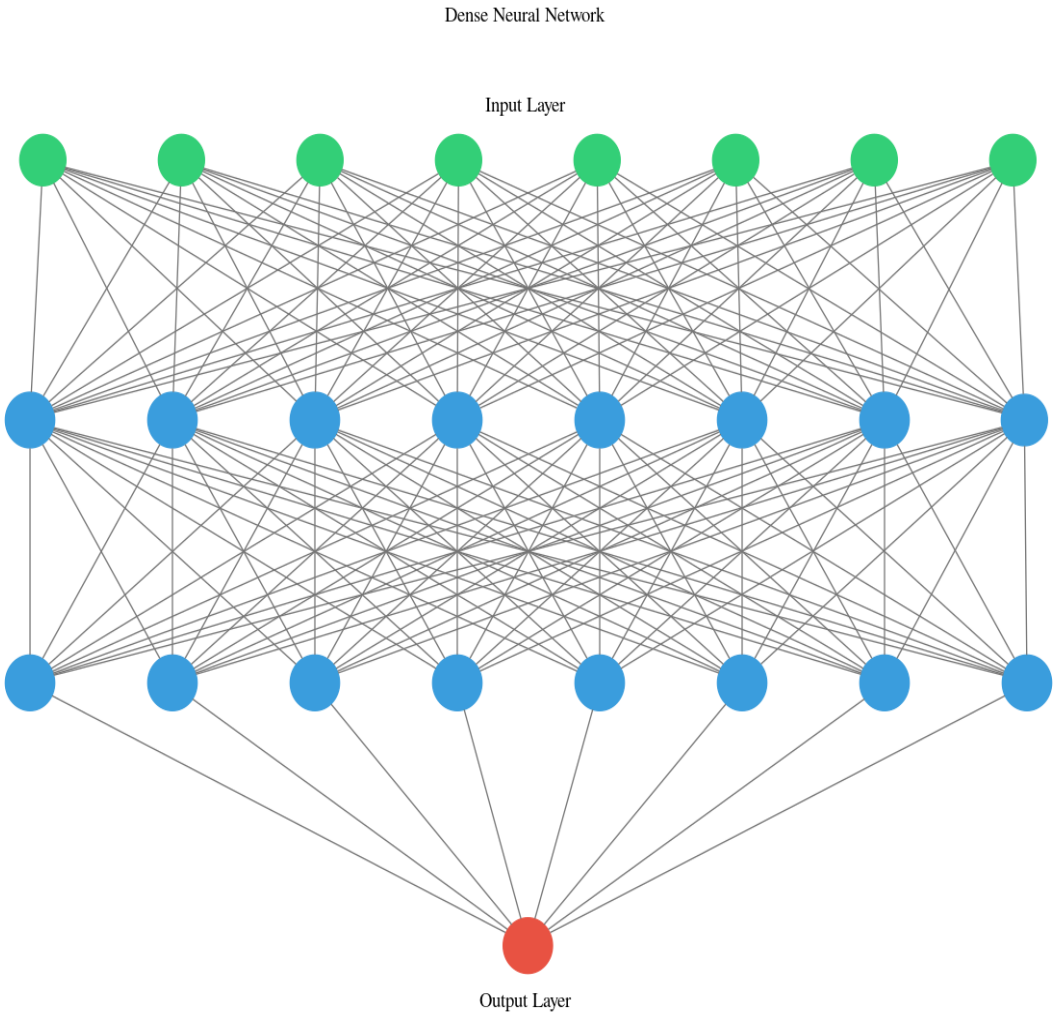

Figure 9: Dense neural network with two hidden layers. Each hidden layer is made up of 8 nodes. Input layer has 8 nodes and output layer is only a single node.

In LIBS the input would consist of the values from the spectrometer, although some sort of data pre-processing can be made before feeding the data to the neural network. Also other information could be given to the neural network if it is available, e.g. laser pulse power or some information about the sample.

3.2 Data

To get started with deep learning we always need data. In LIBS the main source of data is the spectrometer, although we can also generate data with theoretical values. National Institute of Standards and Technology (NIST) provides a database and a service which can calculate the theoretical spectra for most of the elements in the periodic table. For deep learning the more data there is the better, provided that the data is of good quality. Because LIBS is measuring a phenomenon based on physics, the possible errors in data come from the measuring process, the setup and the equipment used.

The NIST data used is collected with regular intervals of approximately 0.067nm, but the spectrometer produces data with varying intervals. For the lowest channels the interval is 0.073nm, the highest channels 0.060nm and in the middle it's 0.067nm. In the middle channels the wavelength for NIST data is 322.575nm and spectrometer is 329.322nm, a 6.747nm difference. Figure 10 shows the intervals between spectrometer channels. The NIST intensity data, shown in figure 11, is in arbitrary units, meaning that the values itself are not as important as their relative intensity to other values. The maximum intensity value in the data was 5798156264, almost 6 billion. Spectrometer produces data in 0-65535 range. In spectrometer data noise is present and without corrections every channel usually shows a value of at least 900, while the NIST data with few elements present has 0 for most channels. Figure 12 shows the spectrometer data.

