

**Digital, autonomous, Intelligent and Synchronous system for  
Continuous identification, Optimization and Value Extraction of  
Resources from the end-of-use built environment**



**DISCOVER**

**D2.1 Drill LIBS (XRF): A Small-scale Portable XRF  
Analyzer with Drilling Capabilities**

**WP 2. Autonomous invasive material identification  
system for unresolved cases (M6 - M28)**

**WP Leader: TU Delft**

**Submission date: 31 October, 2025**



**Funded by  
the European Union**

The DISCOVER (GA 101129909) project is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

<b>Project name</b>	Digital, autonomous, Intelligent and Synchronous system for Continuous identification, Optimization and Value Extraction of Resources from the end-of-use built environment
<b>Grant Agreement number</b>	101129909
<b>Funding scheme</b>	Horizon Europe
<b>Project Acronym</b>	DISCOVER
<b>Project starting date</b>	01/06/2024
<b>Project duration</b>	48 months
<b>Deliverable number</b>	2.1
<b>Deliverable title</b>	Optimizing robotic drilling and material extraction for LIBS XRF analysis
<b>Deliverable version</b>	V1
<b>Work Package number</b>	2
<b>Work Package title</b>	Autonomous invasive material identification system for unresolved cases
<b>Due date of delivery</b>	31/10/2025
<b>Actual date of delivery</b>	31/10/2025
<b>Dissemination level</b>	PU - Public
<b>Type</b>	DEM
<b>Editor</b>	Dr. Eng. Mohanad Abukmeil (TU Delft)
<b>Contributor(s)</b>	Perry Posthoorn (TU Delft), Stephan van 't Hof (TU Delft), Yongli Wu (TU Delft), Francesco Di Maio (TU Delft)
<b>Reviewer(s)</b>	Alba Perez Gracia (UPC)
<b>Rights</b>	DISCOVER consortium

<b>Confidentiality</b>	
<b>Does this report contain confidential information?</b>	Yes <input type="checkbox"/> No <input checked="" type="checkbox"/>
<b>Is the report restricted to a specific group?</b>	Yes <input type="checkbox"/> No <input checked="" type="checkbox"/>

## Document history

Version	Date	Beneficiary	Description
V 1.0	31/10/2025	TU Delft	V1

## List of abbreviations

Abbreviation	Meaning
AAC	Lightweight Concrete
AI	Artificial Intelligence
Ca, Fe, Si, and Al	Calcium, Iron, Silicon, and Aluminum
DL	Deep Learning
KPIs	Key Performance Indicators
LIBS	Laser-Induced Breakdown Spectroscopy
ML	Machine Learning
Near-Infrared	NIR
PPM	Part Per Million
RGB	Read, Green, and Blue
WP	Work Package
XRF	X-ray Fluorescence

## List of tables

Table 1. Material selected for Pokeye testing and evaluation.....	9
Table 2. Key design elements and components .....	11
Table 3. Technical specification of drilling components .....	12
Table 4. Key electronic and electrical components of the drilling platform.....	15
Table 5. XRF test set classification KPIs.....	19
Table 6. NIR test set classification KPIs.....	22
Table 7. RGB Test Set Classification KPIs.....	23

## List of figures

Figure 1. Vanta Max XRF.....	10
Figure 2. AVANTES NIR.....	10
Figure 3. BASLER RGB.....	10
Figure 4. Modular lab setup design of the Pokeye drilling platform .....	10
Figure 5. Hilti TE 6-22 rotary hammer drill .....	11
Figure 6. Hilti TE-CD drill bit.....	11
Figure 7. Hilti VC 2D-22 vacuum unit.....	11
Figure 8. Sample holder.....	11
Figure 9. Automated workflow and processing stations.....	12
Figure 10. Rendered views of Pokeye drilling platform.....	13
Figure 11. Block diagram of the Pokeye drilling platform control system architecture .....	14
Figure 12. State machine workflow for automated sample processing.....	16
Figure 13. LabVIEW-based GUI for system control and diagnostics .....	17
Figure 14. Resampled 200-feature NIR vector for material classification input.....	21

## Table of contents

About the DISCOVER project.....	5
Abstract .....	6
1. Introduction and Main Objectives .....	7
2. Mechanical Design of the XRF Drilling Platform .....	9
2.1. Mechanical Design of the XRF Drilling .....	10
2.2. Automated Workflow and Processing Stations .....	12
3. Drilling Platform Control and Communication .....	13
3.1. System Architecture and Components.....	13
3.2. Control Logic and Communication .....	15
3.3. LabVIEW Graphical User Interface (GUI).....	17
4. Dataset collection and AI-based Identification.....	18
4.1. XRF Data collection and AI-based identification .....	18
4.1.1. XRF Dataset Collection and Preprocessing.....	19
4.1.2. XRF AI-based Material identification and classification .....	19
4.2. NIR Data Collection and AI-Based Identification.....	20
4.2.1. NIR Spectroscopy: Principles and Applicability.....	20
4.2.2. Data Acquisition and Spectral Processing.....	21
4.2.3. NIR AI-Based Material Identification and Classification.....	22
4.3. RGB Data Collection and AI-Based Identification.....	22
4.3.1. RGB Data Acquisition and Image Processing.....	23
4.3.2. RGB AI-Based Material Identification and Classification.....	23

## About the DISCOVER project

DISCOVER intends to develop an autonomous, synchronous, continuous and intelligent identification and data analysis system for materials and products in existing end-of-life built works. The proposed approach will provide key stakeholders, including academia research performers, along with construction industry representatives, with data-driven insights to make deconstruction more efficient, optimise the use of resources, improve the environmental footprints and enhance the circularity of construction and demolition, unlocking the potential of end-of-life built works, which will become material banks. The expected outcomes include an autonomous robotic platform coupled with continuous identification tools to scan built works and provide synchronous quantitative and qualitative data from different materials, including complex and concealed elements. Artificial intelligence algorithms will allow a rapid analysis of the properties and characteristics of components, and feed the automated scan-to-BIM model creation. The multi-dimensional BIM, including selective demolition processes, labour productivity, and technical planning, will become a Digital Twin of the demolition site optimised by social, economic, and environmental multi-criteria assessments. This approach will highly contribute to increase significantly the supply of traceable and sustainable construction materials and products to enhance their wider market acceptance, following the waste hierarchy. The social impacts of digital transformation in the construction sector will be considered, and new professional development tools for the relevant stakeholders will be proposed. The system will be tested in four different real demolition sites (Spain, Portugal, Poland and Belgium), offering a complete range of built work typologies and wide geographical coverage to demonstrate the replicability potential of DISCOVER, increasing the project dissemination capacity and awareness among the construction sector.

