Metallic Elements Dissipation Avoided by Life cycle design for Steel

Deliverable 4.2

SysConfDeploy

Project details

Project Title	**Metallic Elements Dissipation Avoided by Life cycle design for Steel**		
Project Acronym	**MEDALS**		
Grant Agreement No.	101138516	**Duration**	48 months
Project Start Date	01-07-2024	**Project End Date**	30-06-2028

Document details

WP:	**4**	**WP Leader:**	UL
WP Title:	**Data-driven decision making (Big data storage, handling and use)**		
Task:	**4.2**	**Task Leader:**	UL
Task Title:	System configuration for deployment of the machine learning model for live decision making.		
Deliverable No.	**D4.2**		
Deliverable Title	SysConfDeploy		
Dissemination level	PU		
Written By	Ehsan Namjoo (UL)		
Contributing beneficiary(ies)	UL, LSA		
Approved by	Björn Glaser (KTH)		
Status	Draft		
Date	18 December 2025		

Funded by the European Union

Table of Contents

List of Figures

Funded by the European Union

1 Introduction

This document outlines the system configuration required for deploying machine learning models in a live decision-making environment. It addresses three key areas: the specifications of the real LIBS spectra samples supplied to UL by LSA, which provide essential background for selecting models capable of handling this type of data; the candidate machine learning models and their computational complexity, assessed to determine their suitability for integration into the LSA system; and the estimated processing time required for live decision-making based on the characteristics of the provided sample data. These topics are presented in the dedicated chapters on LIBS Sample Specifications, Machine Learning Model Evaluation, LSA System Integration, and Real-Time Processing Requirements.

2 General specifications of real LIBS data

The first step in designing a machine learning model is to investigate the data on which the model will be trained and to identify its core characteristics. LSA has provided UL with 56 LIBS spectra (multiple readings) corresponding to the standard grades listed in Figure 1. The information presented in this table can also be obtained from the BAS official website, which offers reference materials, including spectroscopic standards used for calibration [1]. Figure 2 also presents a LIBS reading of the standard grade SS-CSM 469.

Funded by the European Union

No.	C (%)	Si (%)	Mn (%)	P (%)	S (%)	Cr (%)	Mo (%)	Ni (%)	Sn (%)	Pb (%)	Zn (%)	Cu (%)	Fe (%)	Al (%)
469	0.279	0.421	0.598	0.015	0.020	11.93	----	0.246	----	----	----	----	----	----
470	0.153	0.335	0.235	0.024	0.035	17.68	----	0.369	----	----	----	----	----	----
471	0.095	0.326	0.417	0.018	0.023	23.85	----	0.960	----	----	----	----	----	----
472	0.227	1.050	1.020	0.032	0.029	15.82	0.661	1.950	----	----	----	----	----	----
473	0.172	0.604	0.494	0.019	0.030	9.06	0.950	----	----	----	----	----	----	----
551	----	0.018	----	1.01	----	----	----	0.76-	8.92	0.79	0.74	87.4	0.20-	0.052
552	----	0.019	----	0.77	----	----	----	0.56-	9.78	0.63	0.35	87.7	0.10-	0.023
553	----	0.022	----	0.68	----	----	----	0.44-	10.8-	0.47	0.49	87.0	0.056	0.017
554	----	0.038	----	0.41	----	----	----	0.22-	11.3-	0.34	0.22	87.4	0.022	0.005
555	----	0.036	----	0.18	----	----	----	0.11-	12.1-	0.24	0.16	87.1	0.010	<0.005
556	----	<0.005	----	0.10	----	----	----	0.014	13.2-	0.16	0.09	86.4	0.004	<0.005

Figure 1: Elemental concentrations of standard grades SS-CRM 469-473, SS-CSM 551-556

