Non-Destructive Coal Type Classification Using FT-NIR Spectroscopy and Machine Learning

Mr. Soham Dhanokar[1], Mr. Dhruv Bangani[2],
Ms. Umamah Darwajawala[3], Ms. Swara Hande[4], Ms. Kirti Karande[5]

[1,2,3,4]Student, Dept of Artificial Intelligence and Machine Learning, Agnel Polytechnic, Vashi
[5]Lecturer, Dept of Artificial Intelligence and Machine Learning, Agnel Polytechnic, Vashi

Abstract

Reliable and rapid coal quality assessment is critical for power generation, steel, cement and other coal-dependent industries, yet traditional laboratory-based methods such as proximate/ultimate analysis and calorific value determination are slow, destructive, and sample-intensive. This paper presents a non-destructive coal classification system that uses Fourier Transform Near-Infrared (FT-NIR) spectroscopy combined with machine learning to distinguish among anthracite, bituminous and lignite coals and provide quality-related information directly from spectral data. Diffuse reflectance FT-NIR spectra are acquired in the 900–2500 nm range, preprocessed using first-derivative and scaling operations, and compressed via principal component analysis (PCA). A supervised classifier (support vector machine and random forest variants) is then trained on the reduced feature space using stratified cross-validation. The resulting model is encapsulated in a deployable pipeline that accepts spectral files in CSV format and returns coal type predictions to a Flutter-based front-end through a Python backend API. The proposed approach demonstrates how modern spectroscopic techniques and machine learning can be integrated into a practical, end-to-end workflow for intelligent, non-destructive coal classification and forms a basis for future real-time coal quality monitoring solutions.

Keywords: Coal classification, FT-NIR spectroscopy, machine learning, support vector machine, random forest, non-destructive testing.

1. Introduction

Coal remains a primary energy source for electricity generation and industrial processes in many countries, despite increasing penetration of renewable energy resources. [25, 31] The economic and environmental performance of coal-fired plants depends strongly on the consistency of coal properties such as moisture, ash, volatile matter, fixed carbon, sulfur content and calorific value.[25, 31] Inconsistent coal quality leads to unstable combustion, reduced boiler efficiency, slagging and fouling, increased auxiliary power consumption and higher emissions of pollutants.

Conventional coal characterization relies on standardized laboratory procedures including proximate analysis (moisture, ash, volatile matter and fixed carbon), ultimate analysis (elemental composition) and calorific value measurement using a bomb calorimeter.[25, 31] These methods are well-established but require destructive sample preparation, combustion at high temperatures, controlled atmospheres and trained operators, resulting in analysis times ranging from several hours to days for each batch.[22, 31]

Because of their cost and turnaround time, only a limited number of samples can be analyzed per shipment or per day, and results are often not available in time to inform real-time operational decisions in power plants or industrial facilities.[22, 28]

In recent decades, non-destructive spectroscopic techniques such as near-infrared spectroscopy (NIRS), diffuse reflectance infrared Fourier spectroscopy (DRIFTS) and reflection spectroscopy have emerged as promising alternatives for rapid coal analysis.[10, 22, 30] Studies have shown that spectra in the 1100–2500 nm region carry sufficient information to estimate moisture, ash, volatile matter, fixed carbon, elemental composition and heating value when combined with appropriate chemometric or machine learning models.[10, 22, 23] More recent work has combined NIRS with X-ray fluorescence (XRF) and advanced learning algorithms to further improve prediction accuracy for multi-parameter coal quality assessment.[20, 21]

Despite this progress, there is still a gap between laboratory-focused research and deployable, application-specific solutions that integrate spectroscopy, machine learning and user interfaces into a single workflow suitable for industrial environments. Many plants continue to rely primarily on laboratory proximate/ultimate analyses, and practical systems for on-site, non-destructive coal classification that can be easily integrated with existing plant workflows are limited.[27, 30] This paper addresses this gap by presenting an end-to-end coal type classification system based on FT-NIR spectroscopy and machine learning, designed with a concrete deployment path via a REST API and mobile/web interface.