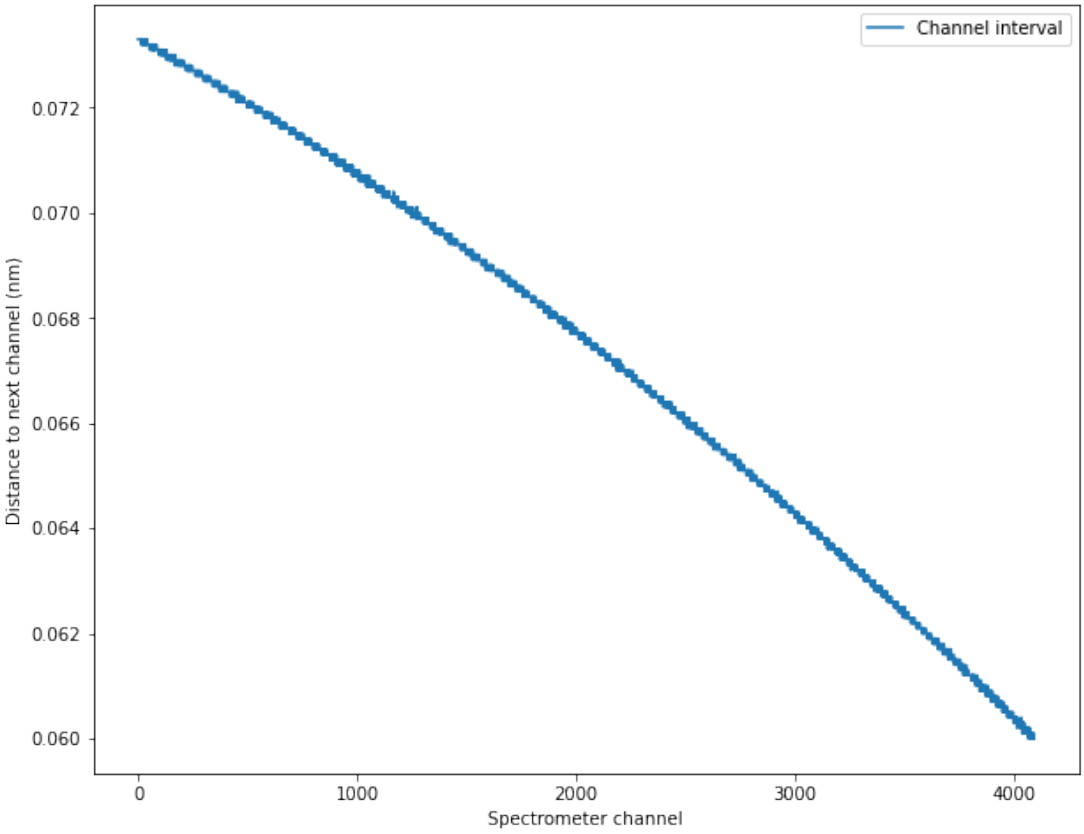

Figure 10: Channels of the spectrometer are not evenly spaced

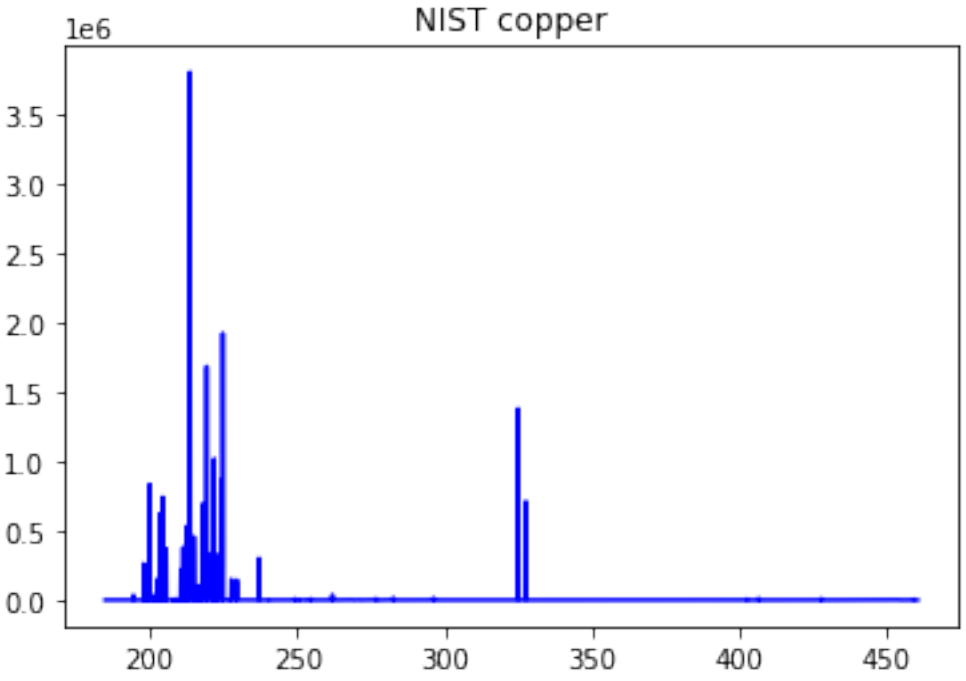

Figure 11: Clean spectrum of copper from NIST data

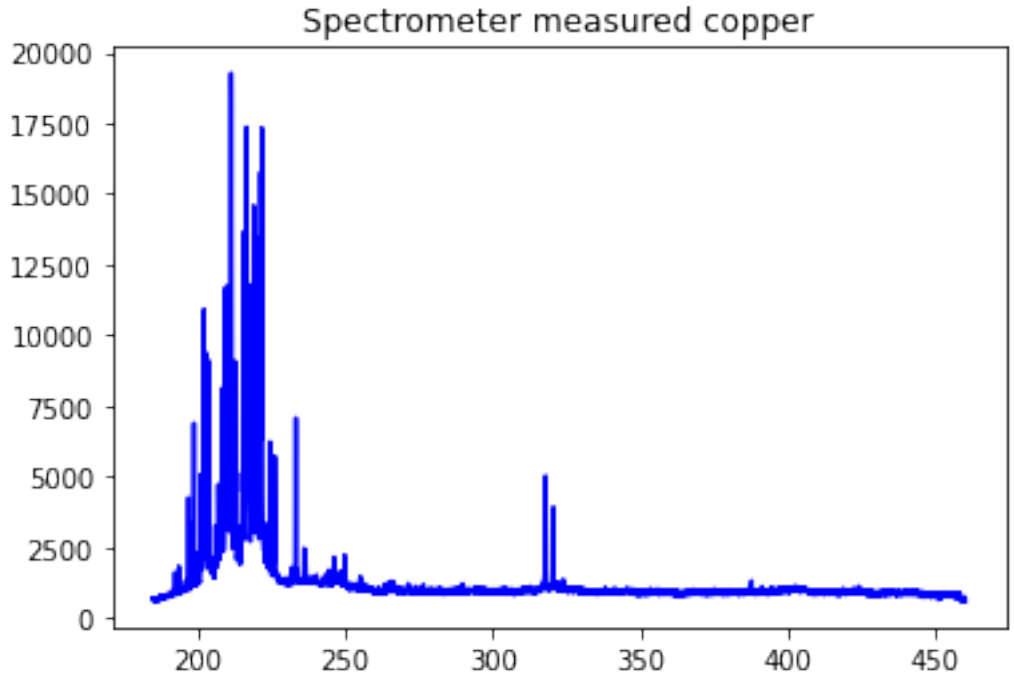

Figure 12: 99,9% copper measured with SAMK LIBS equipment

The spectrum also changes depending on electron temperature and electron density. The NIST data was collected with electron temperature of 1.0 electronvolt and electron density (cm^{-3}) of 1e17. At the time of writing the LIBS project at SAMK hasn't looked into measuring the electron temperature or density of observed plasma.