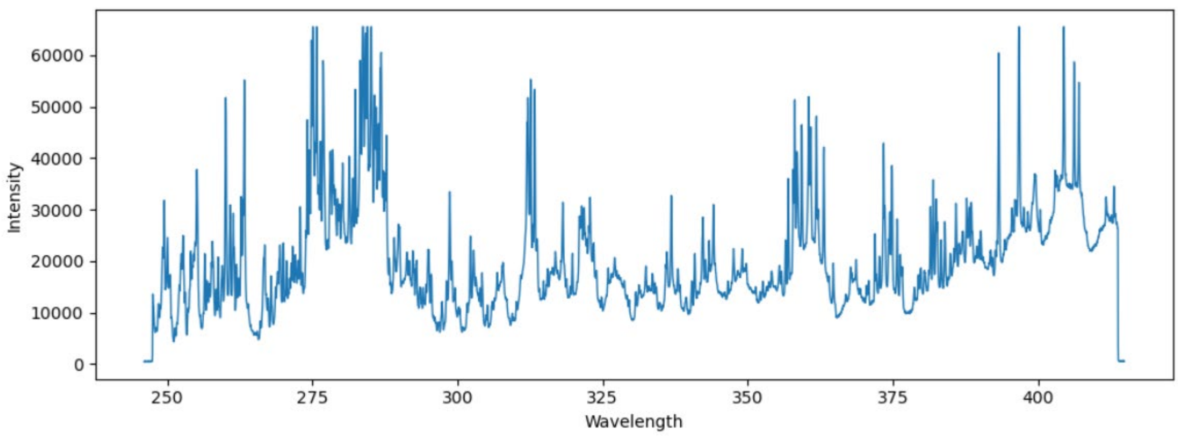

Figure 2: A sample spectrum from data provided by LSA (standard grade SS-CRM 469)

The following characteristics were observed across all spectra:

- **Wavelengths:** The set of wavelengths present in all spectra is identical. Each LIBS spectrum for a given standard grade contains 4,094 intensity values corresponding to 4,094 wavelength points.

Funded by the European Union

- **Sampling period:** The sampling period is not constant; it varies between 0.035 nm and 0.047 nm.
- **Sampling pattern:** Although the sampling period is non-uniform, the sampling pattern remains consistent across all spectra. Figure 3 illustrates the distribution of the sampling period (in nm) versus the index for all spectra.
- **Spectral range:** The spectral range for all samples lies between 246.018 nm and 414.849 nm.

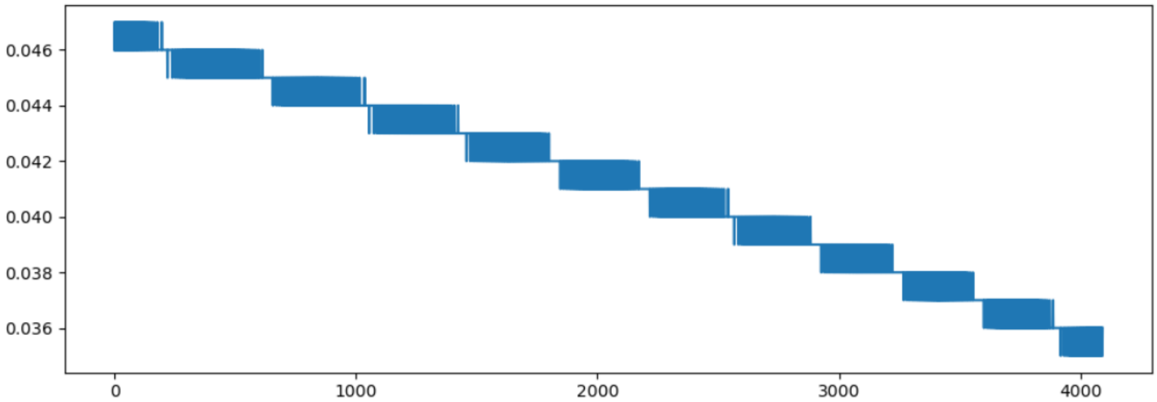

Figure 3: Distribution of sampling periods

In the next section, several candidate machine learning models are introduced, and their suitability in terms of computational complexity is evaluated.