## Abstract

This report details the development of the Pokey drilling robotic platform for material characterization in the built environment, as part of deliverable 2.1. The system integrates a robotic drilling platform with three sensing modalities: X-ray Fluorescence (XRF), Near-Infrared (NIR), and RGB imaging. It is designed to autonomously collect and analyze drill dust samples from 17 distinct classes of construction materials. XRF was chosen over the initially proposed LIBS for its superior stability, reproducibility, and easier robotic integration. The mechanical design features a modular frame, automated sample transport, and Hilti drilling components for robust operation. A LabVIEW-based control system manages the workflow, communicating with motor drivers and sensors. The collected multi-modal data is used to train and evaluate AI models for real-time, high-accuracy material classification.

## 1. Introduction and Main Objectives

Task 2.1-2.3, “Optimizing robotic drilling and material extraction for LBIS XRF analysis” and developing a portable analyzer, focuses on developing and refining a fully automated and efficient system for invasive material characterization in the built environment. The work builds upon previous tasks in WP2 and integrates mechanical, electronic, and artificial intelligence (AI) components into a single, cohesive experimental platform.

The primary objective of this task is to design and optimize a robotic drilling and material extraction system capable of collecting representative samples from construction and demolition materials. These samples are analyzed using a combination of three sensing modalities including handheld X-ray Fluorescence (XRF), Near-Infrared (NIR), and RGB imaging, operating simultaneously to enable multi-modal material identification. Each sensing technology provides a unique and complementary contribution to the overall classification system:

- XRF provides quantitative elemental composition data (e.g., calcium (Ca), iron (Fe), silicon (Si), and aluminum (Al), etc.) ideal for identifying inorganic building materials such as concrete, sand-lime brick, ceramics, metals, gypsum, wood, and glass.
- NIR captures molecular bond (chemical functional groups) information from Carbon-Hydrogen (C-H) and Oxygen-Hydrogen (O-H) bonds, offering high discrimination for organic materials like plastics and wood, as well as hydration features in gypsum and concrete.
- RGB imaging provides visual, structural, and morphological features of the drilled material, such as color, texture, and particle size distribution, which enhance classification accuracy and allow detection of mixed or contaminated samples.

This multimodal integration ensures robust, complementary data acquisition where each sensor compensates for the limitations of the others. For example, XRF is highly accurate for elemental materials but cannot detect organic compounds perfectly; NIR excels at detecting molecular features but is less effective for metals and glass; RGB provides spatial and visual context that bridges both domains.

Originally, Laser-Induced Breakdown Spectroscopy (LIBS) was proposed for analyzing the elemental composition of materials within demolition spaces. However, extensive experimental evaluation performed by our team revealed several performance and integration limitations that led to the adoption of XRF spectroscopy as the primary analytical technique.

- Measurement stability and quantitative reproducibility: XRF provides highly stable and reproducible quantitative elemental concentration measurements with less variation respecting the surface condition, texture, and the surrounding atmosphere. In contrast, LIBS exhibits significant shot-to-shot variability caused by inconsistencies in plasma formation and energy coupling. These fluctuations directly compromise measurement reliability and degrade the accuracy of AI-based classification models we propose for real-time material identification.
- Enhanced robotic integration and field reliability: From a system engineering perspective, XRF offers straightforward integration and superior compatibility with autonomous and mobile platforms such as Pokeye robot proposed works package 2.

Unlike LIBS, XRF does not require optical alignment, precise focusing mechanisms, or high-power pulsed lasers. This simplifies integration with robotic drilling systems, enhances field reliability, and eliminates plasma-related instability while reducing mechanical complexity and maintenance frequency.

- **Non-Destructive multi-sensor capability:** XRF's non-destructive nature allows the same drill dust sample to undergo sequential or parallel analysis using additional sensors such as RGB camera and NIR spectrometers. Conversely, LIBS is inherently destructive, albeit on a micro-scale. This distinction has significant implications for sequential analysis. The multi-sensor (multi-modal in the AI context) approach enables comprehensive material characterization without altering or damaging the drilled powder sample of demolishing materials, ensuring consistency across measurement modalities and facilitating multimodal data fusion.
- **Optimized Input for AI-based material classification:** XRF provides a quantitative elemental concentration dataset for critical construction-related elements such as Ca, Fe, Si, and Al. Such a structured numerical dataset serves as an optimized input for machine learning (ML) and deep learning models, significantly outperforming the qualitative and often noisy spectral signatures generated by LIBS. Consequently, AI-driven classification models trained on XRF data achieve higher robustness, precision, and generalization performance in identifying and classifying material types.

The aforementioned sensors, including XRF, NIR, and RGB, are fully integrated into the Pokeye drilling platform, which is designed to operate autonomously within complex demolition environments. The platform is designed to help Pokeye navigate into various building conditions to identify and characterize a predefined set of 17 distinct material classes, representing the most common materials encountered in demolition and deconstruction scenarios across the European building stock.

These materials were carefully selected to ensure broad coverage of construction categories, including mineral-based materials, metals, polymers, and insulating materials. This diversity allows comprehensive testing of the platform's sensing, data collection, and AI classification capabilities, ensuring robustness across both inorganic and organic materials.

Table 1 highlights the material classes selected for Pokeye drilling platform optimization, dataset collection, and multi-modal AI evaluation.

No.	Material Type	Category	Justification
1	Concrete	Mineral	Common structural material widely used in construction and found in demolition sites.
2	Lightweight Concrete (AAC)	Mineral	Porous structural variant of concrete with distinct mechanical and spectral properties
3	Sand-lime Brick	Mineral	Typical masonry material with a characteristic lime-based composition
4	Ceramic Brick	Mineral	Common fired clay material representing traditional wall components
5	Composite (Mixed ceramic and sand-lime)	Mineral	Represents heterogeneous building materials encountered in mixed structures
6	Rubble Stone	Mineral	Irregular natural stone used in historical and foundation structures
7	Gypsum	Mineral	Common interior finishing and plastering material

No.	Material Type	Category	Justification
8	Wood	Organic	Represents natural organic material used in
9	Glass	Inorganic	Common building material used in windows and partitions
10	Aluminum	Metal	Representative metallic material used in frames and structures
11	Steel	Metal	Core structural element for reinforcement and frameworks
12	Stroyfoam	Polymer	Common insulation material
13	Poly(methyl methacrylate) (PMMA)	Polymer	Transparent plastic alternative to glass
14	Polycarbonate (PC)	Polymer	Durable transparent plastic used in glazing and panels
15	Polyvinylchloride (PVC)	Polymer	Widely used plastic in piping and window profiles
16	High-Density Polyethylene (HDPE)	Polymer	Representative of dense plastic materials found in construction
17	Glass Wool	Insulation	Typical thermal insulation material

*Table 1. Material selected for Pokeye testing and evaluation.*

The remainder of this report is organized as follows: [Section 2](#) highlights the mechanical design of the Pokeye drilling platform, including all sensors and mechanical components. [Section 3](#) reviews the electronic, electrical, and control systems of the proposed drilling platform. [Section 4](#) presents dataset collection and AI model development.