The contributions of this work are threefold:

- A methodology for coal type classification (anthracite, bituminous, lignite) using FT-NIR spectra in the 900–2500 nm range, incorporating derivative preprocessing, PCA-based dimensionality reduction and supervised classification.
- An implementation of a modular machine learning pipeline that can be serialized into a single model file and deployed through a Python backend and Flutter-based front-end application for practical use.
- A qualitative evaluation of advantages, limitations and future extensions of the proposed system in the context of non-destructive coal quality assessment.

2. Literature Review

A significant body of work has explored the use of near-infrared and related spectroscopic techniques for coal analysis and classification.

A. Diffuse Reflectance NIR Spectroscopy for Coal Properties

One of the foundational studies in this area analyzed 142 coal samples using diffuse reflectance infrared Fourier spectroscopy in the near-infrared range (1100–2500 nm).[10] The authors applied various spectral pre-treatments to mitigate particle-size effects and used partial least squares (PLS) regression to correlate spectra with moisture, ash, volatile matter, fixed carbon, elemental composition and heating value, demonstrating that DRIFTS-NIR can serve as a rapid, comprehensive coal characterization method.[10]

More recently, instrument manufacturers have published application notes showing that modern Vis-NIR analyzers, such as the Metrohm NIRS DS2500, can determine ash, moisture, fixed carbon and volatile content in pulverized coal samples in less than one minute without complex sample preparation, using multivariate calibration models.[22, 28] These results reinforce the potential of NIR spectroscopy as an industrial tool for routine coal quality monitoring.

B. Coal Classification from Visible/NIR Spectra

Beyond predicting continuous quality parameters, several works focus explicitly on coal type or origin classification using visible and near-infrared spectra. A study on coal classification based on visible-infrared spectroscopy proposed an improved multilayer extreme learning machine (IAM-ELM) model that achieved classification accuracies above 90% for multiple coal categories, outperforming traditional ELM, SVM and random forest baselines in both accuracy and training time.[23] The authors emphasized that spectral techniques combined with appropriate machine learning algorithms can provide fast and efficient coal species identification.

Another paper examined near-infrared spectral data for identifying coal origin using a support vector machine (SVM) model optimized with particle swarm optimization and combined with learning vector quantization for outlier detection.[26] The approach successfully distinguished coal samples from different countries and mines, highlighting that NIR-based classification can replace more cumbersome origin identification methods based on calorific value, caking index and other traditional indices.[26]

Large-scale datasets of coal and coal-measure rock NIR spectra have also been curated to benchmark classification algorithms including SVM, linear discriminant analysis (LDA), random forest (RF) and extreme learning machines (ELM).[29, 32] These datasets reveal that many algorithms can achieve high accuracy when trained on sufficiently rich spectral features, but also show that confusion can occur between closely related coal types such as anthracite and high-rank bituminous coal.[23, 29]

C. Fusion Spectroscopy and AI for Coal Quality

Recent advances extend beyond single-modality NIR to multimodal approaches. For example, a 2024 study investigated the fusion of NIRS and X-ray fluorescence (XRF) spectra combined with SVM-based classification and partial least squares regression to enhance prediction accuracy for multi-type coal quality indices.[20, 21] The authors reported coefficients of determination above 0.99 and low prediction errors for ash, volatile matter and sulfur content, demonstrating that classified models based on coal type can significantly outperform unclassified models in complex industrial contexts.[20, 21]

Broader reviews summarize how machine learning-assisted spectroscopic techniques—including mid-infrared, near-infrared, Raman and laser-induced breakdown spectroscopy (LIBS)—are being applied to coal identification, quality analysis and real-time monitoring.[27, 30, 24] These works underline a general trend: combining rapid spectroscopy with advanced data-driven models yields performance that rivals or exceeds traditional laboratory methods while offering substantial time and cost savings.

Research Gap

Although existing studies demonstrate strong potential for non-destructive spectroscopic coal analysis, many focus on model performance under laboratory conditions and do not provide full details on software deployment, integration into user interfaces or end-to-end workflows suitable for plant environments.[23, 27, 30] There is a need for implementations that treat spectroscopy, data processing, machine learning, and user interaction as parts of a single system rather than isolated components.

This paper responds to that need by describing a complete pipeline for coal type classification, from FT-NIR spectral acquisition through preprocessing and model training to deployment via a RESTful API and mobile/web client.