The spectrometer data was gathered by shooting five consecutive shots at one point then moving the sample by some tens to hundreds of microns. The multiple shots are done so that the first shots clean the surface of the sample from impurities.

3.2.1 Data preprocessing

As the NIST data is theoretical data based on calculations it is perfect data. The challenge is making it as close to the spectrometer data as possible to make it more useful for transfer learning. Three different preprocessing techniques were used to make the NIST data more similar to the real data, added noise, clipped high values and scaling.

First the NIST data was clipped at value 1000000, 1 million, since some of the elements have very high maximum intensities compared to the other elements and even if those elements were encountered in samples, the channels with the high

intensities would over saturate and stop at maximum value of 65535, making them have no further relevant information.

Secondly noise was added. Since 1 million was now the maximum value for the NIST data and 65535 for spectrometer and the spectrometer values were showing approximately 600 for channels with no peaks the similar noise for the clipped NIST data would be around 9000. For each channel with values less than 10000 in the NIST data a random number between 1 and 9000 was added.

Thirdly the NIST data was scaled between 0 and 1. Scaling data often helps the performance of several algorithms.

Also shifting the NIST data wavelengths to match the spectrometer was tried. It was done by interpolating the intensities based on the wavelengths. 1000 models were trained for each data, with the shifting and without it. Without shifting the average MSE was 6.805989e-4 while the shifted one was 6.787509e-4, a 1.848e-6 difference. The difference is minuscule and such difference could come from random initialization of the weights.

3.2.2 Data labeling

The NIST data was collected with in an array of numbers in which each index corresponded to a certain element. The percentages were added as numbers between zero and one. Same kind of labeling was used for the spectrometer data. Earliest tests used a network which tried to predict several elements at once, but this proved to be challenging and the network was changed to predict a single element. For this the labels would be cut to a single element so that the label would be just a single number between zero and one, i.e. earlier the labels were [0.4, 0.5, 0.1] if the sample contained 40% copper, 50, aluminum and 1% zinc, but now the label would be just [0.4] for the copper and for each element own network would be trained.

3.3 Transfer learning

Transfer learning refers to the situation where what has been learned in one setting (i.e., distribution P 1) is exploited to improve generalization in another setting (say distribution P 2) (Goodfellow et al, 2016, p. 536). Transfer learning is often used e.g. in object detection where a network is trained with hundreds of thousand or more images with some classes for example different kinds of animals or objects like cars, airplanes etc. and then that network is trained with other kinds of classes, e.g. different kinds of fruits. With this kind of training better results can be achieved with fewer images. Transfer learning was tested on LIBS data and it was found to be useful.

3.4 Error metric

For calculating the error mean squared error (MSE) was chosen.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

MSE is the average squared difference between the observed and predicted values. Since the difference is squared bigger errors are penalized more than smaller errors. When we take the root of our MSE we get the average error in percentage points because our labels are provided as percentages.

3.5 Network architecture

Several different network architectures were tried. First tests were done trying to predict the percentage of all elements in the sample. This did not produce good results so it was changed so that the network would only predict a single element and then for every element a network would be trained and the predictions combined.

Initial tests for single elements were made on networks that were deeper but narrower. For all hidden dense layers a L2 kernel regularizer was used to reduce overfitting. All hidden layers used tanh activation function. The default weight

initializer was used, which is glorot uniform, it produces values between -1 and 1. Adam optimizer was used for all networks to reduce the time it takes to tune parameters as finding a good architecture and training parameters takes considerable time already (Adam, 2017). A few tests were done with the adam learning rate and lowering the initial learning rate seemed useful, although not much testing was done on that since adam updates any parameter with an individual learning rate. This means that every parameter in the network has a specific learning rate associated. But the single learning rate for each parameter is computed using lambda (the initial learning rate) as an upper limit. In earlier tests dropout layers were used but were later dropped.

First network tests were done with 6 hidden layer of which second was a dropout layer with 10% value. A dropout layer randomly drops connections by a percentage provided. For the second tests another dropout layer was added between third and fourth dense layers with 5% drop rate. In third network third dropout layer was added and the dropout layers were placed between first and second, second and third and finally third and fourth dense layers with rates of 5%, 7% and 5%. Several tests were done with the third configuration with different epoch, batch and neuron settings. Each dense layer had the same amount of neurons. The output layer was a single neuron dense layer with linear activation for every network tested.

Later a test with reinforcement learning algorithm Advantage Actor Critic (A2C) to find a better architecture. Although it was utilized more as a random search it yielded results. The algorithm preferred networks without dropout layers and wider rather than deeper networks. After couple days of searching for a network, the best result was with a network of just 2 hidden layers, with 16384 neurons in the first and 128 neurons in the second.

3.6 Training

The datasets were initially split to training and validation sets with 90% for training and 10% for validation. Later when spectrometer data was available, the NIST data was split 99% for training and 1% for validation and the spectrometer data was split

90% training and 10% validation. This was done so that less information would be lost on the NIST data for transfer learning. The data was randomly split with a set random seed so that the same data would always be used for training/validation.

3.6.1 Epoch

A one epoch occurs when all of the dataset is being passed through the neural network. As the weights are updated gradually, more epochs mean a better fit for the data. The risk with too many epochs is overfitting, which means that the model will not generalize to unseen data.