## 2. Mechanical Design of the XRF Drilling Platform

This section details the mechanical design of the integrated AI-based drilling and multi-sensor analysis platform developed for autonomous material identification.

The proposed drilling platform combines three synchronized sensors: a Vanta Max XRF spectrometer (as depicted in Figure 1), an AVANTES NIR spectrometer with a wavelength range of 900–1700 nm and a resolution of 2.0–50 nm (as shown in Figure 2), and a BASLIR RGB camera with a lens distance of 300 mm and a field of view (FOV) of approximately 48 × 41 mm, providing around 51 pixels per millimeter (as shown in Figure 3). All components are mounted on a modular ITEM aluminum frame.

A key feature of the design is the automated sample handling system, which uses a linear rail transport mechanism to sequentially move drill dust samples through distinct processing stations. The design prioritizes robustness for harsh demolition environments, precise sensor alignment to ensure high-quality data acquisition, and a compact footprint for seamless integration with the mobile Pokeye robot platform. All mechanical components were carefully selected to minimize maintenance requirements and ensure reliable field serviceability.



Figure 1. Vanta Max XRF.



Figure 2. AVANTES NIR.



Figure 3. BASLER RGB.

## 2.1. Mechanical Design of the XRF Drilling

The platform's design, as shown in Figure 4, integrates several key subsystems into a cohesive unit. The primary structure is a rigid frame constructed from ITEM aluminum profiles, chosen for their modularity and strength. A Hilti TE 6-22 rotary hammer drill depicted in Figure 5, modified for remote operation, is mounted on the robotic arm's end effector using a custom 3D-printed bracket. Construction material samples are collected using a hollow Hilti TE-CD drill bit (see Figure 6) that channels drill dust through a vacuum hose to a dust cyclone separator. The cyclone employs centrifugal force to separate solid

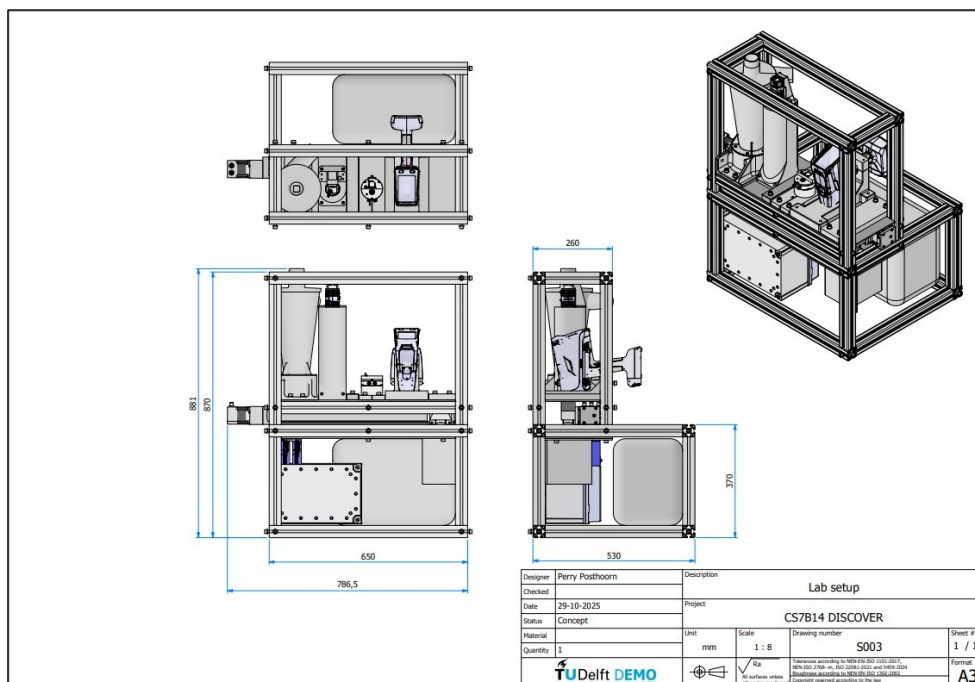


Figure 4. Modular lab setup design of the Pokeye drilling platform.

particles from the air stream: heavy drill dust is forced outward and downward into a collection funnel, while clean air exits through the top to the Hilti VC 2D-22 vacuum unit that is shown in Figure 7. This separation protects the vacuum motor from clogging while delivering concentrated drill dust directly into the sample holder positioned beneath the cyclone outlet.

The core of the automated workflow is a sample holder, shown in Figure 8 on a cart that travels along a linear rail. This holder features a vertically adjustable bottom to compress the sample, ensuring its top surface is consistently aligned for analysis. The entire sample processing area is enclosed within a 3 mm stainless steel box, which contains dust while providing strategically placed openings for each sensor and a disposal chute at the bottom. The following Table 2 summarizes the key design parameters and components of the platform.

Parameter	Specification
Overall Dimensions (H x W x D)	879 mm x 786.5 mm x 530 mm
Frame Material	ITEM Aluminum Profile
Enclosure Material/ Sampel Holder	3 mm Stainless Steel
Drilling Unit	Hilti TE 6-22 Rotary Hammer Drill
Drill Bit Type	Hilti TE-CD Hollow Drill Bit
Vacuum Unit	Hilti VC 2D-22
Vacuum Unit	XRF, NIR, RGB Camera
Transport Mechanism	Linear Rail with Motorized Cart

Table 2. Key design elements and components.



Figure 5. Hilti TE 6-22 rotary hammer drill.



Figure 6. Hilti TE-CD drill bit.



Figure 7. Hilti VC 2D-22 vacuum unit.

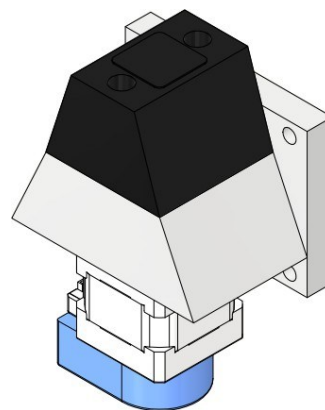


Figure 8. Sample holder.

Table 3 below outlines the key specifications for the primary drilling components used in the platform.

Component	Specification	Value
Hilti TE 6-22 Rotary drill	Speed	1050 rpm
	Energy	2.6 Joule
	Weight (without battery)	2.84 kg
Hilti TE-CD 16mm drill bit	Diameter	16 mm
	Max depth	400 mm

Table 3. Technical specification of drilling components.