3 Candidate machine learning models

Based on the information presented in the previous section, this part of the report introduces candidate machine learning models suitable for handling the LIBS spectrum data and evaluates the computational complexity of each model for processing an input spectrum.
because the wavelength values and sampling pattern are identical across all provided data, and under the assumption that future data supplied to the UL team will follow these same characteristics, we consider the following conditions when selecting candidate machine learning models:

- **Feature vectors**: Each sample is represented by a feature vector of length d, containing the intensity values (For the data provided by LSA, $d = 4094$).
- **Dataset**: The dataset consists of intensity values associated with the observed wavelengths and has dimensions $N \times d$, where N is the number of samples. The dataset may include both real and synthetic samples.
- **Number of classes**: For this report, the number of classes is considered to be equal to the number of distinct standard grades provided by LSA. However, the final model will likely be applied to the steel grades previously supplied to UL by Kobolde, and other partners.

Funded by the European Union

- **Order of intensities**: The intensity values for each LIBS spectrum will be ordered according to their corresponding wavelengths to avoid any potential mismatch.
- **Scaling:** The intensity values are typically very large in both the NIST LIBS samples and those provided by LSA. To reduce storage requirements and to facilitate efficient model training, all data will be scaled to the range (0, 1).
- **Nonlinearity of LIBS spectrum:** As observed, the LIBS spectra exhibit highly non-linear behaviour. Therefore, machine learning models known for effectively handling such data will be considered.

Given the nature of the machine learning task, and considering that the wavelength values remain fixed across the entire spectral range, the following classifiers will be initially considered for fast inference of steel grade classification from LIBS spectra:

- **K Nearest Neighbour (KNN)**

  It can capture complex, non-linear relationships in high-dimensional data without assuming any underlying distribution. However, its prediction time can be high for large datasets due to distance calculations.
- **Support Vector Machine (SVM) with RBF kernel**

  Non-linear kernels (e.g., RBF) allow SVMs to model non-linear patterns present in LIBS spectra. SVMs are effective for high-dimensional data, making them well suited for spectra with thousands of intensity values.
- **Decision Tree (DT)**

  Due to their fast evaluation and ability to capture nonlinear dependencies, decision trees are well suited for extracting characteristic spectral patterns in LIBS measurements.
- **Random Forest (RF)**

  Random Forests are well-suited to LIBS spectra because they can model the non-linearities and high dimensionality typical of this data. Their ensemble structure also provides robustness to noise while helping to limit overfitting, an important consideration when the available training data has relatively low variability. In cases where spectral variability is limited, data augmentation or similar techniques can be used to increase diversity in the training set and further reduce the risk of overfitting.

- **Gradient Boosting Machine (GBM)**

  GBMs are highly effective at capturing subtle, non-linear relationships in complex datasets, such as LIBS spectra. Their iterative approach allows for precise discrimination between steel grades, even when the differences in spectral features are small. In this report we utilize XGBoost as a powerful member of the GBM family.

4 LSA System Overview

The ML model for LIBS spectrum classification will be required to be integrated into the LSA sorting system. Figure 4 shows a general overview of the LSA system from the operator's point of view.