3.6.2 Batch

Batch size determines how many samples of the dataset we pass through the network at a time. A larger batch size means faster training, but sometimes that is limited by the hardware. Also if the batch size is very large that might degrade our network since the weights are usually upgraded after each batch is passed through the network.

3.7 Datasets

Datasets in this project are divided into the NIST datasets and the spectrometer dataset.

3.7.1 NIST datasets

The first NIST dataset had a total of 14960 samples. It consisted of 15 elements, carbon, magnesium, aluminum, silicon, titanium, vanadium, iron, cobalt, nickel, copper, manganese, lithium, zinc, oxygen and phosphorus. Samples consisted of one to three element combinations. With single element samples of 100%, two-element

samples from 1% to 99% in 1% steps and three-element combinations of [33, 34, 33], [40, 20, 40], [20, 40, 40], [40, 40, 20], [60, 20, 20], [20, 60, 20], [20, 20, 60], [80, 10, 10], [10, 80, 10], [10, 10, 80].

The second NIST dataset consists of 24 different elements, carbon, magnesium, aluminum, silicon, titanium, vanadium, iron, cobalt, nickel, copper, manganese, lithium, zinc, oxygen, phosphorus, platinum, gold, silver, iridium, tungsten, palladium, tin, technetium and yttrium. Altogether 53660 NIST data samples was collected and used. They consist of combinations of 1 to 3 different elements. The different sample compositions were a single element 100% samples, two-element samples increasing from 1% to 99% with 1% steps and three-element samples with the following combinations [33, 34, 33], [40, 20, 40], [20, 40, 40], [40, 40, 20], [60, 20, 20], [20, 60, 20], [20, 20, 60], [80, 10, 10], [10, 80, 10], [10, 10, 80], [15, 30, 55], [30, 55, 15], [55, 15, 30].

3.7.2 Spectrometer dataset

The spectrometer dataset used consist of third shots of the samples. Most samples are from sheets of copper which had between 99.995% and 99.96% copper in them, the rest was unknown. Other samples included sheets of brass with 85.0% copper and rest zinc, a zinc bars with 63.10%-63.60% copper and 35.40%-35.80% zinc and the rest a combination of silicon, phosphorous, lead and bismuth. Also 99.99% reference samples of carbon, magnesium, aluminum, silicon, titanium, vanadium, iron, cobalt, nickel, copper, zinc and tin were included. The samples are shown in figures 13 and 14.

Figure 13: Copper and brass samples

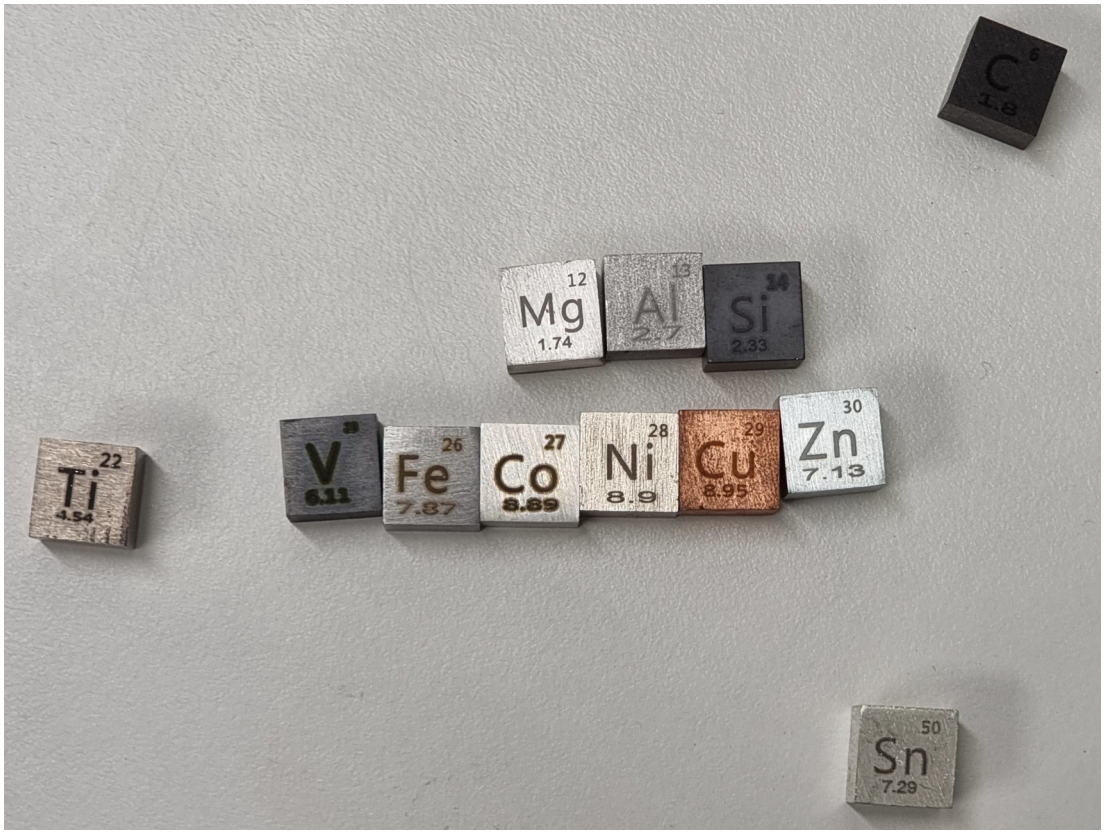

Figure 14: 99.99% reference samples

4 RESULTS

4.1 First network tests

For training the networks different numbers of epochs, batch sizes and neurons were used. Because of the random initialization of the weights, each network was trained 100 times and the average MSE was calculated. A combination of [2, 4, 6, 8, 10] epochs, [16, 64, 512, 1024] batch sizes and [64, 256, 512, 1024, 2048] neurons. The lowest MSE of 0.005047 was achieved with the largest values of 10 epochs, 1024 batch size and 2048 neurons. Initial learning rate for the adam optimizer was 1e-4. Data that was used was only NIST data as data from spectrometer wasn't available at the time. The first NIST dataset was used.