## 2.2. Automated Workflow and Processing Stations

As shown in Figure 9, the platform operates by moving the sample holder sequentially through six dedicated stations to perform a complete multi-modal analysis. After the final station, the sample is discarded, and the holder returns to the first station to begin a new cycle. The automated workflow is highlighted in Figure 9 and proceeds as:

1. **Collection:** The sample holder is positioned under the dust cyclone to collect a fresh sample of drill dust.
2. **Visual Imaging:** The holder moves to the RGB camera station, where a high-resolution image of the sample's surface of the drilled powder is captured for material classification based on powder visual features.
3. **Compression:** At this station, the adjustable bottom of the sample holder is raised to compress the powder, creating a flat and dense surface for optimal spectroscopic measurements.
4. **NIR Spectroscopy:** The sample is subjected to NIR spectroscopy for wavelength-dependent chemical characterization and recognition.

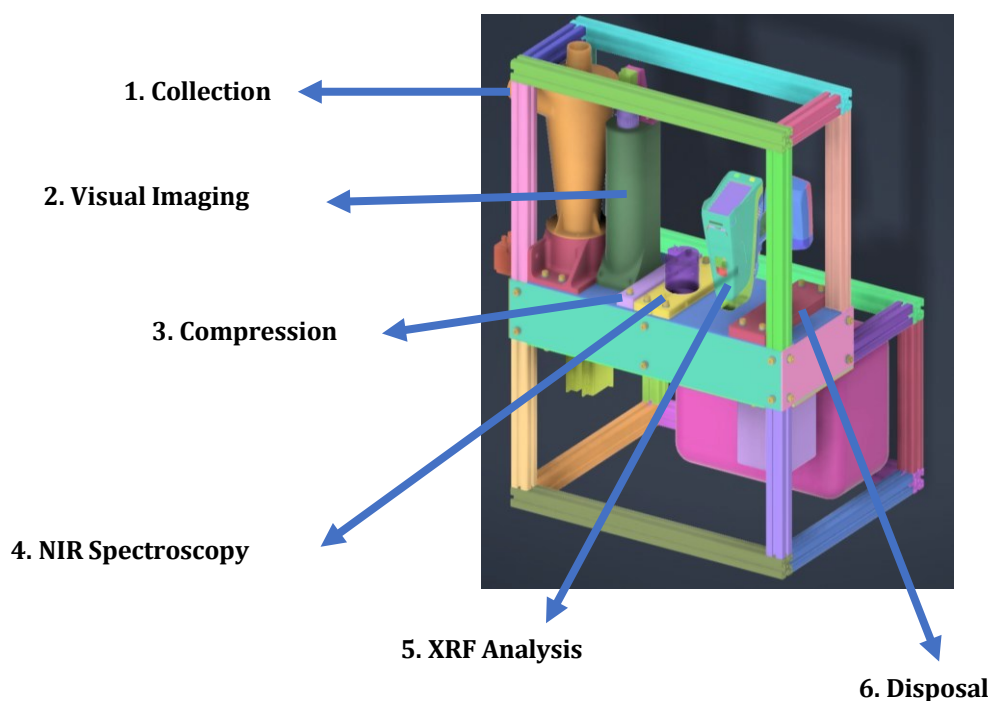


Figure 9. Automated workflow and processing stations.

5. **XRF Analysis:** The holder then moves to the XRF spectrometer, which performs a non-destructive analysis to determine the sample's elemental composition that used for material identification.
6. **Disposal:** Finally, the bottom of the sample holder retracts, and the analysed dust is discarded into a collection bin below.

After the final disposal station, the sample holder automatically returns to the first station to begin a new cycle, ensuring continuous operation. The sample holder is enclosed within a 3mm stainless steel housing featuring strategically positioned openings at the top for each analytical component (dust cyclone, RGB camera, compression mechanism, NIR spectrometer, and XRF analyzer) and an opening at the bottom for automated sample disposal into a collection bin. Figure 10 below presents various rendered views of the Pokey Drill platform, providing a comprehensive visualization of its design.



*Figure 10. Rendered views of Pokeye drilling platform.*

The control electronics and motion control systems are housed in a separate electrical cabinet, ensuring a clean separation between the analytical environment and the control infrastructure. Section 3 below highlights the electronic, control, and communication design aspects of the Pokeye drilling platform.

### **3. Drilling Platform Control and Communication**

This section describes the control, communication, and software architecture developed for the Pokeye drilling platform. The system integrates hardware, electronics, and software components into a unified automation framework that enables fully autonomous material collection, analysis, and disposal. The entire platform is managed via a LabVIEW-based control system, which coordinates the operation of sensors, actuators, and peripheral devices in real time.

#### **3.1. System Architecture and Components**

The core of the electronic control system is built around two Nanotec C5-E-1-81 motor drivers, which serve as the central control units of the drilling platform. These controllers provide precise motion control for the system's two stepper motors and manage digital I/O for auxiliary devices such as the Hilti drill and vacuum unit.

The drill is operated directly via a digital output channel, while the vacuum cleaner requiring higher current switching, is activated through a relay module. The platform's communication backbone is established using the Modbus TCP protocol, ensuring deterministic, reliable, and scalable data transfer between the LabVIEW software and the

motor controllers. All system components are interconnected as illustrated in Figure 11 which presents the high-level block diagram of the control and communication architecture.

