

Cite this article

Folgozo-Bullejos C., García-Cortes V, Iturrioz JA, Galán Saiz P and Benito-Del-Valle L (2026) Assessment of in-line spectroscopy and CNNs for CDW fine fraction classification. *Proceedings of the Institution of Civil Engineers – Construction Materials* **179**(5): 361–372, <https://doi.org/10.1680/jcoma.25.00148>

Research Article

Paper 2500148

Received 01/12/2025; Accepted 16/06/2026

Published with permission by Emerald Publishing Limited under the CC-BY 4.0 license. (<http://creativecommons.org/licenses/by/4.0/>)

Assessment of in-line spectroscopy and CNNs for CDW fine fraction classification

Carlos Folgozo-Bullejos

Basque Research and Technology Alliance (BRTA), TECNALIA, Bizkaia, Spain (Orcid:0009-0002-8642-9635) (corresponding author: carlos.folgozo@tecnalia.com)

Verónica García-Cortes

Basque Research and Technology Alliance (BRTA), TECNALIA, Bizkaia, Spain (Orcid:0000-0002-9810-9973)

Jon Ander Iturrioz

Basque Research and Technology Alliance (BRTA), TECNALIA, Bizkaia, Spain (Orcid:0000-0001-6911-4530)

Pable Galán Saiz

Basque Research and Technology Alliance (BRTA), TECNALIA, Bizkaia, Spain (Orcid:0000-0003-3404-7774)

Leire Benito-Del-Valle

Basque Research and Technology Alliance (BRTA), TECNALIA, Bizkaia, Spain (Orcid:0000-0002-2152-1281)

Effective management of construction and demolition waste (CDW), particularly its fine fraction (<4 mm), faces technical challenges due to its heterogeneity and limited recyclability. Conventional characterisation and sorting technologies are ill-suited to handle this fraction, as its fine particle size and compositional heterogeneity exceed their operational limits, resulting in widespread landfill disposal. Exploiting this fraction as a secondary raw material requires accurate, real-time classification or regression methods. Unlike previous studies that used individual spectroscopic instruments, this study leverages a novel multi-sensor system capable of in-line analysis using four complementary techniques: laser-induced breakdown spectroscopy (LIBS), Raman, ultraviolet–visible (UV–Vis) and near-infrared (NIR) on fine CDW. The system was tested on 12 standard materials, including cementitious materials, recycled aggregates, and organic and polymeric compounds, resulting in a total of 49 152 spectra across the four modalities. One-dimensional (1D) convolutional neural networks (CNNs) were trained for multiclass classification. LIBS achieved the highest micro F1 scores (0.999), followed by UV–Vis (0.990), Raman (0.989) and NIR (0.869). This work demonstrates that the multi-sensor system captures material-specific spectral features that, coupled with 1D-CNN classifiers, allow reliable discrimination among fine CDW classes. The system’s ability to acquire data continuously and in-line provides a viable pathway towards scalable, automated characterisation in industrial settings.

Keywords: artificial intelligence/CDW/circularity/CNN/spectroscopy/waste

1. Introduction

Construction and demolition waste (CDW) constitutes the largest solid waste stream globally, representing $\approx 40\%$ of total waste in the European Union (Cristobal-García *et al.*, 2024). A considerable fraction consists of fine particles (<4–5 mm), produced during crushing and comminution processes. This fraction accumulates heterogeneous materials – including residual cement paste, gypsum, ceramics, wood and glass – that complicate their characterisation and limit their valorisation. Consequently, they are often stockpiled or downcycled, resulting in inadequate utilisation of valuable resources and increased environmental risks, such as sulphate leaching. Current separation technologies lack the precision and selectivity needed to efficiently process such complex mixtures. Although emerging techniques, such as the heating-air classification system, have shown potential for the selective separation of fine recycled concrete particles (Moreno-Juez *et al.*, 2020), their implementation in real CDW recycling contexts remains limited.

Over the past decade, the use of optical sensors and machine learning algorithms in the recycling sector has grown (Kroell

et al., 2022), with various spectroscopic techniques applied primarily to coarse CDW. Among them, near-infrared (NIR) spectroscopy – especially within the short-wave infrared range – has shown promising results for real-time classification of mixed demolition waste and recycled aggregates, often supported by multivariate statistical models such as principal component analysis (PCA) and partial least squares discriminant analysis (Vegas *et al.*, 2015; Vitek *et al.*, 2025; Xiao *et al.*, 2019). This progress has enabled industrial prototypes, like the WiserSort system (ICEBERG project, GA No. 869336), designed for sorting coarse materials such as concrete, ceramics and gypsum. However, large-scale adoption remains limited due to challenges in integrating these technologies into diverse and cost-sensitive industrial settings.

Other techniques, such as laser-induced breakdown spectroscopy (LIBS), have achieved high accuracy (>99%) in pre-industrial settings when combined with clustering algorithms and three-dimensional scanning for real-time quality assessment in recycled coarse aggregates (Chang *et al.*, 2025b, 2025a). Similarly, Raman spectroscopy, known for its molecular specificity, has recently been applied at laboratory level to the quantification of fine

ceramic and cementitious particles in recycled powders, showing strong correlation with reference laboratory techniques (Marín-Cortés *et al.*, 2024).

Despite these advances, most approaches focus on single spectroscopic modalities and linear classification models, limiting generalisability across variable CDW compositions. In contrast to coarse fractions, the fine fraction has received comparatively less attention, despite its relevance for circularity, and poses greater challenges due to smaller particle size, contamination and spectral interference. In parallel, deep learning algorithms, particularly convolutional neural networks (CNN), have emerged as powerful tools for analysing high-dimensional spectral data (Wu *et al.*, 2023). These models are capable of learning complex patterns from high-dimensional data sets and have outperformed conventional chemometric methods in various applications. Nonetheless, their integration with in-line spectroscopy for the classification of fine CDW is still underexplored.

This study addresses this gap by evaluating a multi-sensor framework that combines four complementary in-line spectroscopic techniques: NIR, LIBS, Raman and ultraviolet–visible (UV–Vis) with one-dimensional (1D)-CNN classification models for the identification of fine CDW materials. Representative materials found in CDW were individually measured to construct a labelled spectral data set, enabling models to learn from pure, unambiguous spectral signatures. Different network architectures and PCA-based outlier-detection strategies are systematically evaluated to characterise the discriminative capability of each sensor independently.

Specifically, the present work makes four contributions to the field: (i) it introduces a multi-sensor spectroscopic platform for in-line analysis of fine CDW, a fraction largely overlooked in previous sorting studies; (ii) it establishes a large labelled spectral

data set of 49 152 acquisitions covering 12 representative material classes across four complementary techniques; (iii) it benchmarks independent 1D-CNN classifiers per sensor; and (iv) it demonstrates the added value of combining (i)–(iii) into a unified workflow for fine-fraction CDW classification.

The results indicate the potential of this approach for in-line classification, although further development is needed for industrial deployment. Furthermore, these contributions provide a reference framework for future data fusion strategies in automated CDW processing – particularly relevant for compositionally complex mixtures where single-sensor models are expected to show limited discriminative capacity.

2. Materials and methodology

2.1 Sample selection and characterisation

The materials analysed in this study included: cement paste (R1-CP), limestone sand (R2-LS), siliceous sand (R3-SS), ceramic fraction (R4-CF), recycled limestone concrete aggregate (R5-RLCA), siliceous mortar (R6-SM), gypsum (R7-G), wood, glass powder (R9-GP), polyethylene (R10-P) and iron (R11-I) (see Figure 1). Two forms of wood, fibres (R8-FW) and chips (R8-W), were evaluated separately. All materials were processed to obtain a particle size below 4 mm, except for R10-P, which consists of spherical pellets, and R8-W. Samples R1-CP, R5-RLCA, R6-SM, and R7-G were synthesised under controlled laboratory conditions by mixing the relevant constituents (e.g., cement, sand, gypsum and water), allowing the mixtures to cure, and subsequently crushing the hardened products to the desired particle size. The remaining samples were commercially sourced and subjected to mechanical conditioning (crushing and/or sieving) to meet the granulometric requirements.