Table 1: Top 10 results of the first network tests

Epochs	Batch size	Neurons	MSE
10	1024	2048	0.0050479
8	1024	2048	0.0052536
6	512	2048	0.0054870
8	512	2048	0.0056326
4	512	2048	0.0058435
10	512	2048	0.0060021
6	1024	2048	0.0061080
10	1024	1024	0.0075520
8	1024	1024	0.0076527
6	512	1024	0.0081316

From these tests it was clear that a larger network with more training would provide better results.

4.2 Second network tests

For the second network the number of epochs, batch and neurons were all increased based on the results of the first tests. The combination was [8, 9, 10, 11, 12, 13, 14, 15] for epochs, [128, 256, 512, 1024, 2048] for batch size and [1024, 2048, 4096] for neurons. Initial learning rate was the same as in the first tests, 1e-4. The first NIST dataset was used.

Table 2: Top 10 results of the second network tests

Epochs	Batch size	Neurons	MSE
15	1024	4096	0.0017068
11	512	4096	0.0017098
10	512	4096	0.0017173
9	512	4096	0.0017249
8	512	4096	0.0017372
14	1024	4096	0.0017431
12	512	4096	0.0017926
12	1024	4096	0.0018023
13	1024	4096	0.0018185
14	512	4096	0.0018369

It seems that the main factor in performance is the size of the network, the bigger the better.

4.3 Third network tests

For the third network tests the batch sizes were kept the same as before but number of neurons and epochs were increased. For neurons the options were [3500, 4096, 5120] and the epochs [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21]. Initial learning rate was 1e-4. The first NIST dataset was used.

Table 3: Top 10 results of the third network tests

Epochs	Batch size	Neurons	MSE
10	512	5120	0.0013283
20	1024	5120	0.0013343
19	1024	5120	0.0013525
21	1024	5120	0.0013600
18	1024	5120	0.0013684
17	1024	5120	0.0013770
11	512	5120	0.0013885
16	1024	5120	0.0014000
12	512	5120	0.0014078
14	512	5120	0.0014525

4.4 Fourth network tests

For the fourth tests the batch sizes remained the same but the number of epochs were changed to [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21] and the neurons were tested only with 8192. Initial learning rate was 1e-4. The second NIST dataset was used which most likely affected the observed MSE.

Table 4: Top 10 results of the fourth network tests

Epochs	Batch size	Neurons	MSE
15	512	8192	6.125e-4
14	512	8192	6.208e-4
12	512	8192	6.342e-4
17	512	8192	6.390e-4
9	256	8192	6.418e-4
8	256	8192	6.447e-4
19	512	8192	6.454e-4
7	256	8192	6.465e-4
16	512	8192	6.479e-4
13	512	8192	6.551e-4

4.5 Fifth network tests

The fifth network tests were done with a different kind of network which consisted only of 2 hidden layers of 16384 and 128 neurons. Different batch sizes and epochs were not tested. Instead the effects of different kinds of preprocessing on the data was tested and additional tests with another test dataset were conducted. Each test was ran 100 times and the average MSE recorded. Tests were done so that first a network was trained with NIST data and tested with NIST data and the spectrometer data. Then that model was retrained with spectrometer data, i.e. transfer learning as used, and tested with spectrometer data validation set and additional copper data also gathered with the spectrometer. The preprocessing done on NIST data included clipping the highest values, adding noise and scaling the data between 0 and 1. No preprocessing was done on the spectrometer data. Initial learning rate was 5e-5.

Table 5: Effects of different preprocessing techniques on MSE

Preprocessing	NIST trained NIST tested	NIST trained spectrometer data tested	Transfer learning	Transfer learning copper tested
No preprocessing	0.0074788	0.5687736	2.478e-4	3.812e-4
Clip	0.0063265	0.5567484	6.291e-4	6.820e-4
Clip + noise	0.0010603	0.6233914	8.16e-5	2.480e-4
Clip + noise + scaling	**1.219e-4**	**0.1264416**	**4.08e-5**	**1.619e-4**

The best results in each of the tests were achieved with including all of the three preprocessing techniques. Also training a network with only spectrometer data was tested which resulted in MSE of 1.047e-4 with the validation set and 3.417e-4 with the copper test set. From this we can conclude that transfer learning clearly helps the network to learn.

Although most predictions are very close to correct, some are far off as can be seen from tables 6 and 7.

Table 6: Bad predictions from the network

Predicted	Label	Prediction – Label
0.081178136	0.999	-0.9178
0.012133906	0.999	-0.9869
-0.0107268	0.999	-1.0097
-0.014411053	0.999	-1.0134
0.0063202777	0.999	-0.9927
-0.017494997	0.999	-1.0165
0.06240226	0.999	-0.9366
0.06865663	0.999	-0.9303
0.0061571696	0.999	-0.9928
0.0462044	0.999	-0.9528

Table 7: Better predictions from the network

Predicted	Label	Label – prediction
0.997506	0.999	-0.0015
0.8655777	0.85	0.0156
0.9973346	0.999	-0.0017
0.9941029	0.999	-0.0049
-0.015855422	0.0	-0.0159
0.61475134	0.631	-0.0162
0.0017584956	0.0	0.0018
0.9904094	0.999	-0.0086
0.99488956	0.999	-0.0041

A closer look at the data shows that some of the spectra are very different from each other even if they are supposed to be almost the same. In figure 15 a large difference in intensities between samples can be seen.