Funded by the European Union

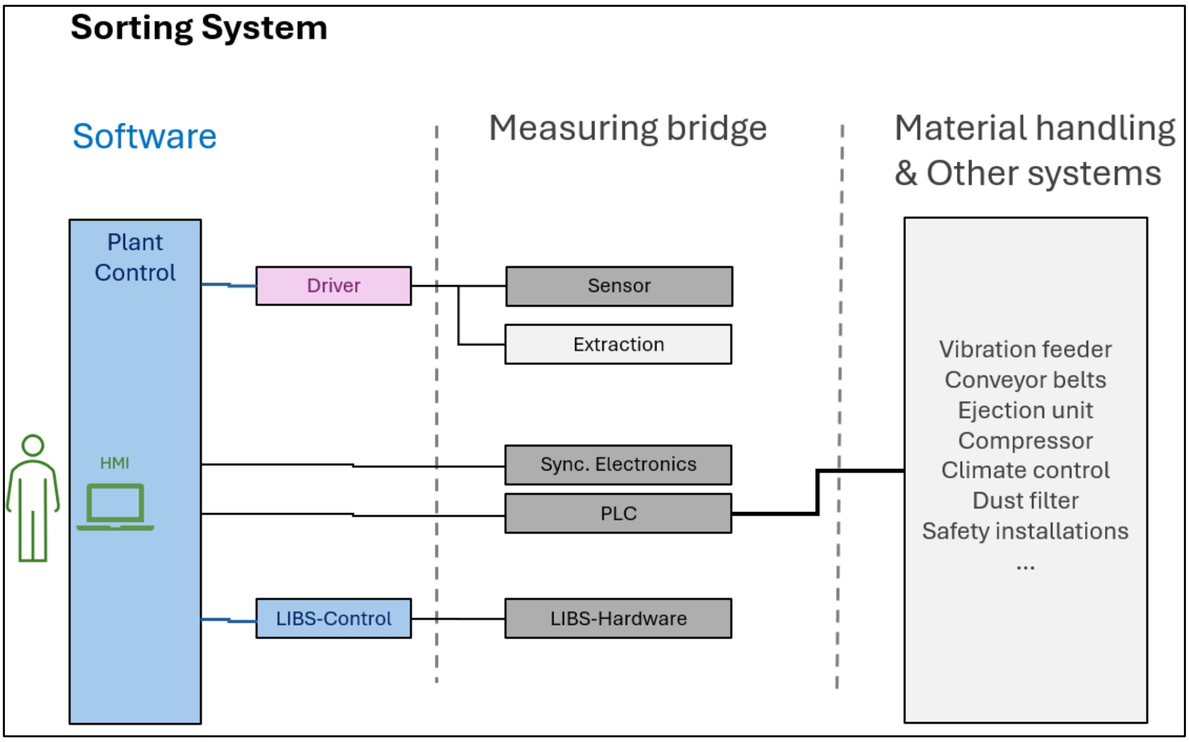

Figure 4: Schematic overview of individual hardware and software dependencies of the system from the operator's point of view.

The LIBS-Control application starts all applications that control the components of the LIBS sensor of the measuring bridge and the machine user interface (HMI). Table 1 shows an overview of the applications that communicate with LIBS-Control and their functions.

Table 1: Software applications for plant control

Application	Function
LIBS-Control	Startup of the driver applications and the **HMI,** synchronization of the **LIBS sensor** components for the control of the laser beam source, spectrometer system, and classification of the measurement data.
Laser Driver	Driver Application for the Laser Beam Source.
Classificator	Determination of the material class is based on **the measurement data.**
Plant control	Provision of the user interface **(HMI),** control of the 3D scanner, and the synchronization of electronics for positioning the laser beam on the

Funded by the European Union

	measurement objects, connection of the **HMI** and **the LIBS sensor** applications to the automation control (**PLC**).

The driver applications communicate with **LIBS Control** via the integrated interface drivers and sometimes directly with each other.
The control principle is based on a state machine that is implemented in **LIBS-Control** and in all driver applications. After receiving a status request from the higher-level process control (**PLC),** the drivers set the respective connected hardware or software components to the requested state. The necessary interfaces, commands and processes to achieve the desired state of the controlled component are integrated into each driver.
The UL developed machine learning model will replace the “Classificator” block and will be implemented on an Industrial PC with the other software components of the system. Currently the classificator block operates with a latency of several ms and in order to maintain the 50Hz operating rate of the LSA system any replacement classifier must also operate on a similar short timescale.

Funded by the
European Union

5 Computational consideration, storage requirements, and complexity

This section presents a general overview of the computational complexity of the candidate model capable of performing this classification task, followed by an estimation of the processing time required for each model to perform a single live evaluation.

5.1 Computational complexity

In this report, computational complexity refers specifically to the amount of computation required for a model to generate a prediction for a single input. It is important to note that we do not consider training-time complexity; all machine-learning models are assumed to be trained offline and deployed solely as a classifier.

A perfectly uniform comparison across all candidate models is not entirely feasible, as each classifier has unique structural properties that prevent fully equivalent evaluation conditions. Nevertheless, some general observations can be made. A decision tree with a reasonable depth typically offers the fastest evaluation, since prediction only involves traversing a single path from the root to a leaf, resulting in very low computational resource requirement [2].