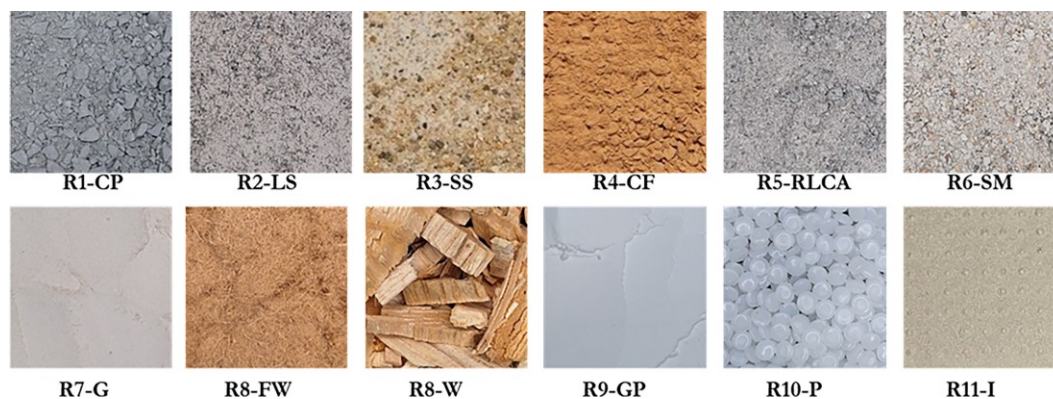


Figure 1. Construction and demolition waste materials characterised and classified using spectroscopic and deep learning methods (R1-CP, cement paste; R2-LS, limestone sand; R3-SS, siliceous sand; R4-CF, ceramic fraction; R5-RLCA, recycled limestone concrete aggregate; R6-SM, siliceous mortar; R7-G, gypsum; R8-FW, wood fibres; R8-W, wood chips; R9-GP, glass powder; R10-P, polyethylene; R11-I, iron)

The chemical composition was determined by X-ray fluorescence (XRF) using a Thermo Scientific ARL PERFORM'X 4200 spectrometer. This instrument features a 4200 W solid-state X-ray generator, operating at up to 70 kV and 140 mA. Mineralogical phases were identified by X-ray diffraction (XRD) employing a Philips X'Pert PW 1820 powder diffractometer (Philips Analytical), equipped with a Cu K α radiation source ($\lambda = 0.154$ nm), operating at 40 kV and 40 mA. Data were collected over a 2θ range of 2–75°, with a step size of 0.02° and a counting time of 2 s per step.

2.2 Spectroscopy techniques

2.2.1 Multi-sensor equipment

A multi-sensor equipment was used to perform the spectral measurements. It integrates four spectroscopy techniques (LIBS, Raman, NIR and UV–Vis) and a conveyor belt to place the materials under study. The optical system is formed by quartz optical fibres to drive the light from the source to the physical point of analysis and to collect the light from the material after the interaction. Furthermore, the optical system is mounted on a movable head on the axis transverse to the direction of movement of the conveyor belt. The conveyor and the head are coordinated to enable point-by-point analysis. Before measuring, the equipment was calibrated using reference materials, white PTFE (polytetrafluoroethylene) for UV–Vis and NIR, aluminium for LIBS and sulfur for Raman.

LIBS uses a pulsed neodymium-doped yttrium aluminium garnet (Nd:YAG) laser (providing pulses up to 60 mJ and <7 ns width at 1064 nm) and two spectrometers to cover a spectral range [226–631] nm. Samples were measured in powdered form. To avoid the laser reaching the sample holder surface, due to sample scattering after each pulse, the spectrum of each point is obtained from the average of five pulses using an integration time of 1 ms.

Raman spectra were measured with an excitation laser emitting at 784.8 nm with a power of 810 mW and width of 54 pm. A cadmium zinc telluride (CZT) spectrometer with a mean focal length of $f \approx 200$ mm and a back-thinned charge-coupled device (CCD) refrigerated to -40°C is used to capture the spectra which cover the spectral range [100–2100] cm^{-1} (Raman shift). One measurement per point with an integration time of 4 s was applied.

NIR measurements were carried out using a tungsten incandescent filament of 5 W with electronically stabilised current source. The spectrometer has a thermoelectrically cooled InGaAs detector covering the range [900–1700] nm with a resolution of 8 nm. An integration time of 10 ms and three measurements per point was applied.

UV–Vis spectra were performed using a pulsed Xenon lamp of 2 W (providing pulses up to 20 mJ) with a lineal complementary metal oxide semiconductor (CMOS) that cover the spectral range

[190–700] nm with a spectral resolution of 6 nm. An integration time of 2 ms and three measurements per point was applied.

The measurement parameters of all sensors were optimised to ensure reliable band/peak detection. Longer integration times yielded no significant spectral improvement, so shorter acquisition parameters were selected to reduce total measurement time, which increases the number of measurements per unit time and thus enhances industrial scalability.

2.2.2 Data and preprocessing

High-quality data sets are required to capture sufficient variability for effective artificial intelligence (AI) model training. To address this, the equipment is configured to carry out a point matrix analysis over an area with known dimensions. This procedure consists of measuring a set of spatially distributed points in the form of a matrix and equally spaced over the surface material. In each point, all the spectroscopy techniques hit the sample one by one automatically. A 16×16 matrix is selected resulting in 256 spectra over a square sample holder with a side length of 100 mm. To generate the spectral data set, four holders of each material were analysed resulting in 12 288 measurements per technique.

To improve the quality of the spectra and the signal–noise (S/N) ratio, a pre-processing step was performed before training. For LIBS, standard normal variate (SNV) was first applied independently to each spectrum acquired from the two spectrometers comprising the instrument. After that, a baseline correction was applied with asymmetric least square (ALS) to all the signal ($\lambda_m = 10^8$, $p = 10^{-2}$). For Raman, the spectral pre-processing pipeline includes dark signal subtraction, wavelength range adjustment, Savitzky–Golay (S–G) filtering (window length = 11, polynomial degree = 3), ALS-based baseline correction ($\lambda_m = 10^8$, $p = 10^{-2}$) and, finally, SNV normalisation was applied to minimise the influence of scattering and scaling differences. For NIR and UV–Vis reflectance data, preprocessing involved dark signal subtraction, comparison with a reference spectrum, smoothing using S–G filter (using Raman parameters) and SNV. The pre-processing steps were uniformly applied across all model training.

To prevent data leakage between partitions and ensure an unbiased evaluation of model generalisation, the data set was split at the sample holder level: two holders per material class were assigned to training (50%), one to validation (25%) and one to testing (25%). To identify outliers, PCA was fitted on the combined training and validation data set. Outliers in the training subset were flagged and excluded using the PCA reconstruction error (inverse projection), with the 98th percentile used as the threshold.

2.3 Model architectures

The goal of this work consists of generating a model able to classify samples from the fine fraction of CDW. To this end, a

comparison of different architectures of a 1D-CNN is investigated. For the classification task, four architectures have been used (CNN2, CNN3, CNN4 and CNN5), which share the same base architecture but vary in the number of convolutional blocks. The upper limit of five blocks was determined by the spectral interval of NIR data (≈ 250 wavelengths), as a sixth block would collapse the feature dimension to one, creating an information bottleneck. Each of these blocks, as shown in Figure 2, consists of a 1D convolution layer, batch normalisation, rectified linear unit (ReLU) activation, and 1D max pooling, drawing from architectures presented by Picon *et al.* (2025) and Ta *et al.* (2023).