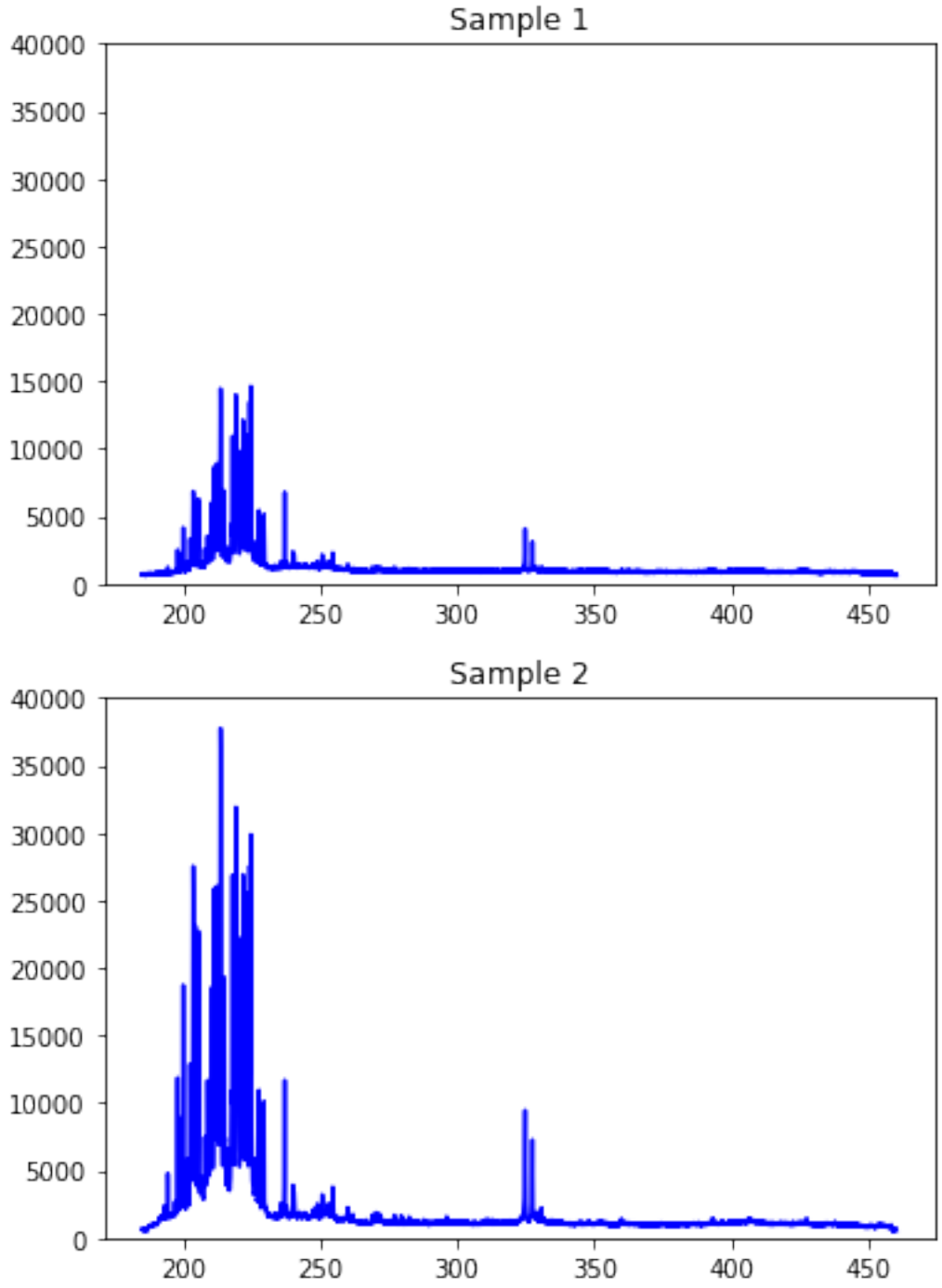

Figure 15: Two >99,9% copper samples

4.6 Additional test

Another test was performed to see if data could be created by combining two >99.9% samples. The test was done by combining measured copper and zinc samples. 50 samples of each were combined while multiplying them by a factor and adding them together. Each of the samples were multiplied with factors of 0.1 to 0.9 with 0.1 steps, i.e. the copper sample would be multiplied by 0.4 while the zinc sample was multiplied with 0.6. The different copper and zinc samples were not mixed between different factors.

This test provided worse results than without the mixed samples. With the samples the MSE was 1.412-04 and without them 9.540e-05. The test was run a few times and every time the results were similar, without the added samples the network performed better.

It could be that the samples used for these tests were not optimal and more carefully selected samples could provide better results. Also mixing the samples with more variety could be useful. It was also hypothesized that different compositions of the elements would result in different electron temperatures, which would affect the measured spectra.

5 LIBS EQUIPMENT

The LIBS equipment at SAMK, shown in figures 16 and 17, consist of the laser, spectrometer, computer for controlling the equipment, surrounding box for cover, two linear motors for moving the samples in two axes and a webcam inside the box used for setting the sample to correct position. Also optical tube with two lenses to focus the laser beam are used and another lens to collect the light from the plasma to the spectrometer via an optical fiber cable.