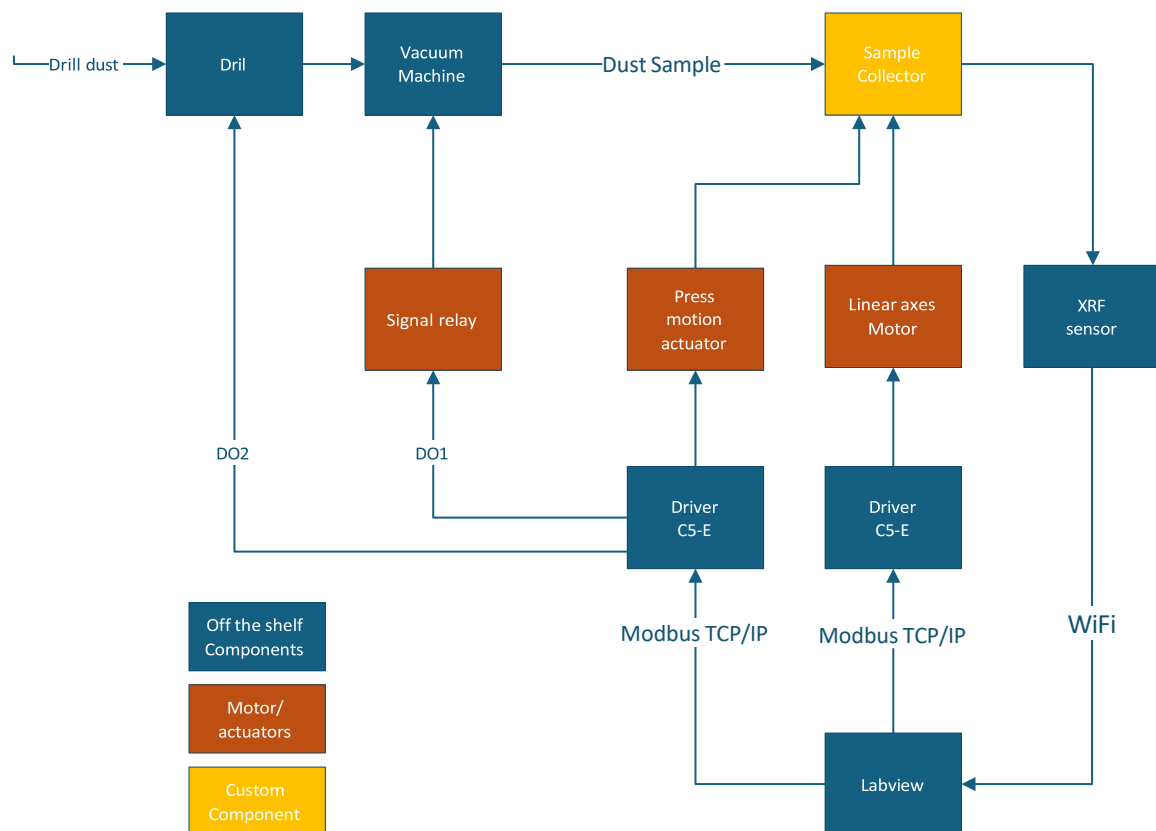


Figure 11. Block diagram of the Pokeye drilling platform control system architecture.

The platform integrates a range of electromechanical and analytical components to achieve its automated workflow. The following provides a detailed breakdown of these key components, specifying their model, power requirements, and primary function within the system (see Table 4).

Component	Brand	Type / Model	Max Power (W)	Function
Drill	Hilti	TE 6-22	550	Rotary drilling and material collection
Vacuum	Hilti	VC 2D-22	330	Drill dust extraction and filtration
XRF Spectrometer	Evident	Vanta VCA	70	Elemental composition analysis
Motor Drivers (x2)	Nanotec	C5-E-1-81	3.6 each	Stepper motor and peripheral device control
Linear Axis Motor	Nanotec	ASA5618M42-E3	15	Drives linear axis (horizontal motion)
Sample Holder Motor	Nanotec	LGA351S12-B-UIAP-019	10	Controls vertical compression and release

Component	Brand	Type / Model	Max Power (W)	Function
Linear Axis	Festo	ELGC-BS-KF-60-500-12P	—	Horizontal sample transport between stations
Relay Module	Finder	38.52.7.024.0050	—	Switches high-power devices (vacuum, drill)
NIR Spectrometer	Avantes	NIR256/512-1.7-EVO	4	Chemical/molecular composition analysis
RGB Camera	Basler	ACA2440-35u	2.7	Visual and structural imaging
Ring Light	Basler	Light Ring-700D-White	6	Uniform illumination for imaging station

Table 4. Key electronic and electrical components of the drilling platform.

### 3.2. Control Logic and Communication

The drilling platform operates as a fully automated, closed-loop system controlled by LabVIEW. The software communicates with the Nanotec motor controllers via Modbus TCP, enabling synchronized coordination of motion control and peripheral devices. Two stepper motors drive the motion subsystems, each equipped with encoders for real-time position feedback and precise motion execution:

- Motor 1 (Horizontal Axis): Drives the Festo linear rail, moving the sample holder between the six processing stations (collection, compression, RGB, NIR, XRF, and disposal).
- Motor 2 (Vertical Axis): Controls the sample holder lift mechanism, enabling compression of drill dust during preparation and release after scanning.

The platform's workflow is structured as a state machine architecture within LabVIEW. See Figure 12, which shows the machine architecture including XRF. For the other sensors, we are integrating them and will report on them in Task 2.4. Each operational phase is represented by a distinct state, allowing for modular development, debugging, and scalability. This proposed state machine diagram illustrates the automated workflow of the drilling platform's sample processing sequence. Starting from the "Idle" state, the system receives a "Scan sample" trigger and transitions through the following states:

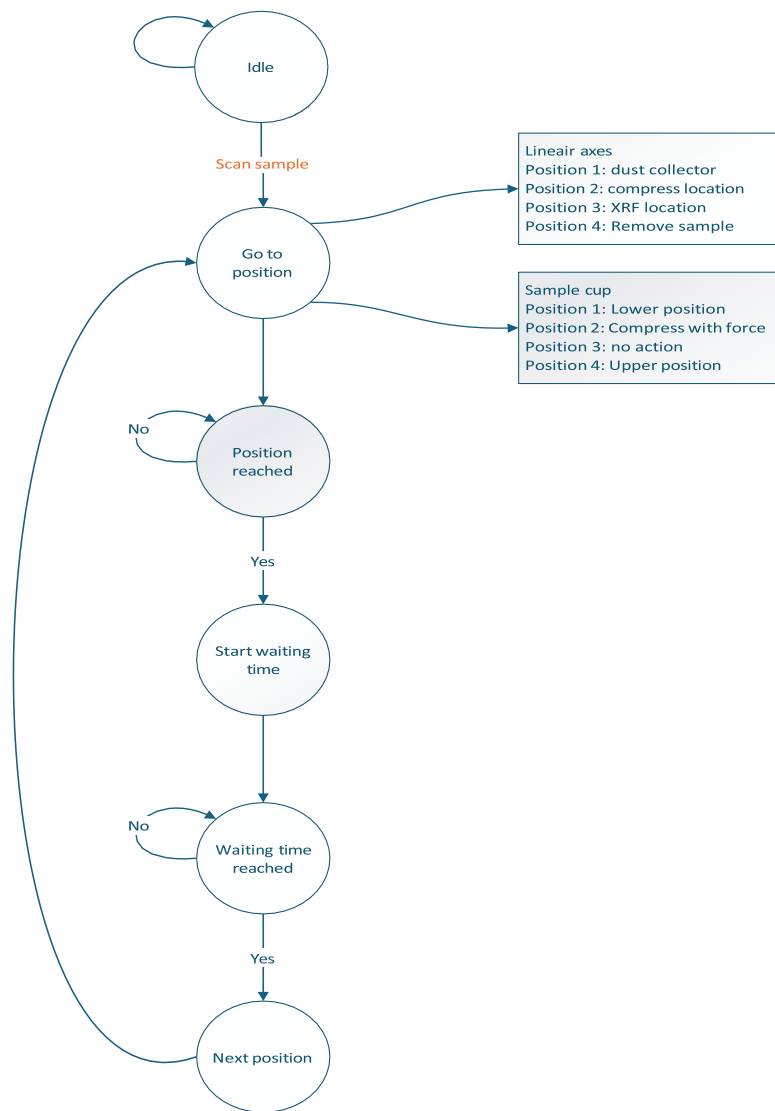


Figure 12. State machine workflow for automated sample processing.