Batch normalisation is included to stabilise training and accelerate convergence. The output of these blocks results in embeddings that represent the features of the input signal. These embeddings pass through a flatten layer, a linear layer and a SoftMax activation to produce the probability of the input belonging to a specific class.

To determine the class corresponding to the input signal, the model's output undergoes post-processing. This involves applying thresholds to optimise the classification results. Each class has its own threshold, which is obtained using the validation set. The threshold values range from 0 to 1 and are adjusted in increments of 0.001 to maximise the micro-F1 score (m-F1). This metric was chosen to assess both accuracy and sensitivity in the balanced data set, as it provides a unified measure of performance across all classes. Ultimately, the thresholds that achieve the highest m-F1 score for the validation set are selected, ensuring the most accurate classification. When multiple classes exceed their thresholds, or none do, the class with the highest SoftMax probability is assigned by way of argmax operation.

The number of epochs and batch size were fixed to 100 and 48, respectively. Among the two tested learning rate values, 10^{-4} and 10^{-5} , the optimal value was selected per technique: 10^{-4} for Raman, NIR and UV-Vis, and 10^{-5} for LIBS. All the experiments were implemented with the PyTorch (version 2.4.0) and PyTorch

Lightning (version 2.4.0) deep learning libraries and the torchvision (version 0.19.0) package. The study was performed on an Ubuntu 22.04 server with an NVIDIA TITAN X graphics processing unit (GPU).

3. Results and discussion

3.1 Chemical and mineralogical characterisation

The combined XRF and XRD analyses confirmed the heterogeneous nature of the CDW fine fractions, reflecting their diverse material origins. The main chemical compositions (expressed as equivalent oxides) and mineralogical phases are summarised in Figure 3 and Table 1, respectively.

A strong correlation was observed between chemical and mineralogical results. Siliceous materials, characterised by high silicon dioxide contents, consistently exhibited quartz and feldspar phases. In contrast, calcium oxide-rich samples correlated with the presence of calcite, portlandite and calcium silicates, typical of cementitious residues. The gypsum-based sample (R7-G) showed a marked sulfur trioxide enrichment, corresponding to the predominant presence of bassanite. Furthermore, significant amorphous contents were detected in cement-based and ceramic-rich fractions, likely associated with partially hydrated binders or vitrified phases formed during material processing. The alignment between chemical and mineralogical data ensures the representativeness of the selected materials and establishes a reliable baseline for this study.

3.2 Spectral characterisation

3.2.1 Raman

As observed in previous studies (Marín-Cortés *et al.*, 2023), averaging Raman signals – commonly used to enhance S/N – can hide characteristic bands or distort their relative intensities, potentially reducing the model's accuracy for classification. Therefore, pre-processed spectra were used for model training to preserve spectral variability, while average spectra were presented to facilitate interpretation.

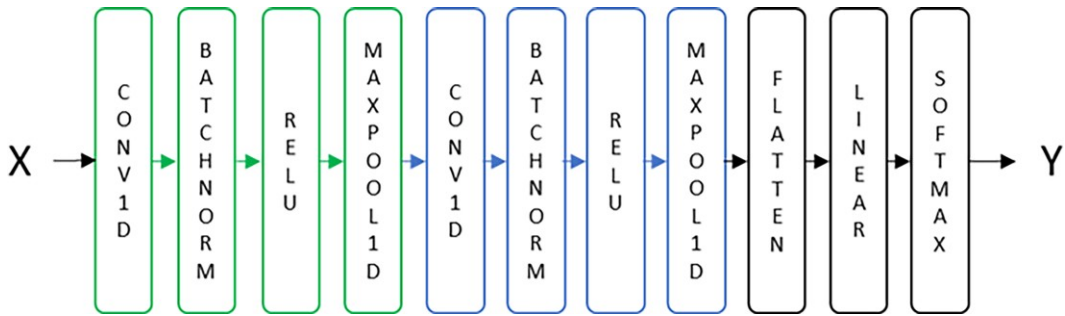


Figure 2. Two convolutional neural network block-based architecture (CNN2)

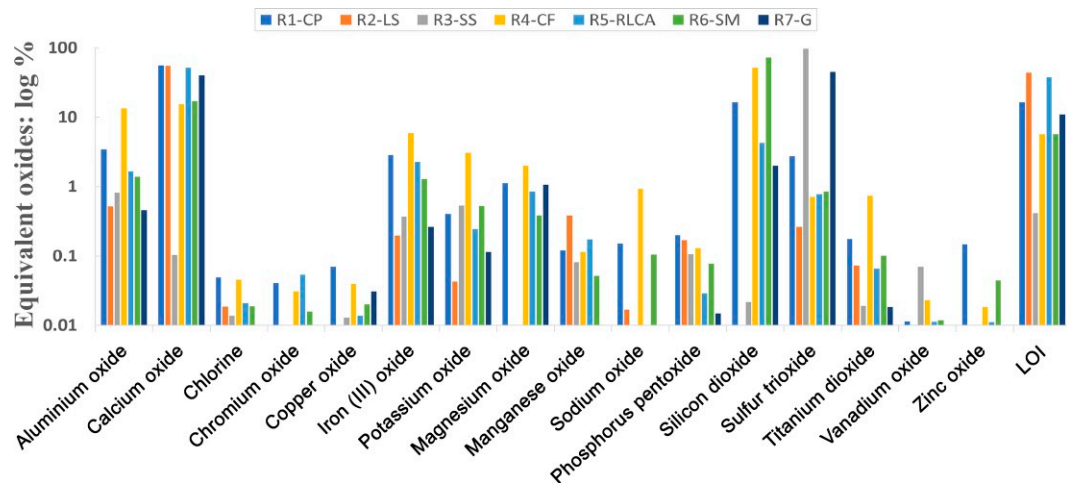


Figure 3. Chemical composition determined by X-ray fluorescence and expressed as wt. % equivalent oxides of inorganic materials (R1-CP, cement paste; R2-LS, limestone sand; R3-SS, siliceous sand; R4-CF, ceramic fraction; R5-RLCA, recycled limestone concrete aggregate; R6-SM, siliceous mortar; R7-G, gypsum; LOI, loss on ignition)

Table 1. Mineralogical phases identified by X-ray diffraction (XRD)

Mineral phase	R1-CP	R2-LS	R3-SS	R4-CF	R5-RLCA	R6-SM	R7-G
Albite, (Na,Ca)Al(Si,Al) ₃ O ₈	—	—	•	•	—	—	—
Alite, Ca ₃ SiO ₅	•	—	—	—	—	—	—
Anortoclase, (Na,K)(Si ₃ Al)O ₈	—	—	—	—	—	••	—
Bassanite, CaSO ₄ ·1/2H ₂ O	—	—	—	—	—	—	•••••
Belite, Ca ₂ SiO ₅	•	—	—	—	—	—	—
Calcite, CaCO ₃	••	•••••	—	••	•••••	•	•
Quartz, SiO ₂	—	—	•••••	•••••	—	•••••	•
Dolomite, CaMg _{0.77} Fe _{0.23} (CO ₃) ₂	—	•	—	—	•	—	•
Microcline, KAlSi ₃ O ₈	—	—	•	—	—	—	—
Moscovite, KAl ₃ Si ₃ O ₁₀ (OH) ₂	—	—	•	••	—	—	—
Portlandite, Ca(OH) ₂	••••	—	—	—	••	•	—
Amorphous phase	•••	—	—	—	••	•	•

Note(s): mineralogical phases identified by XRD (• = trace; •• = low; ••• = moderate; •••• = high; ••••• = very high) in inorganic materials; R1-CP, cement paste; R2-LS, limestone sand; R3-SS, siliceous sand; R4-CF, ceramic fraction; R5-RLCA, recycled limestone concrete aggregate; R6-SM, siliceous mortar; R7-G, gypsum