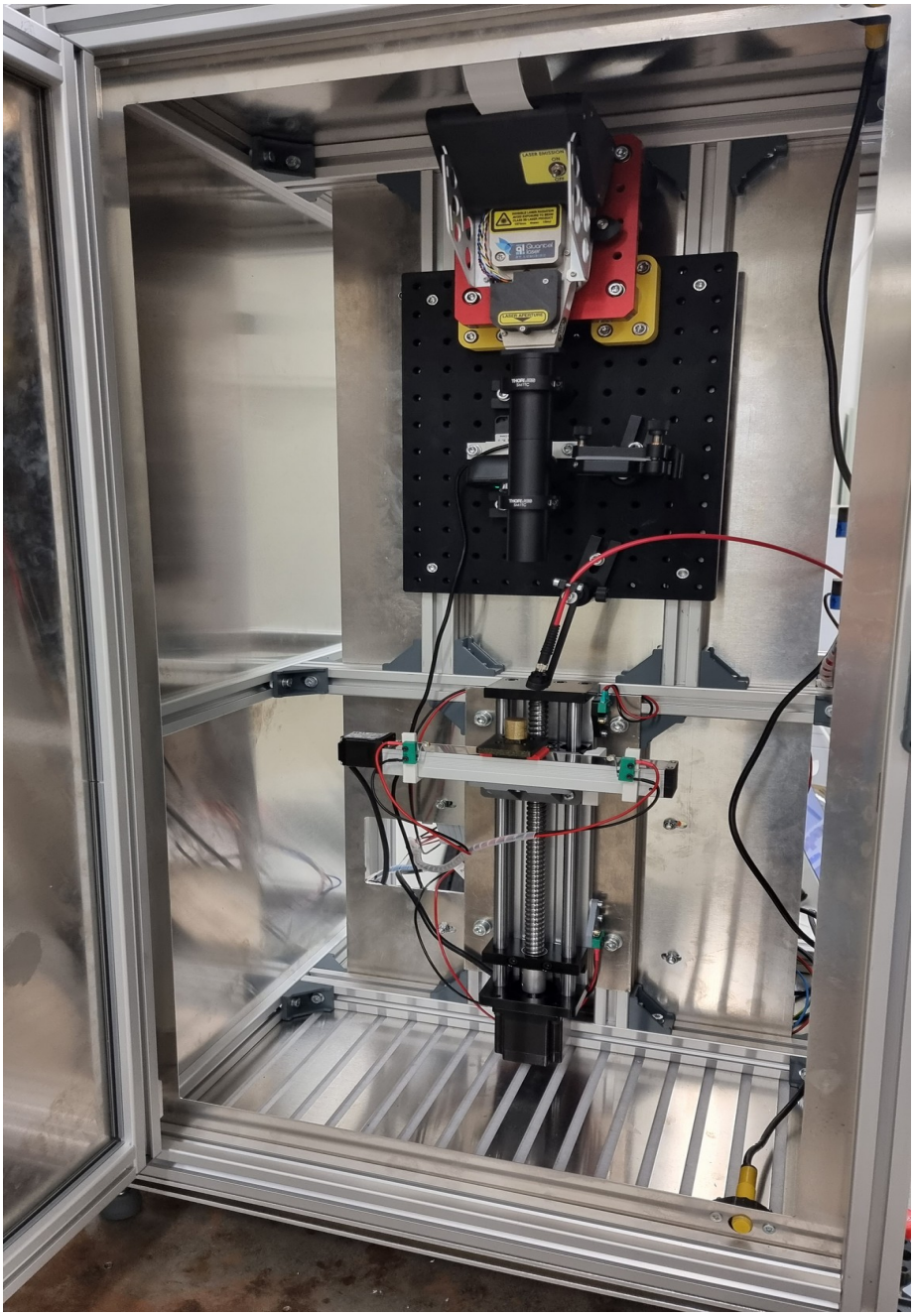

Figure 16: Inside of the box with the laser, optical tube, spectrometer lens and the two linear motors visible

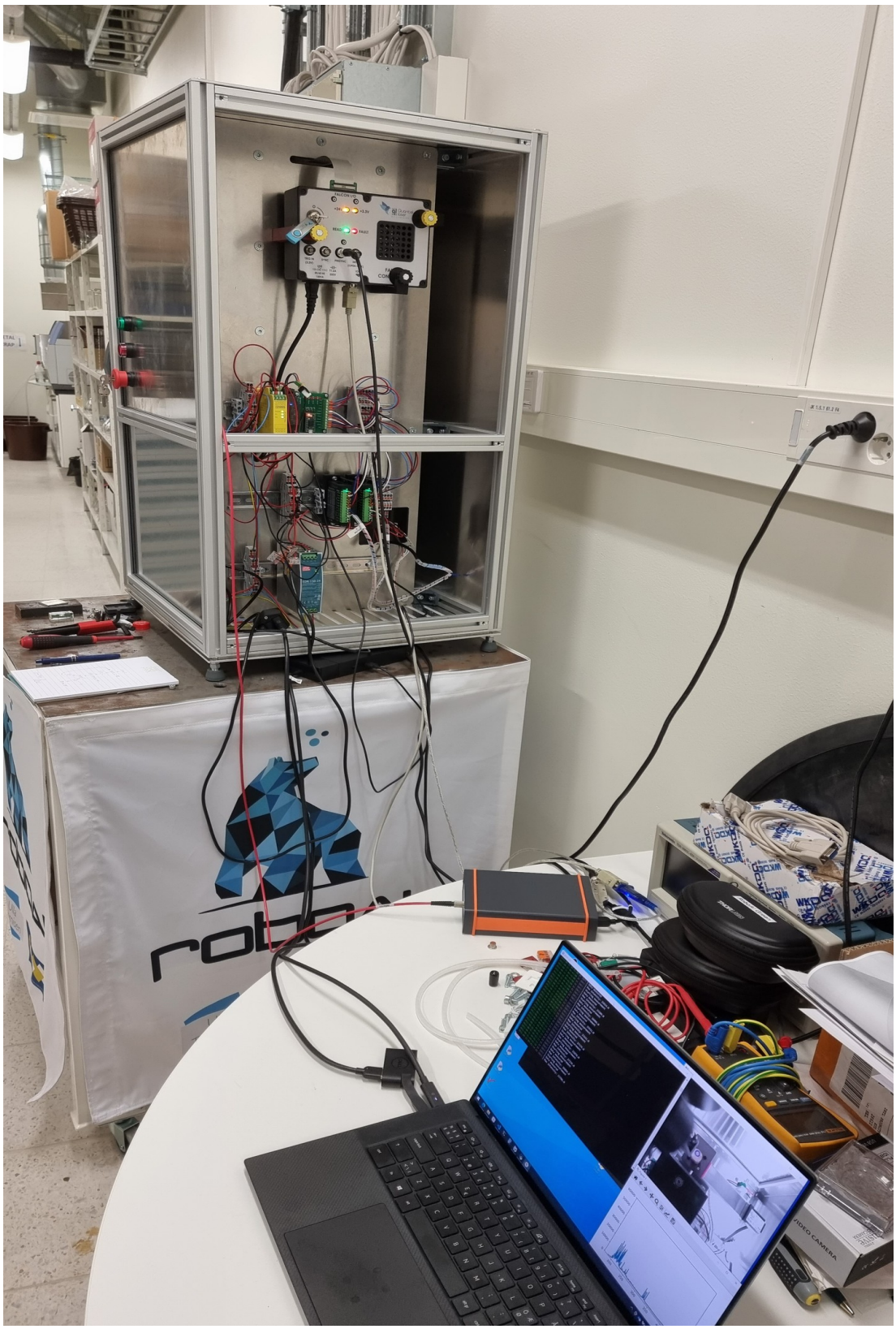

Figure 17: Back of the box with the spectrometer and computer visible

6 FUTURE WORK

At the time of writing the available samples are limited with most samples being close to 100% of one element and some samples with 85% copper,15% zinc and 63% copper and 35% zinc. The accuracy of the models would most likely improve with more diverse samples and the testing of the models would be more reliable if the testing set included samples with contents not included in the training set.