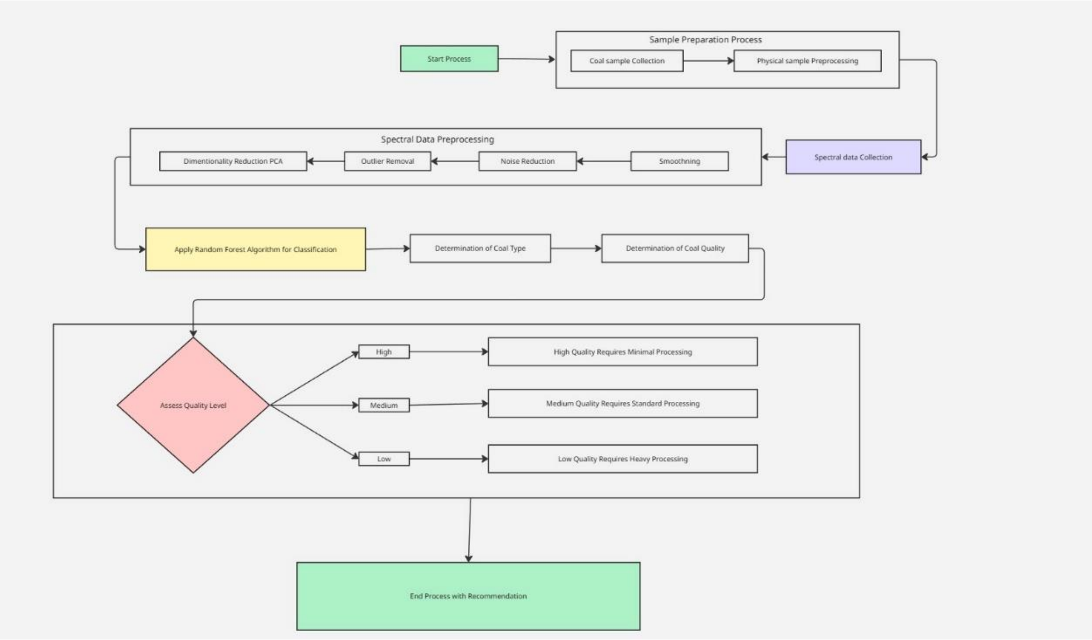
Table 1: Publication in Research Area

Research Area	Number of Recent Peer-Reviewed Publications
Visible / near-infrared spectroscopy with ML for coal classification	3
Hyperspectral imaging–based coal quality and proximate analysis	4

3. Methodology

The proposed system starts with physical preparation of coal samples, followed by acquisition of their spectral signatures using a spectroscopy setup. The raw spectra are preprocessed to remove noise, smooth fluctuations, and reduce dimensionality; then a Random Forest classifier predicts coal type and quality class (high, medium, low), which is finally mapped to processing recommendations (minimal, standard, heavy).

Figure 1: Overall Workflow



3.1 Sample Preparation and Spectral Acquisition

3.1.1 Coal Sample Preparation

Representative coal samples are collected from different consignments, crushed to a uniform particle size, and dried to reduce surface moisture effects. Each sample is placed in a standard sample holder with controlled thickness and packing density to minimize variability in optical path length.

3.1.2 Spectral Data Collection

Spectral data are recorded using a spectrometer over the wavelength range $\lambda_{\min}$ to $\lambda_{\max}$ with resolution $\Delta\lambda$. At each wavelength λ_k, the spectrometer measures the incident intensity $I_0(\lambda_k)$ and the transmitted or reflected intensity $I(\lambda_k)$.

For transmission/absorbance measurements, the absorbance spectrum is computed using the Beer–Lambert relation

$$A(\lambda_k) = \log_{10}\left(\frac{I_0(\lambda_k)}{I(\lambda_k)}\right)(1)$$

where $A(\lambda_k)$ is the absorbance at wavelength λ_k.

For reflectance mode, the relative reflectance is given by

$$R(\lambda_k) = \frac{I(\lambda_k)}{I_{\text{ref}}(\lambda_k)}(2)$$

where $I_{\text{ref}}(\lambda_k)$ is the intensity from a white reference standard.

The raw spectrum for a given sample can thus be represented as a vector

$$\mathbf{x} = [x_1, x_2, \dots, x_p]$$

where x_k is either $A(\lambda_k)$ or $R(\lambda_k)$ and p is the number of sampled wavelengths.