The average Raman spectra of all samples are shown in Figure 4. Several characteristic features can be observed across the data set. R1-CP presents a low intensity spectrum and lacked Raman bands despite XRD confirmation of calcite and portlandite. Calcite, identified as the dominant mineral phase in R2-LS and R5-RLCA, shows its characteristic Raman bands at 283.7 cm⁻¹ (translational lattice mode), 710.5 cm⁻¹ (symmetric (CO³)²⁻ deformation) and 1087.5 cm⁻¹ (symmetric (CO³)²⁻ stretching) (Gunasekaran *et al.*, 2006). In R5-RLCA, calcite bands are reduced in intensity due to the mixing of materials which decreases the aggregate concentration. Quartz was identified in R3-SS and R6-SM, by bands at 208.9 cm⁻¹ and 465 cm⁻¹ (silicon–oxygen–silicon symmetric stretching vibrations) (Liu *et al.*, 2022), but no clear signals were observed in R4-CF, likely due to signal attenuation or fluorescence. Fluorescence interference was notable in R3-SS and R6-

SM, affecting interpretability. Gypsum shows a characteristic band at 1010.7 cm⁻¹ (symmetric stretch SO⁴ tetrahedra) (Buzgar *et al.*, 2009), though weak due to fluorescence. Glass exhibits a broad band centred around 1400 cm⁻¹, attributed to photoluminescence, consistent with previous findings using a 784.8 nm laser (Tuschel, 2016). R10-P shows characteristic bands detected at 1061.4, 1081.1, 1129.2, 1170.4, 1295.3, 1269.9, 1418.6, 1440.4 and 1461.0 cm⁻¹, corresponding to vibrational modes of C-C and CH² bonds (Furukawa *et al.*, 2006). In wood samples, the spectra are dominated by fluorescence, masking characteristic bands. As expected, the metal sample R11-I does not exhibit Raman activity due to the absence of vibrational modes. Overall, Figure 4 illustrates the strengths of Raman spectroscopy in identifying specific mineral phases, while also highlighting its limitations in highly fluorescent or metallic samples.

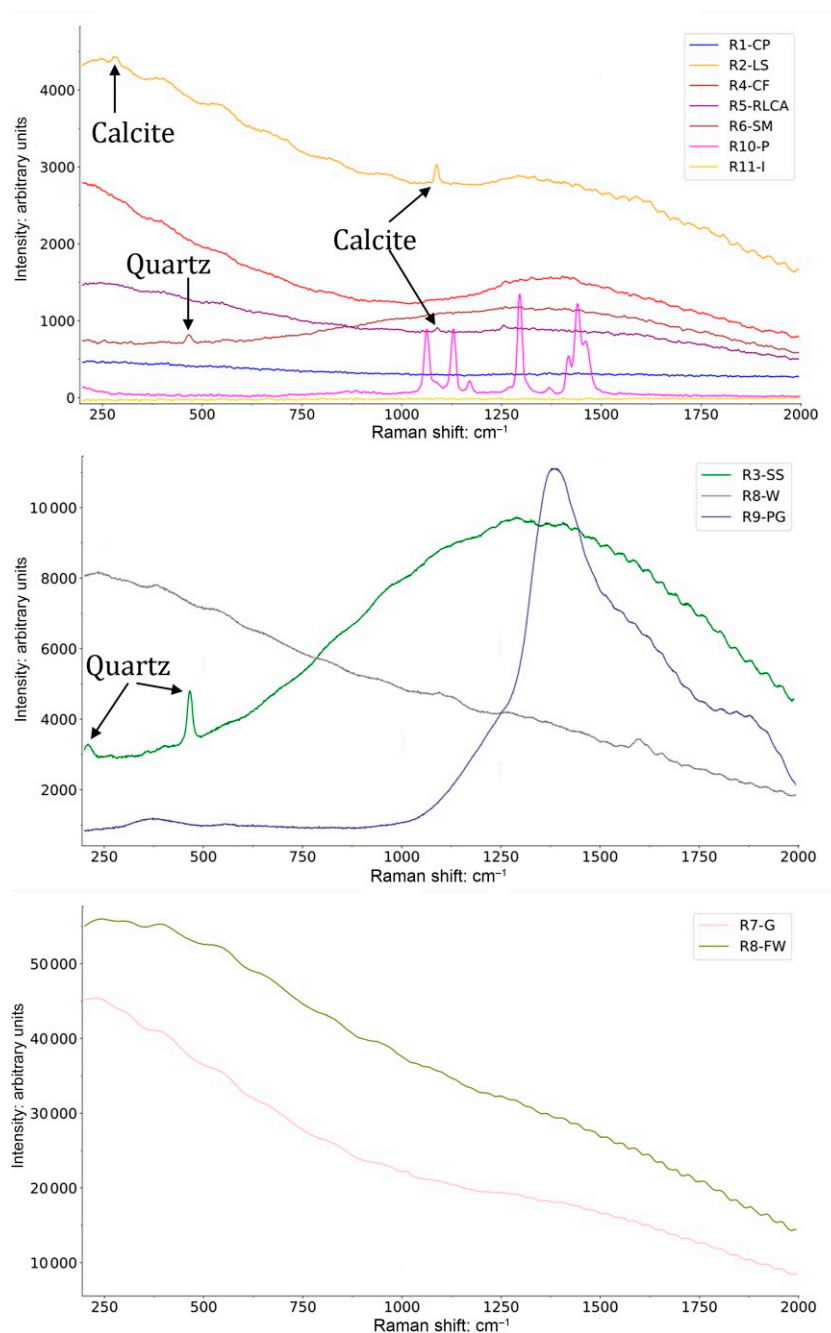


Figure 4. Average Raman spectra without pre-processing of all materials (R1-CP, cement paste; R2-LS, limestone sand; R3-SS, siliceous sand; R4-CF, ceramic fraction; R5-RLCA, recycled limestone concrete aggregate; R6-SM, siliceous mortar; R7-G, gypsum; R8-FW, wood fibres; R8-W, wood chips; R9-GP, glass powder; R10-P, polyethylene; R11-I, iron)

3.2.2 LIBS

The average LIBS spectra are presented in Figure 5. Inorganic materials exhibit similar peak distributions, suggesting comparable elemental compositions, with variations in line intensity reflecting differences in concentrations. These observations are consistent with

XRF data, which served to validate the elemental profiles obtained by LIBS. On the other hand, organic samples lack prominent emission lines because they are composed mainly of carbon, hydrogen and oxygen. Woods show two spectral lines close to 400 nm that may be associated with the presence of calcium. Its origin could be natural or due