1. Go to position: The system moves to one of four predefined positions along the linear axes:
  - Position 1: Dust collection
  - Position 2: Compression location
  - Position 3: XRF location
  - Position 4: Remove sample
2. Position reached: A decision state that verifies whether the motor has reached the target position using encoder feedback. If not reached (NO), it loops back; if reached (YES), it proceeds.
3. Start waiting time: Initiates a timer for sensor stabilization or mechanical settling.
4. Waiting time reached: Another decision state that checks if the required waiting period has elapsed. If not (NO), it continues waiting; if yes (YES), it advances.
5. Next position: Transitions to the next station in the sequence or returns to Idle after completing all positions.

For the Sample cup workflow, the vertical motor controls compression and release:

- Position 1: Lower position (collection)
- Position 2: Compress with force
- Position 3: No action (imaging/spectroscopy)
- Position 4: Upper position (disposal)

This modular state machine architecture, implemented in LabVIEW with Modbus TCP communication to Nanotec motor controllers, ensures synchronized motion control, high repeatability, and seamless integration of the six processing stations along the Festo linear rail system.

### 3.3. LabVIEW Graphical User Interface (GUI)

A dedicated Graphical User Interface (GUI), depicted in Figure 13, was developed in LabVIEW to provide complete control, monitoring, and debugging functionality for the drilling platform. The GUI serves as both the main operational dashboard for automated runs and a diagnostic tool for manual component testing.

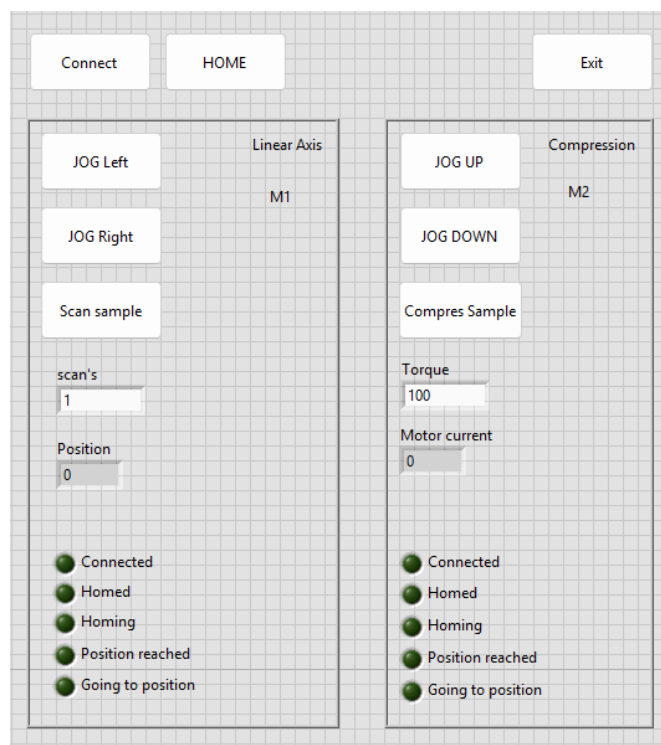


Figure 13. LabVIEW-based GUI for system control and diagnostics.

The GUI Key Features including:

- Automated Operation: A central “Scan Sample” button initiates the complete predefined state machine sequence, automating the sample collection, compression, scanning, and disposal processes.
- Manual Control: “Jog” buttons allow manual movement of both motors, enabling calibration, testing, and troubleshooting during setup.

- **Status Indicators:** Real-time system feedback displays the connection status, homing status of the motors, and readiness of sensors, ensuring operational safety and stability.
- **Monitoring Panel:** The interface provides live tracking of motor positions, active states, and error codes for efficient system management.
- The GUI design emphasizes intuitive operation and clear visualization, allowing both expert users and maintenance personnel to interact confidently with the system.

#### 4. Dataset collection and AI-based Identification

The LabVIEW control system manages synchronized data collection from three sensors, each producing outputs in distinct formats. The XRF spectrometer records elemental concentrations in either .csv or .json files, the NIR spectrometer generates spectroscopic data in the Raw8 format, and the RGB camera captures and stores a high-resolution image of each drilled powder sample in a designated location. This integrated multimodal setup enables AI-based material identification directly within the system machine learning models. The following process details the dataset collection pipeline and the development of AI models to achieve accurate and robust material classification.

To assess the performance of these models, several Key Performance Indicators (KPIs) were used. These KPIs provide quantitative measures of how effectively the AI models identify materials based on the data collected from each sensor modality.

- **Accuracy:** Measures the overall proportion of correctly classified samples among all predictions.
- **Precision:** Indicates the proportion of true positive predictions among all positive predictions made by the model.
- **Recall:** Represents the proportion of actual positive samples correctly identified by the model to evaluate its sensitivity.
- **F1-Score:** The harmonic mean of precision and recall, providing a balanced measure of both metrics.
- **ROC-AUC:** Represents the area under the Receiver Operating Characteristic curve, showing the model's ability to distinguish between classes across probability thresholds.
- **PR-AUC:** Measures the area under the Precision-Recall curve, assessing performance in imbalanced datasets by balancing precision and recall.

##### 4.1. XRF Data collection and AI-based identification

The XRF module of the Pokeye drilling platform provides quantitative elemental composition data that serve as input features for AI-based material classification. Each drilled powder sample collected by the robotic platform is analyzed using a handheld XRF spectrometer integrated within the system. The XRF records the emitted characteristic X-ray intensities of each element and converts them into concentration values expressed in

parts per million (ppm), providing a non-destructive and highly reproducible elemental profile for each material.

A total of 17 representative construction and demolition materials were analyzed, encompassing minerals, metals, polymers, organics, and insulation materials, as outlined in Table 1. To ensure robust and statistically representative data, 15 independent samples were collected per material, resulting in a dataset of 255 XRF spectra. Each spectrum captures the concentration of 36 detectable elements, including major construction-related elements such as Ca, Si, Al, Fe, K, Ti, and others.

#### 4.1.1. XRF Dataset Collection and Preprocessing

Each powder sample is subjected to XRF analysis following a standardized measurement protocol integrated within the Pokeye platform's automated workflow. After drill dust collection and compression in the sample holder, the holder is positioned precisely beneath the XRF analyzer at the designated XRF station. The measurement parameters were optimized for construction material analysis: 60-second measurement duration, 40 kV X-ray tube voltage (optimized for Geochem elements), auto-adjusted tube current, and fixed standoff distance with contact sensor for positioning repeatability.

The raw XRF output provides concentration values ppm for 36 elements: Mg, Al, Si, P, S, K, Ca, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Y, Zr, Nb, Mo, Ag, Cd, Sn, Sb, Ba, W, Hg, Pb, Bi, Th, U, and LE (Light Elements composite).