Currently the positioning of the sample is done by taking measurements of the sample and moving along the vertical axis recording the maximum intensity, figure 18, and when the intensity starts to lower going back to the position with the highest recorded intensity. This is time consuming and the maximum intensities also fluctuate as seen in the image below. That leads to measurements happening from different distances which might affect results.

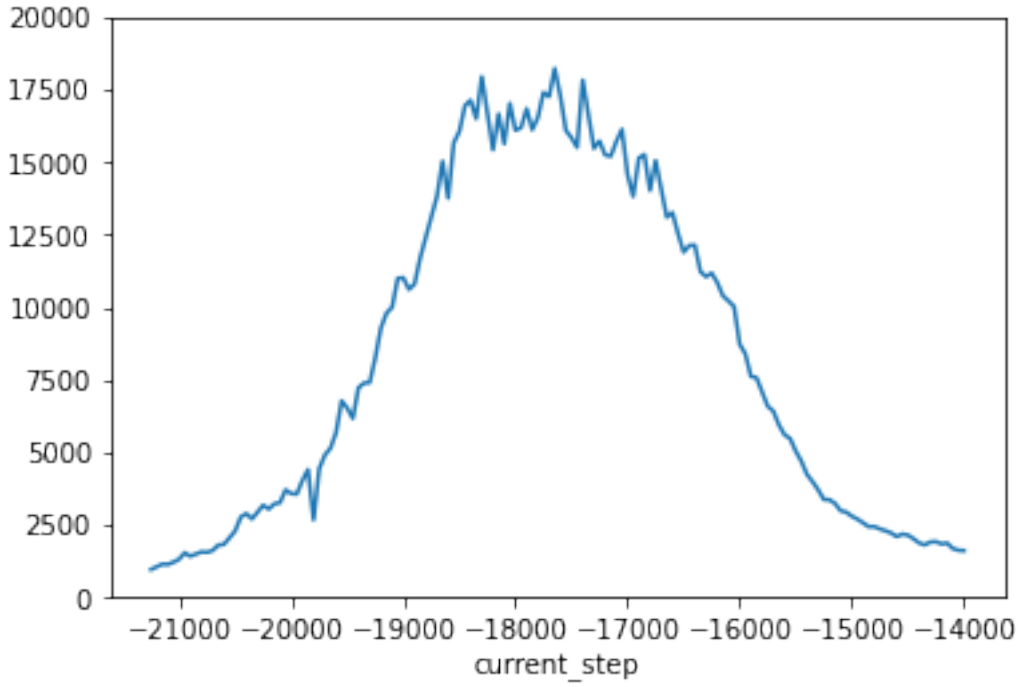

Figure 18: Y-axis: the maximum intensity of the spectrum. X-axis: the number of steps from the highest position

Some sort of distance measuring device will be added in the future so the measurements will always happen from the same distance.

One possibility is to also test networks with different kinds of layers like one dimensional convolutional neural network and to try even wider or deeper dense networks. Also combining the networks that are trained to predict the percentage of a single element to a single network and see if that improves the predictions.

More work on finding out if the networks generalize well to unseen data should be made. At the moment it is not clear if the networks recognize the important relativities between peaks. The results do suggest that the tested networks might focus on the absolute intensities instead.

7 CONCLUSION

Analyzing the spectra with neural networks could be viable and so far the results suggest predictions within 1% to be possible. This method requires taking a lot of samples of the elements that are to be predicted but once that is done no expertise on LIBS is required to use the models. The tests performed also suggest that changes in measurements, such as spectrometer lens positioning, heavily affect the predictions.

REFERENCES

Cremers, D. A., & Radziemski, L. J. (2013). Handbook of Laser-Induced Breakdown Spectroscopy: Second Edition. https://doi.org/10.1002/9781118567371

Takabe, H. (2020). *The Physics of Laser Plasmas and Applications - Volume 1*. https://doi.org/10.1007/978-3-030-49613-5

Hitz, C. B., Ewing, J., & Hecht, J. (2012). Introduction to Laser Technology: Fourth Edition. https://doi.org/10.1002/9781118219492

Silfvast, W. T. (2004). *Laser Fundamentals*. Cambridge University Press. https://doi.org/10.1017/CBO9780511616426

Noll, R. (2012). *Laser-Induced Breakdown Spectroscopy*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20668-9

Goodfellow lan, Bengio Yoshua, C. A. (2016). Deep Learning - Ian Goodfellow, Yoshua Bengio, Aaron Courville - Google Books. In *MIT Press*.

Avantes. (2022). Optical Spectrometers: an introduction. Retrieved August 24, 2022, from https://www.avantes.com/support/theoretical-background/introduction-to-spectrometers/

Quantel. (2022). Quantel Falcon 157. Retrieved Semptember 4, 2022, from https://www.quantel-laser.com/en/products/item/falcon.html

NIST. (2022). NIST LIBS Database. Retrieved Semptember 4, 2022, from https://physics.nist.gov/cgi-bin/ASD/lines1.pl?composition=Ni%3A50%3BPd%3A50&mytext%5B%5D=Ni&myperc%5B%5D=50&spectra=Ni0-2%2CPd0-2&mytext%5B%5D=Pd&myperc%5B%5D=50&low_w=200&limits_type=0&upp_w=600&show_av=2&unit=1&resolution=1000&temp=1&eden=1e17&maxcharge=2&min_rel_int=0.01&libs=1

Adam. (2017). ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION. https://arxiv.org/pdf/1412.6980.pdf