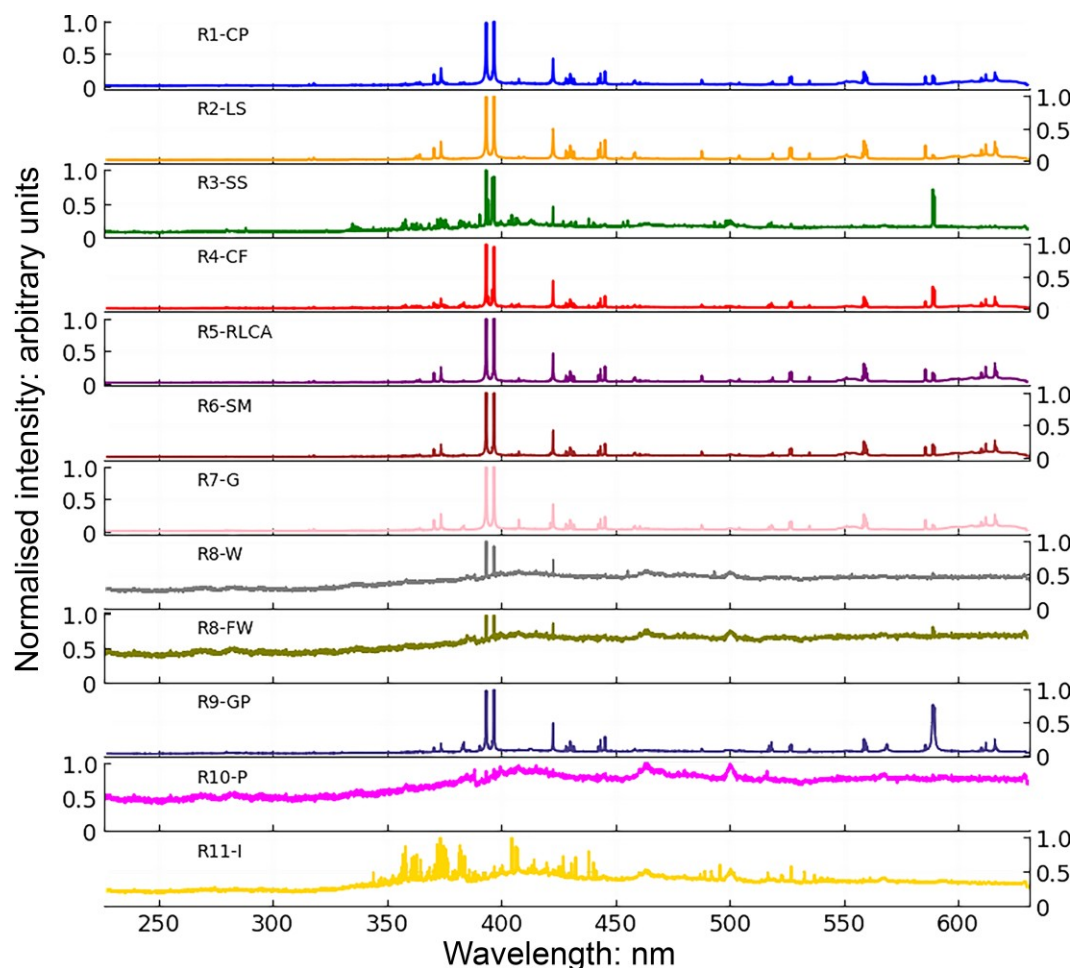


Figure 5. Mean laser-induced breakdown spectroscopy spectra for each material after min–max normalisation (R1-CP, cement paste; R2-LS, limestone sand; R3-SS, siliceous sand; R4-CF, ceramic fraction; R5-RLCA, recycled limestone concrete aggregate; R6-SM, siliceous mortar; R7-G, gypsum; R8-FW, wood fibres; R8-W, wood chips; R9-GP, glass powder; R10-P, polyethylene; R11-I, iron)

to treatments or contamination of the samples. A detailed peak analysis is not included, as the focus of the study is material differentiation rather than quantification or detection of specific elements.

3.2.3 UV–Vis

UV–Vis reflectance spectra of the materials in the wavelength range 190–700 nm are presented in Figure 6. In the range 380–700 nm, the diffuse reflectance spectrum is directly related to the perceived hue of the samples. Materials with a white or greyish hue have a uniform reflectance across this wavelength range such as R1-CP, R2-LS, R5-RLCA, R6-SM, R7-G, R9-GP and R10-P. In contrast, R4-CF, R8-W and R8-FW demonstrate a positive gradient with increasing wavelength, which is correlated with its brownish and reddish hue. In the ultraviolet range (<380 nm), these materials exhibit bands and changes that can be used to differentiate them. These results indicate that UV–Vis is useful for identifying visually distinct materials based on

diffused reflectance behaviour, but its capacity to resolve subtle compositional differences is limited compared with other techniques applied.

3.2.4 NIR

The NIR reflectance spectra of the materials in the wavelength range 1000–1700 nm are presented in Figure 7. While many mineral-based materials display similar, smooth spectral profiles, those containing hydrated phases or organic compounds – such as R1-CP, R5-RLCA, R6-SM, R7-G, R8-W, R8-FW and R10-P – exhibit distinctive absorption features. Cement-based samples (R1-CP, R5-RLCA and R6-SM) exhibited features near 1400 nm, attributed to the first overtone of O–H stretching in hydration products such as portlandite and calcium silicate hydrate (Kotthaus *et al.*, 2014; Ptacek *et al.*, 2021). Gypsum (R7-G) displays distinct absorption bands near 1440 nm and 1750 nm, corresponding to O–H overtones of the water molecules in $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$ (Moreira *et al.*, 2014). Wood samples (R8-W and

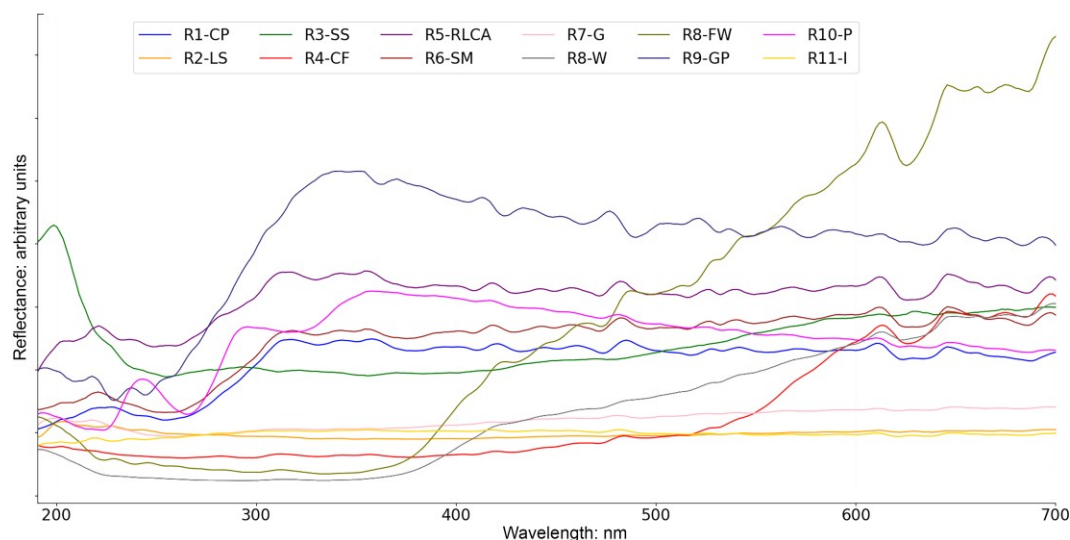


Figure 6. Average ultraviolet–visible reflectance spectra of the materials under study, after dark correction and normalisation to a white polytetrafluoroethylene reference (R1-CP, cement paste; R2-LS, limestone sand; R3-SS, siliceous sand; R4-CF, ceramic fraction; R5-RLCA, recycled limestone concrete aggregate; R6-SM, siliceous mortar; R7-G, gypsum; R8-FW, wood fibres; R8-W, wood chips; R9-GP, glass powder; R10-P, polyethylene; R11-I, iron)

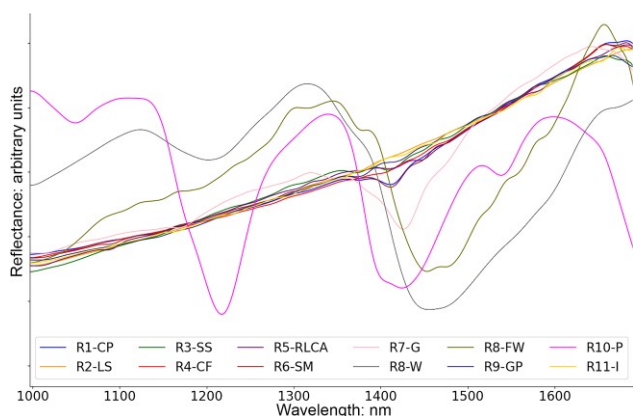


Figure 7. Average near infrared reflectance spectra of the materials under study, after dark correction and normalisation to a white Teflon reference

R8-FW) show a broad absorption band that extends from 1300 nm to nearly 1650 nm, related to O–H and C–H vibrations in cellulose and lignin (Leblon *et al.*, 2013). Polyethylene (R10-P) showed bands at 1150–1210 nm, ≈ 1390 nm and ≈ 1730 nm, corresponding to aliphatic C–H overtones (Mizushima *et al.*, 2012). In contrast, samples like R2-LS, R3-SS and R9-GP exhibited relatively flat reflectance, with no sharp absorption features – consistent with the low NIR activity in this range. Ceramic materials (R4-CF) show modest variability, although no diagnostic bands are observed between 1000 nm and 1700 nm (Kotthaus *et al.*, 2014). Iron (R11-I) was also featureless, as its primary absorptions occur

near 900 nm, falling outside the measured window. These observations confirm that the 1000–1700 nm region provides spectral variability particularly for materials containing hydrated (O–H) or organic functional groups (C–H). However, discrimination among purely mineral or metallic components is limited, as their spectra lack strong overtone bands. This limitation is especially relevant for materials with similar mineralogical composition, as confirmed by complementary XRF and XRD analyses. Whether more subtle spectral variations can be detected through CNNs will be examined in the next section.