Prior to model training, the dataset was preprocessed to enhance data quality. XRF measurements are governed by instrument-specific Limits of Detection (LOD), representing the minimum elemental concentration reliably distinguishable from background noise (typically 10-50 ppm for favorable elements, 200-1000 ppm for light elements). Elemental feature columns for which >60% of measurements fell below the LOD were removed, eliminating 13 elements (V, Co, Se, Y, Nb, Mo, Ag, Cd, Sn, Sb, Ba, W, Hg, Bi). Sporadic below-LOD values in retained features were imputed with zero. This yielded a final feature set of 23 elements: Mg, Al, Si, P, S, K, Ca, Ti, Cr, Mn, Fe, Ni, Cu, Zn, As, Rb, Sr, Y, Zr, Pb, Th, U, LE.

#### 4.1.2. XRF AI-based Material identification and classification

Five ML algorithms were evaluated: Random Forest, Gradient Boosting, XGBoost, Logistic Regression, and Decision Tree. Each model was trained on the 23-element feature set using a 60/20/20 train-validation-test split with stratified sampling (153 training, 51 validation, 51 test samples (See Table 5).

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	PR-AUC
Gradient Boosting	95.31%	96.18%	95.10%	94.99%	99.2%	97.54%
Random Forest	95.31%	96.18%	95.10%	94.99%	98.93%	95.40%
XGBoost	90.63%	87.65%	89.71%	87.68%	99.53%	97.82%
Logistic Regression	85.94%	85.00%	85.78%	83.08%	97.73%	86.55%
Decision Tree	68.75%	60.58%	69.12%	63.06%	93.99%	61.52%

Table 5. XRF test set classification KPIs.

Gradient Boosting and Random Forest were identified as the best-performing classifiers, each achieving 95.31% accuracy with comparable performance across all metrics. Gradient Boosting showed slightly better confidence calibration, with ROC-AUC of 99.68% and PR-AUC of 97.54%, and was therefore selected for production deployment due to its consistent and reliable performance across material categories.

The pre-processed XRF dataset has been archived in both CSV (`xrf_building_materials.csv`) and Pickle (`xrf_building_materials.pkl`) formats, containing 23 elemental features per sample and material labels (1–17). The trained Gradient Boosting model will be integrated into the LabVIEW control system via a Python interface for Pokeye real-time material identification. The workflow involves extracting features from XRF output, generating predictions with confidence scores, and returning results for automated decision-making, enabling autonomous routing, and quality control.

## 4.2. NIR Data Collection and AI-Based Identification

Following XRF elemental analysis, each powder sample undergoes NIR spectroscopy to complement XRF-based identification and address its limitations for organic materials. While XRF effectively distinguishes mineral-based and metallic materials through elemental composition, it exhibits reduced sensitivity for polymers, wood, and other organics composed primarily of light elements (C, H, O, N). NIR spectroscopy captures molecular vibrational signatures, particularly C–H, O–H, and N–H bonds, enabling clear differentiation between organic and inorganic materials. This multi-modal sensing strategy leverages the complementary strengths of both XRF and NIR to achieve comprehensive identification across all 17 material classes.

### 4.2.1. NIR Spectroscopy: Principles and Applicability

NIR spectroscopy operates within the 700–2500 nm wavelength range, detecting absorption features produced by molecular bonds. The NIR region is particularly sensitive to:

- **C–H bonds (~1100 nm, ~1350 nm):** Strong absorption in organic materials such as polymers and wood.
- **O–H bonds (~1200 nm):** Indicative of materials containing water or hydroxyl groups (e.g., gypsum, concrete).
- **N–H bonds (~1500 nm):** Found in certain polymers and organic compounds.

For construction material classification, NIR spectroscopy offers several advantages:

- **Organic material discrimination:** Differentiates polymers (PMMA, PC, PVC, HDPE, Styrofoam) that appear identical under XRF.
- **Wood identification:** Detects cellulose and lignin signatures absent in XRF data.
- **Moisture sensitivity:** Captures water content in porous materials (AAC, gypsum, concrete).
- **Complementarity to XRF:** Provides molecular information orthogonal (independent) to elemental composition.

The Avantes NIR256/512-1.7-EVO spectrometer integrated into the Pokeye platform operates in the 900–1700 nm range, capturing key discriminative regions for construction material classification.

#### 4.2.2. Data Acquisition and Spectral Processing

Each powder sample was measured using NIR spectroscopy following the same protocol and sequence as the XRF analysis, resulting in **255 total samples (15 per material)**. After XRF measurement, the sample holder is transferred to the NIR station, where spectra are recorded under consistent geometry and illumination. Measurement parameters include:

- **Wavelength range:** 900–1700 nm
- **Integration time:** Auto-optimized for signal-to-noise ratio
- **Output format:** RAW8 files
- **Spectral resolution:** intensity values across 900-1700 nm wavelength range

A standardized preprocessing pipeline was applied to enhance signal quality and extract meaningful features:

1. Wavelength cropping to reduced range to 950–1450 nm to eliminate boundary noise.
2. Discriminative Region Selection which is considered key material separation observed at 1050–1200 nm (C–H, O–H overtone; weaker absorptions at multiples of fundamental vibrational frequencies) and 1300–1400 nm (C–H combination bands).
3. Spectral Smoothing by using Savitzky-Golay filter (window = 51, polynomial = 3) applied to reduce noise and preserve peak shapes.
4. Feature Resampling: from the three spectral regions were resampled to 90, 20, and 90 points respectively (C–H overtone, O–H overtone, C–H combination), yielding a 200-feature vector per sample (see Figure 14).

This produced a dataset of 255 samples × 200 spectral features, each corresponding to normalized intensity values at specific wavelengths.