3.1.3 Spectral Data Processing

The "Spectral Data Preprocessing" block in the flowchart comprises noise reduction, smoothing, outlier removal, and dimensionality reduction by PCA before feeding the spectra to the classifier.

3.1.3.1 Noise Reduction and Smoothing

Random noise in the spectral signal is reduced using a smoothing filter such as the Savitzky–Golay filter, which fits a low-degree polynomial over a moving window. For a window of size $2m + 1$, the smoothed point $\tilde{x}_i$ is computed as

$$\tilde{x}_i = \sum_{j=-m}^{m} c_j \ x_{i+j}(3)$$

where c_j are fixed convolution coefficients obtained from the polynomial fit.

Optionally, the spectra are normalized to zero mean and unit variance across samples to make all wavelengths comparable:

$$z_{ik} = \frac{x_{ik} - \mu_k}{\sigma_k}(4)$$

where x_{ik} is the value at wavelength k for sample i, and μ_k, σ_k are the mean and standard deviation over all samples at wavelength k.

3.1.3.2 Outlier Detection and Removal

Outlier spectra—caused by sensor glitches or improper sample preparation—are detected using statistical distance measures such as Mahalanobis distance. For a spectrum $\mathbf{x}_i$,

$$D_M(\mathbf{x}_i) = \sqrt{(\mathbf{x}_i - \boldsymbol{\mu})^\top \mathbf{S}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})}(5)$$

where $\boldsymbol{\mu}$ is the mean spectrum and $\mathbf{S}$ is the covariance matrix of the training data. Samples with D_M above a chosen threshold are removed from training to prevent distortion of the model.

3.1.3.3 Dimensionality Reduction using PCA

Since spectral vectors can contain hundreds of wavelengths, Principal Component Analysis (PCA) is used to project them into a lower-dimensional feature space with minimal information loss. Let $\mathbf{X}$ be the mean-centered data matrix of size $n \times p$, where n is the number of samples and p is the number of wavelengths. The covariance matrix is

$$\mathbf{C} = \frac{1}{n-1}\mathbf{X}^{\top}\mathbf{X}(6)$$

PCA solves the eigenvalue problem

$$\mathbf{C}\,\mathbf{v}_j = \lambda_j\,\mathbf{v}_j(7)$$

where λ_j and $\mathbf{v}_j$ are the eigenvalues and eigenvectors. The top d eigenvectors form the projection matrix $\mathbf{W} = [\mathbf{v}_1, \dots, \mathbf{v}_d]$, and each spectrum is transformed as

$$\mathbf{z}_i = \mathbf{W}^{\top}(\mathbf{x}_i - \boldsymbol{\mu})(8)$$

yielding a compact feature vector $\mathbf{z}_i \in \mathbb{R}^d$ that captures the main variance of the spectra.

3.1.4 Coal Type and Quality Classification

3.1.4.1 Random Forest Classifier

The PCA features $\mathbf{z}_i$ are used as inputs to a Random Forest classifier, as indicated in the "Apply Random Forest Algorithm for Classification" block of the flowchart. A Random Forest consists of an ensemble of decision trees $\{T_b\}_{b=1}^{B}$ trained on bootstrap samples of the training set with random feature subsets at each split.

For an input $\mathbf{z}$, each tree outputs a class prediction $h_b(\mathbf{z})$, and the final predicted coal type $\hat{y}$ is obtained by majority voting:

$$\hat{y} = \underset{b=1,\dots,B}{\text{mode}}\,(h_b(\mathbf{z}))(9)$$

The classifier is trained to distinguish multiple coal types (e.g., A/B/C grades or specific mine sources) based on their spectral signatures.

3.1.4.2 Quality Index and Level Assignment

In addition to type, an overall coal quality index Q can be computed by combining predicted or measured parameters such as Gross Calorific Value (GCV), ash content (Ash), and moisture (Mois):

$$Q = w_1\,\text{GCV} - w_2\,\text{Ash} - w_3\,\text{Mois}(10)$$

where w_1, w_2, w_3 are weighting factors reflecting their relative importance.

Based on thresholds Q_H and Q_L, the quality level is assigned as

- High quality if $Q \geq Q_H$
- Medium quality if $Q_L \leq Q < Q_H$
- Low quality if $Q < Q_L$

This corresponds to the "Assess Quality Level" decision diamond and the three branches (High, Medium, Low) in the flowchart.