3.3 Classification models

The classification of CDW materials was carried out using CNN models trained on the spectral data sets. For each technique, the impact of varying the number of convolutional layers (2–5) and the application of PCA-based outlier detection in the train subset was evaluated. Model performance was assessed using the m-F1, which reflects the overall balance between precision and recall, and confusion matrices, which provide detailed insight into class-specific performance. Table 2 presents the m-F1 scores for all model configurations, while the confusion matrices (Figure 8) were generated using the best performing model for each technique and applied to the corresponding test data set, thereby evaluating the robustness and generalisation capability of the models on previously unseen data obtained under controlled experimental conditions.

Across all modalities, performance on the independent test subsets is consistently high under the present controlled acquisition

Table 2. Micro-F1 score by model and outlier setting

Model	Outliers	CNN2	CNN3	CNN4	CNN5
Raman	—	0.982	0.986	0.989	0.989
	O	0.980	0.983	0.983	0.986
LIBS	—	0.997	0.998	0.999	0.998
	O	0.996	0.994	0.997	0.998
UV-Vis	—	0.989	0.990	0.988	0.984
	O	0.975	0.983	0.976	0.982
NIR	—	0.853	0.869	0.857	0.854
	O	0.852	0.865	0.851	0.868

Note(s): micro-F1 score obtained for each model and spectroscopic technique without (—) or with (O) outlier detection; values in italics are the best metrics achieved for each technique; LIBS, laser-induced breakdown spectroscopy; UV-Vis, ultraviolet-visual; NIR, near infrared

conditions, indicating that the spectral variability captured by the platform is sufficient for reliable class discrimination. The comparison of CNN depths (CNN2–CNN5) shows only marginal

differences for Raman, LIBS and UV-Vis, suggesting that additional model complexity offers limited benefit once the task is well constrained. Likewise, PCA-based outlier detection does not provide systematic gains across techniques and architectures, with performance remaining comparable (and in some cases slightly reduced), indicating that the measured spectra are already largely consistent after preprocessing. Consequently, the discussion below focuses on the few residual confusions visible in the confusion matrices and on the practical boundary conditions for transferability beyond controlled settings.

For LIBS, the best performance was obtained with CNN4 without PCA-based outlier detection (m-F1 = 0.999). The confusion matrix (see Figure 8(a)) shows near-perfect separation on the independent test set, with only a few isolated confusions: mainly within the organic subset (R8-W → R8-FW; R10-P → R8-FW) and a single confusion between two cementitious classes

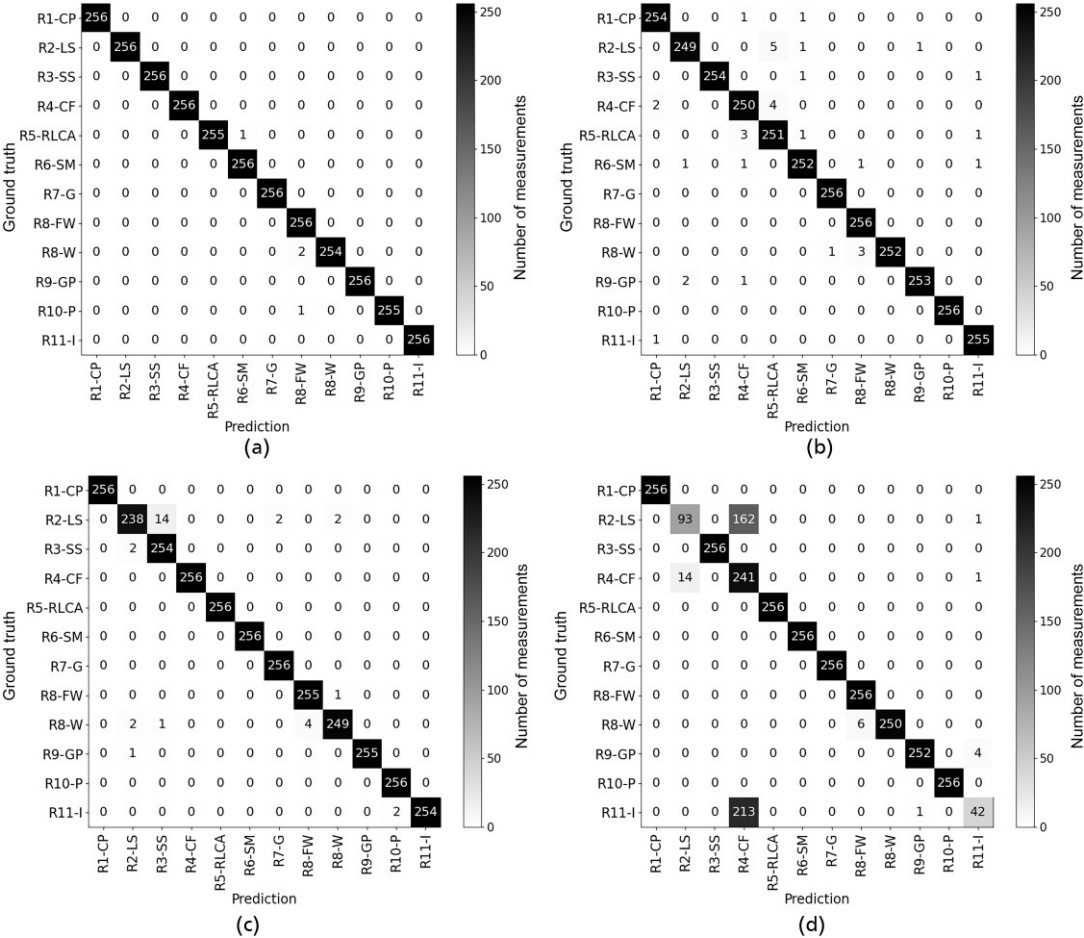


Figure 8. Confusion matrices obtained by applying the independently trained models to their corresponding test subsets: (a) laser-induced breakdown spectroscopy, (b) Raman, (c) ultraviolet-visual and (d) near infrared (R1-CP, cement paste; R2-LS, limestone sand; R3-SS, siliceous sand; R4-CF, ceramic fraction; R5-RLCA, recycled limestone concrete aggregate; R6-SM, siliceous mortar; R7-G, gypsum; R8-FW, wood fibres; R8-W, wood chips; R9-GP, glass powder; R10-P, polyethylene; R11-I, iron)

(R5-RLCA → R6-SM). Regarding organic materials, both wood and polyethylene (PE) are dominated by light elements (C–H–O for wood; predominantly C–H for PE), for which commonly exploited atomic lines lie outside the analysed window [260, 610] nm (e.g. H I $\approx$ 656 nm, O I $\approx$ 777 nm). Nonetheless, the observed performance indicates that the model extracts sufficient discriminative information from the measured range, primarily through distributed spectral cues such as relative intensity patterns, line-to-continuum balance and baseline/continuum behaviour. Weak carbon-nitrogen bounds ($\approx$ 385 to 388 nm) and carbon-carbon bounds (e.g., $\approx$ 516.5 nm) emissions may contribute but are of low contrast. For inorganic classes, the near-perfect performance of LIBS is consistent with the compositional contrasts evidenced by XRF (Figure 3), insofar as LIBS is an elemental emission technique whose spectra encode differences in elemental abundances through class-dependent patterns of emission lines and their relative intensities. Under the present acquisition conditions, discrimination does not necessarily require the appearance of entirely new peaks; rather, it can be supported by reproducible differences in relative line-intensity distributions and line-to-continuum behaviour across the measured window, which are influenced by both composition and plasma conditions (e.g. temperature/density-dependent excitation and matrix-related effects).