A random forest requires evaluating multiple trees, so its prediction-time complexity grows with the number and depth of those trees; however, it remains efficient in practice. XGBoost, likewise, evaluates many trees, and although boosting often involves more trees than a random forest, each tree is usually shallower, resulting in slightly higher but still comparable complexity. Since both random forests and XGBoost rely on ensembles of decision trees, if the ensembles are configured with the same number of trees of identical structure, their evaluation complexity becomes effectively equivalent. SVMs with RBF kernels are generally more computationally demanding at inference time because prediction involves computing kernel values against all support vectors, which can be numerous; however, with a reasonably small support-vector set, SVM evaluation can be quite fast. Finally, KNN incurs a high computational cost, as it requires computing distances between the new input and every instance in the training dataset, resulting in complexity that scales linearly with dataset size.

5.2 Evaluation (inference time) time comparison

In this section, the candidate models are trained offline using the same dataset obtained from the reference samples detailed in Section 2, and their inference times are compared. The evaluations were performed on a Dell Alienware Aurora ACT1250, powered by an Intel Core Ultra 9285 (24 cores, 24 threads, up to 5.4 GHz) with 32 GB of DDR5-5600 RAM. An additional consideration (for later investigation) is whether the tested models fully utilise the available CPU cores and threads, as this depends on the specific implementation and execution of the C++ code. The major parameters of each model are as follows:

Funded by the European Union

Model Parameters Summary

Model	Parameter	Value
Decision Tree	Number of Nodes	525
	Max depth	15
	Number of Leaves	263
	Criterion	Gini Index
Random Forest	Number of trees	200
	Average Depth	18.83
	Averages Leaves	434,72
	Criterion	Gini Index
XGBoost	Number of trees	300
	Max depth	6
SVM (RBF)	Kernal	RBF
	Gamma	Scale
	Number of Support Vectors	3906
KNN	Number of neighbours	20

Table 1. Overview of structural characteristics and configuration parameters for the Decision Tree, Random Forest, XGBoost, SVM (RBF), and KNN models.

The inference times of the candidate models were measured to assess their computational efficiency. Among the models, KNN (k=20) and a SVM with a RBF kernel exhibited the longest average inference times, at approximately 6.34 ms and 6.36 ms, respectively. Random Forest also showed relatively high latency, averaging 28.10 ms per inference. In contrast, XGBoost and Decision Tree models were significantly faster, with average inference times of

Funded by the European Union

0.80 ms and 0.035 ms, respectively, highlighting their efficiency for real-time or large-scale applications. Table 2 shows the inference time for the candidate models in microseconds.

Table 2: The evaluation time for the candidate models

Model	Average Inference Time (µs)
SVM (RBF)	6,357
KNN (k=20)	6,339
Decision Tree	35
Random Forest	28,102
XGBoost	803

6 Conclusion

To allow successful real world deployment, machine learning models must be capable of being integrated into the LSA system with a short latency time of several ms. Preliminary work in this report has detailed inference times for common classifier models trained on the LIBS reference dataset, and it is demonstrated that millisecond scale inference is possible with standard classifiers. Further development work will be undertaken by UL to develop more advanced classifiers that can obtain high accuracy on real world LIBS data, but with a strong focus on minimising inference time in line with LSA system requirements.

The recommendation for the machine learning model integration within the LSA system configuration is to deploy the software portion of the system on a Siemens industrial PC *IPC BX-59A* [3]. This modular PC can be equipped with high performing CPUs core i5/i7/i9 which are suitable for fast inference times of standard classifiers. Additionally this family of IPCs will allow a PCI expansion slot which can be later filled with a GPU to allow acceleration of deep learning models if required. The 400W power supply should be specified to allow a power surplus to be used by the GPU.

References

1- Available at: Bureau of Analysed Samples - BAS
2- Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Chapman & Hall/CRC. ISBN 0412048418.
3-Siemens Simatic Industrial PC IPC BX-56A and IPC-59A

Funded by the European Union