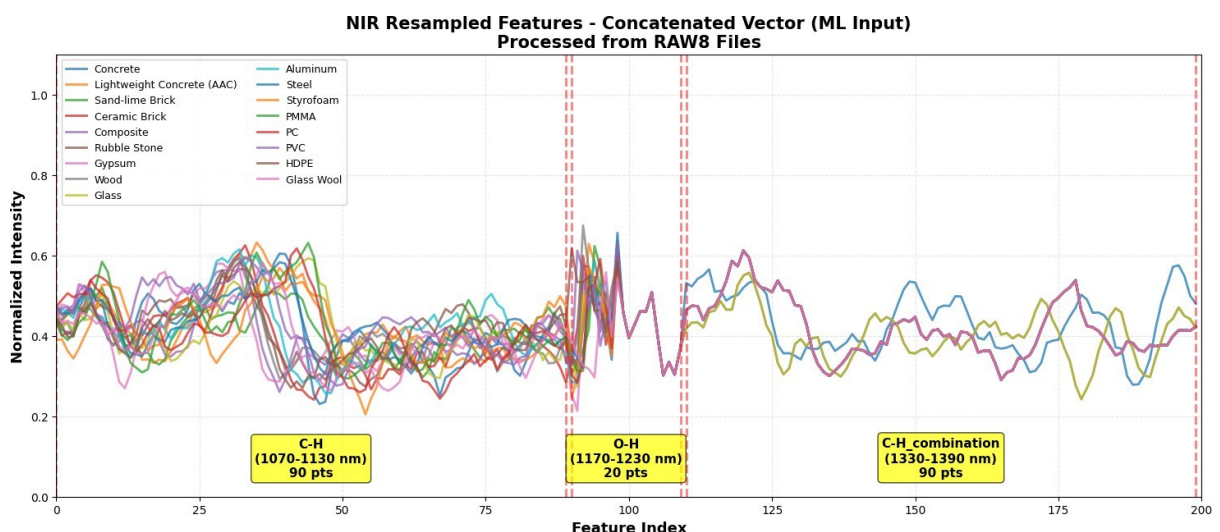


Figure 14. Resampled 200-feature NIR vector for material classification input.

### 4.2.3. NIR AI-Based Material Identification and Classification

Five machine learning algorithms were tested on the NIR dataset: Logistic Regression, XGBoost, Gradient Boosting, Random Forest, and Decision Tree. A 60/20/20 train-validation-test split was applied (153 training, 51 validation, 51 test samples) (see Table 6).

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	PR-AUC
Logistic Regression	95.29%	95.53%	95.29%	95.26%	99.91%	98.69%
XGBoost	95.10%	95.37%	95.10%	95.11%	99.89%	98.65%
Gradient Boosting	94.71%	95.00%	94.71%	94.67%	99.92%	98.81%
Random Forest	88.24%	88.49%	88.24%	87.95%	99.15%	92.54%
Decision Tree	72.16%	74.22%	72.16%	72.15%	96.00%	74.39%

Table 6. NIR test set classification KPIs.

Logistic Regression and XGBoost achieved the highest performance, with accuracies above 95%, indicating that the NIR spectra exhibit strong linear separability among material classes. Logistic Regression, showing the most stable performance and excellent calibration (ROC-AUC = 99.91%, PR-AUC = 98.69%), was selected for deployment due to its simplicity, computational efficiency, and robustness across both organic and inorganic materials. The preprocessed NIR dataset was archived in CSV (`nir_building_materials.csv`) and Pickle (`nir_building_materials.pkl`) formats, containing 200 reflectance features per sample and corresponding material labels (1–17). The trained Logistic Regression model was integrated into the LabVIEW control system through a Python interface, enabling real-time material identification.

During operation, the NIR spectrometer outputs raw spectral data, which are preprocessed via Python to extract 150 spectral features. These are then fed to the Logistic Regression model to generate class predictions and confidence scores. The results are transmitted back to LabVIEW for automated decision-making, allowing autonomous material routing and real-time quality control.

### 4.3. RGB Data Collection and AI-Based Identification

Following XRF analyzer and NIR spectroscopy, each powder sample undergoes RGB imaging to provide visual and morphological information complementary to chemical and molecular data. RGB imaging captures macroscopic (observable) features including color, texture, particle size, and surface morphology, offering an independent source of information for construction material identification. This modality is particularly valuable for detecting heterogeneous or mixed samples and for quality control to ensure consistency in sample preparation. The RGB imaging integration within Pokeye drilling platform provides several key advantages:

- Color discrimination through differentiating materials by characteristic color (e.g., red ceramic brick, white gypsum, yellow wood, gray concrete).
- Texture analysis by capturing surface roughness, particle size distribution, and grain structure.

- Morphological features extraction and detection from particle shapes, agglomeration patterns, and powder homogeneity.
- Complementarity to XRF/NIR by providing visual information independent of chemical composition and molecular structure.

The RGB images are preprocessed to extract meaningful visual features and generate training datasets for deep learning models.

#### 4.3.1. RGB Data Acquisition and Image Processing

Each powder sample was imaged following the same sequential protocol as XRF and NIR analysis, ensuring multi-modal correspondence. Images were captured under consistent illumination and sample positioning.

A preprocessing pipeline was applied to standardize the images for model training:

1. **Sample Center Detection:** Automatically detected powder regions within the holder image and cropped images to exclude metallic edges.
2. **Multi-Crop Data Augmentation:** Four 512×512 sub-images were cropped per sample from different offsets to increase dataset size and capture spatial variability.
3. **Dataset Construction:**
  - 255 unique powder samples × 4 crops = **1,020 images**
  - Image dimensions: 512×512×3 (RGB channels)
  - Pixel depth: 8-bit

Each material class (1–17) was assigned as the label for all images contained in its corresponding subfolder, ensuring a well-structured dataset for deep learning classification.

#### 4.3.2. RGB AI-Based Material Identification and Classification

Four deep learning architectures were evaluated on the RGB dataset: DINOv2, ConvNeXt V2, EfficientNetV2, and MaxViT. A 60/20/20 train-validation-test split was applied (612 training, 204 validation, 204 test images) (see Table 7).

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	PR-AUC
DINOv2	98.55%	98.91%	98.55%	98.51%	99.98%	99.90%
MaxViT	95.65%	96.74%	95.65%	95.53%	99.91%	99.08%
EfficientNetV2	62.32%	62.26%	62.32%	60.15%	95.49%	70.28%
ConvNeXt V2	40.58%	40.32%	40.58%	37.48%	98.00%	50.76%

Table 7. RGB Test Set Classification KPIs.

DINOv2 and MaxViT achieved exceptional performance, with accuracies of 98.55% and 95.65% respectively, demonstrating that deep learning vision transformers effectively capture hierarchical visual features from powder surface images. DINOv2, exhibiting the highest performance and excellent calibration (ROC-AUC = 99.98%, PR-AUC = 99.90%), was selected for deployment due to its superior accuracy, robust feature extraction capabilities, and consistent performance across both visually distinct and similar material classes. The preprocessed RGB dataset was archived in structured folder format, with

512×512-pixel images organized in subfolders by material class (1–17), containing 1,020 total images (4 crops per sample × 255 samples). The trained DINOv2 model was prepared for integration into the LabVIEW control system through a Python interface, enabling real-time material identification in the Pokeye platform.

During operation, the Basler RGB camera captures high-resolution images (2448×2048 pixels) of the powder sample. The image is preprocessed via Python to detect the sample center and extract a 512×512-pixel crop from the center region. This cropped image is fed to the DINOv2 model to generate material class predictions with confidence scores. The results are transmitted back to LabVIEW for automated decision-making, enabling autonomous material routing, real-time quality control, and multi-modal sensor fusion with XRF and NIR predictions.