For the UV–Vis technique, the best classification performance was achieved using a CNN with three convolutional layers, reaching a m-F1 score of 0.990. The confusion matrix (see Figure 8(c)) indicates high accuracy and sensitivity for most classes, showing that UV–Vis spectra provide sufficient discriminative information under control acquisition conditions. This discriminative information is primarily linked to diffuse reflectance behaviour associated with sample hue in the visible range, complemented in some materials by a limited number of additional UV–Vis spectral features, which together support the separation of most classes. However, some misclassifications remain (e.g., R2-LS versus R3-SS and R8-W versus R8-FW), consistent with partially overlapping optical responses. A small fraction of these residual errors may also reflect point-to-point sampling variability and occasional acquisition artefacts (e.g., local surface heterogeneity or edge effects), which can subtly alter reflectance-based cues.

The Raman technique reaches its maximum m-F1 value (0.989) when using four or five convolutional layers. The confusion matrix derived from the independent test set (Figure 8(b)), obtained with CNN4, confirms the stability and robustness of the model when applied to unseen spectra acquired under the same experimental conditions. Nevertheless, a limited number of misclassifications occurred between spectrally similar material classes (e.g., R4-CF versus R5-RLCA, R8-W versus R8-FW or R2-LS versus R5-RLCA). In the preliminary inspection of the spectra (see Section 3.2 above), not all materials exhibited clearly

distinctive Raman bands under the present acquisition conditions, partly due to shared mineral phases and partly due to measurement-related factors such as fluorescence background and limited S/N ratio, which can obscure weaker Raman features. The strong performance of the Raman convolutional models indicates that classification is not exclusively driven by isolated characteristic Raman bands. Instead, 1D-CNN is likely exploiting distributed spectral information, including subtle variations in spectral shape, baseline, fluorescence background and local intensity relationships along the spectral axis (Liu *et al.*, 2017). Such features may arise from matrix effects, microstructural differences or phase composition which, while not easily interpretable in terms of individual Raman assignments, remain physically consistent and reproducible under controlled experimental conditions. This likely explains why the remaining misclassifications occur mainly in cases where distinctive Raman bands are absent or weak, or where characteristic features overlap across different material matrices.

Unlike the other techniques, NIR yields the lowest overall classification performance. The highest m-F1 score obtained was 0.869, achieved using a CNN3 model. The confusion matrix (see Figure 8(d)) reveals clear differentiation for materials exhibiting characteristic bands, such as organic compounds (R8-W, R8-FW, R10-P) and those with hydrated phases (R1-CP, R5-RLCA, R7-G). In addition, R3-SS and R9-GP are still satisfactorily differentiated despite lacking prominent NIR absorption bands, and R1-CP and R6-SM can also be separated despite sharing the $\approx$ 1400 nm feature, suggesting that the model exploits subtle, distributed spectral variations across the measured range (e.g., local shape differences and weak baseline/reflectance cues) rather than relying exclusively on a small number of distinctive bands. The few residual errors occur between classes characterised by relatively flat, feature-poor responses in the analysed window, particularly R4-CF versus R11-I and R4-CF versus R2-LS, where discrimination relies mainly on subtle slope/level differences rather than on diagnostic absorption bands. Notably, R11-I is confused with R4-CF rather than with R2-LS, which is consistent with the XRF results indicating a higher iron content for R4-CF than for R2-LS; this relative iron-richness may contribute to reduced spectral contrast and increased overlap in this wavelength range. These outcomes are consistent with the physics of NIR reflectance: the 1000–1700 nm region is dominated by relatively weak overtone features, while more diagnostically informative combination bands typically occur at longer wavelengths ($>$ 1700 nm). Extending the spectral coverage beyond 1700 nm would therefore be expected to improve discrimination among the currently ambiguous mineral classes.

From an industrial perspective, the strong classification performance reported here should be interpreted within the context of controlled acquisition conditions. In real CDW processing environments, several factors may challenge direct transferability,

including moisture variations, surface contamination or dust, mixed-material contact and particle overlap within the measurement spot, as well as process-related constraints such as conveyor speed and reduced acquisition time. These effects can alter signal stability, reduce S/N ratio or introduce composite spectral responses that differ from laboratory-prepared samples. In addition, long-term deployment requires regular sensor maintenance and calibration to mitigate instrumental drift and ensure consistent data quality. While the multi-sensor approach adopted here offers complementary and partially redundant information that may help mitigate some technique-specific limitations, further validation on heterogeneous, dynamically evolving CDW streams is required to fully assess robustness and operational boundaries in industrial settings. Within these boundaries, the proposed framework shows strong potential as a foundation for more intelligent and efficient CDW processing systems.

4. Conclusions

This study assessed the potential of four in-line spectroscopic techniques (LIBS, Raman, UV–Vis and NIR), combined with 1D-CNN models, for classifying representative materials from the fine fraction (<4 mm) of CDW. Using 12 controlled material classes measured under standardised laboratory conditions, all techniques achieved good performance on independent test subsets, although with clear differences between modalities.

LIBS provided the highest performance (m-F1 = 0.999), followed by UV–Vis (m-F1 = 0.990) and Raman (m-F1 = 0.989), showing that these three techniques can reliably discriminate most classes under the present acquisition conditions. NIR showed the lowest overall performance (m-F1 = 0.869) but remained effective for materials exhibiting O–H and/or C–H related spectral structure, while most errors were concentrated in a small subset of feature-poor inorganic classes. Overall, the results confirm that the proposed multi-sensor framework provides a strong basis for in-line classification of fine CDW, while also indicating that the four modalities are complementary rather than interchangeable.

Further work is required to validate the models on heterogeneous and mixed CDW streams, assess robustness under more realistic processing conditions and explore sensor-fusion strategies to improve transferability to industrial environments.

CRedit authorship contribution

C.F.-B.: writing – original draft, review and editing, conceptualisation, methodology, formal analysis, data curation, visualisation; V.G.-C.: writing – original draft, review and editing, conceptualisation, methodology, formal analysis, data curation; J.A.I.: writing – original draft, review and editing, conceptualisation, methodology; P.G.: writing – original draft, review and editing, conceptualisation, methodology; and L.B.d.V.: writing – original draft, review and editing, conceptualisation, methodology.

Data availability

The data and code underlying this study are not publicly available as they were acquired and developed within the framework of a proprietary research project. Restrictions apply to their availability. Further information and access requests may be directed to the corresponding author.

Funding

This research has been supported by the Elkartek Programme, Basque Government (Spain) (SMART-EYE, KK-2023/00021) and the CIRCULess EU H2020 project (Grant no. 101138220). The dataset used in this work was generated within the framework of the SMART-EYE project, while the application and evaluation of the models were carried out as part of the CIRCULess project. The authors' views are reflected in this work, and the commission is not responsible for its content.

REFERENCES

- Buzgar N, Buzatu A and Sanislav IV (2009) The Raman study on certain sulfates. *Analele Stiintifice Ale Universitatii "Al. I. Cuza" Din Iasi, Geologie* **55**(1): 5–23.
- Chang C, Di Maio F, Bheemireddy R et al. (2025a) Enhancing quality inspection efficiency and reliability of unscreened recycled coarse aggregates (RCA) streams using innovative mobile sensor-based technology. *Developments in the Built Environment* **21**: 100611, [10.1016/j.dibe.2025.100611](https://doi.org/10.1016/j.dibe.2025.100611).
- Chang C, Di Maio F, Bheemireddy R et al. (2025b) Rapid quality control for recycled coarse aggregates (RCA) streams: Multi-sensor integration for advanced contaminant detection. *Computers in Industry* **164**: 104196, [10.1016/j.compind.2024.104196](https://doi.org/10.1016/j.compind.2024.104196).
- Cristobal-García J, Caro D, Foster G et al. (2024) *Techno-economic and Environmental Assessment of Construction and Demolition Waste Management in the European Union: Status Quo and Prospective Potential*. Publications Office of the European Union, Luxembourg. [10.2760/721895](https://doi.org/10.2760/721895).
- Furukawa T, Sato H, Kita Y et al. (2006) Molecular structure, crystallinity and morphology of polyethylene/polypropylene blends studied by Raman mapping, scanning electron microscopy, wide angle X-ray diffraction, and differential scanning calorimetry. *Polymer Journal* **38**(11): 1127–1136, [10.1295/polymj.PJ2006056](https://doi.org/10.1295/polymj.PJ2006056).
- Gunasekaran S, Anbalagan G and Pandi S (2006) Raman and infrared spectra of carbonates of calcite structure. *Journal of Raman Spectroscopy* **37**(9): 892–899, [10.1002/jrs.1518](https://doi.org/10.1002/jrs.1518).
- Kotthaus S, Smith TEL, Wooster MJ and Grimmond CSB (2014) Derivation of an urban materials spectral library through emittance and reflectance spectroscopy. *ISPRS Journal of Photogrammetry and Remote Sensing* **94**: 194–212, [10.1016/j.isprsjprs.2014.05.005](https://doi.org/10.1016/j.isprsjprs.2014.05.005).
- Kroell N, Chen X, Greiff K and Feil A (2022) Optical sensors and machine learning algorithms in sensor-based material flow characterization for mechanical recycling processes: a systematic literature review. *Waste Management* **149**: 259–290, [10.1016/j.wasman.2022.05.015](https://doi.org/10.1016/j.wasman.2022.05.015).
- Leblon B, Adedipe O, Hans G et al. (2013) A review of near-infrared spectroscopy for monitoring moisture content and density of solid wood. *The Forestry Chronicle* **89**(05): 595–606, [10.5558/tfc2013-111](https://doi.org/10.5558/tfc2013-111).
- Liu J, Osadchy M, Ashton L et al. (2017) Deep convolutional neural networks for Raman spectrum recognition: a unified solution. *The Analyst* **142**(21): 4067–4074, [10.1039/C7AN01371J](https://doi.org/10.1039/C7AN01371J).

- Liu H, Kaya H, Lin YT, Ogrinc A and Kim SH (2022) Vibrational spectroscopy analysis of silica and silicate glass networks. *Journal of the American Ceramic Society* **105**(4): 2355–2384, [10.1111/jace.18206](https://doi.org/10.1111/jace.18206).
- Marín-Cortés S, Fernández-Álvarez M, Enríquez E and Fernández JF (2024) Experimental characterization data on aggregates from construction and demolition wastes for the assistance in sorting and recycling practices. *Construction and Building Materials* **435**: 136798, [10.1016/j.conbuildmat.2024.136798](https://doi.org/10.1016/j.conbuildmat.2024.136798).
- Marín-Cortés S, Fernández-Álvarez M, Moure A, Fernández JF and Enríquez E (2023) Chemometric-driven quantification of construction and demolition waste using Raman spectroscopy and SWIR: enhancing sustainability in the ceramic sector. *Resources, Conservation and Recycling* **199**: 107259, [10.1016/j.resconrec.2023.107259](https://doi.org/10.1016/j.resconrec.2023.107259).
- Mizushima M, Kawamura T, Takahashi K and Nitta KH (2012) In situ near-infrared spectroscopic studies of the structural changes of polyethylene during melting. *Polymer Journal* **44**(2): 162–166, [10.1038/pj.2011.100](https://doi.org/10.1038/pj.2011.100).
- Moreira LCJ, Teixeira AdS and Galvão LS (2014) Laboratory salinization of Brazilian alluvial soils and the spectral effects of gypsum. *Remote Sensing* **6**(4): 2647–2663, [10.3390/rs6042647](https://doi.org/10.3390/rs6042647).
- Moreno-Juez J, Vegas J, Gebremariam AT, García-Cortés V and Maio F (2020) Treatment of end-of-life concrete in an innovative heating-air classification system for circular cement-based products. *Journal of Cleaner Production* **263**: 121515, [10.1016/j.jclepro.2020.121515](https://doi.org/10.1016/j.jclepro.2020.121515).
- Picon A, Galan P, Bereciartua-Perez A and Benito-del-Valle L (2025) On the analysis of adapting deep learning methods to hyperspectral imaging. Use case for WEEE recycling and dataset. *Spectrochimica Acta. Part A, Molecular and Biomolecular Spectroscopy* **330**: 125665, [10.1016/j.saa.2024.125665](https://doi.org/10.1016/j.saa.2024.125665).
- Ptacek L, Strauss A, Hinterstoisser B and Zitek A (2021) Curing assessment of concrete with hyperspectral imaging. *Materials* **14**(14): 3848, [10.3390/ma14143848](https://doi.org/10.3390/ma14143848).
- Ta S, Alizadeh S, Samavedham L and Ray AK (2023) Deep chemometrics using one-dimensional convolutional neural networks for predicting crude oil properties from FTIR spectral data. *The Canadian Journal of Chemical Engineering* **101**(12): 6688–6700, [10.1002/cjce.25076](https://doi.org/10.1002/cjce.25076).
- Tuschel D (2016) Selecting an excitation wavelength for Raman spectroscopy. *Spectroscopy* **31**(3): 14–23.
- Vegas I, Broos K, Nielsen P, Lambertz O and Lisbona A (2015) Upgrading the quality of mixed recycled aggregates from construction and demolition waste by using near-infrared sorting technology. *Construction and Building Materials* **75**: 121–128, [10.1016/j.conbuildmat.2014.09.109](https://doi.org/10.1016/j.conbuildmat.2014.09.109).
- Vítek S, Zbírál T and Nežerka V (2025) Using hyperspectral imaging to identify optimal narrowband filter parameters for construction and demolition waste classification. *Resources, Conservation and Recycling* **215**: 108123, [10.1016/j.resconrec.2025.108123](https://doi.org/10.1016/j.resconrec.2025.108123).
- Wu TW, Zhang H, Peng W, Lü F and He PJ (2023) Applications of convolutional neural networks for intelligent waste identification and recycling: a review. *Resources, Conservation and Recycling* **190**: 106813, [10.1016/j.resconrec.2022.106813](https://doi.org/10.1016/j.resconrec.2022.106813).
- Xiao W, Yang J, Fang H, Zhuang J and Ku Y (2019) Development of online classification system for construction waste based on industrial camera and hyperspectral camera. *PloS One* **14**(1): e0208706–16, [10.1371/journal.pone.0208706](https://doi.org/10.1371/journal.pone.0208706).

How can you contribute?

To discuss this paper, please email up to 500 words to the editor at support@emerald.com. Your contribution will be forwarded to the author(s) for a reply and, if considered appropriate by the editorial board, it will be published as discussion in a future issue of the journal.

Proceedings journals rely entirely on contributions from the civil engineering profession (and allied disciplines). Information about how to submit your paper online is available at www.emeraldgrouppublishing.com/journal/jcoma, where you will also find detailed author guidelines.