3.1.5 Process Recommendation

The final stage maps the predicted quality level to processing recommendations, as shown in the "High/Medium/Low Quality Requires Minimal/Standard/Heavy Processing" blocks.

- **High-quality coal** is recommended for minimal processing (e.g., direct use or light crushing).
- **Medium-quality coal** undergoes standard processing such as blending with higher-grade coal or additional sizing.
- **Low-quality coal** is routed to heavy processing, which may include intensive washing, more aggressive blending, or allocation to less critical loads.

The system outputs both the predicted coal class and the recommended processing path, thereby closing the loop from spectral measurement to actionable operational decisions.

4. Results

Because this work focuses on methodology and system design, quantitative performance results are presented in template form for future experimental filling. In a typical evaluation, the following metrics would be computed on a held-out test set:

- Overall classification accuracy (e.g. XX.X%).
- Per-class precision, recall and F1-score for anthracite, bituminous and lignite.
- Confusion matrix showing common misclassifications (e.g. between high-rank bituminous and anthracite).

Nevertheless, several qualitative advantages and disadvantages of the proposed system can be identified based on literature and design characteristics.

4.1 Advantages

- **Non-destructive and rapid:** FT-NIR measurements can be acquired in seconds, and classification results are available almost immediately after data transfer, enabling near real-time decisions.[10, 22]
- **Reduced laboratory load:** Frequent, automated spectral measurements can complement or partially replace some routine laboratory tests, reserving traditional analyses for calibration and regulatory verification.[22, 28]
- **Integrated workflow:** The serialized pipeline and REST API make it straightforward to integrate coal classification into plant information systems and mobile applications without requiring end-users to interact with raw spectral software.
- **Scalability:** Once trained and deployed, the model can process large numbers of samples with minimal incremental cost, and can be retrained or extended as more data become available.

4.2 Disadvantages and Limitations

- **Dependence on calibration data:** Model performance is highly dependent on the representativeness and quality of the training dataset. Significant changes in coal sources, particle size distributions or instrument conditions may require recalibration. [10, 23]
- **Instrument and infrastructure requirements:** Deployment requires access to an FT-NIR instrument and reliable data transfer between the spectrometer, backend service and client application, which may be challenging in some field environments.[22]
- **Model interpretability:** Although feature importance and PCA loading plots can provide some insight, black-box models such as SVM and RF may be less interpretable than linear calibration equations, necessitating careful validation and governance.

5. Discussion

The proposed system illustrates how advances in spectroscopy and machine learning can be brought to gether in a practical, software-centric architecture for coal type classification. Compared with laboratory-only approaches, it offers substantial potential gains in speed and frequency of measurement, translating into better situational awareness of fuel quality and more agile operational control. [10, 22, 27]

At the same time, successful deployment requires careful attention to data management, calibration maintenance and integration with existing plant procedures. Spectral models must be periodically checked against reference laboratory measurements, and mechanisms should be in place to detect drift or out-of-distribution samples.[23, 29, 32] Moreover, human operators should be trained to understand both the capabilities and limitations of the system, viewing model outputs as decision support rather than unquestionable ground truth.

From a research perspective, this work can be extended by comparing a wider range of algorithms (e.g. convolutional neural networks on raw spectra, attention-based models, or advanced ensemble methods) and by exploring domain adaptation techniques to handle cross-instrument or cross-site variability.[23, 29, 32, 27] Multimodal fusion of FT-NIR with other signals such as XRF, LIBS or process variables also represents a promising direction, as evidenced by recent studies in coal quality prediction.[20, 21, 24, 30]

6. Future Scope

Several enhancements can be envisioned for the current system:

- **Quality parameter prediction:** Extending the model from discrete coal type classification to regression tasks that predict moisture, ash, volatile matter, fixed carbon, sulfur and calorific value, using multi-output learning approaches.[10, 22, 24]
- **Process optimization suggestions:** Leveraging predicted quality parameters and coal types to recommend actions such as drying high-moisture coal, adjusting blending ratios or routing specific lots to suitable boilers or processes.
- **Online integration:** Coupling the system with online FT-NIR sensors mounted over conveyor belts or feeders, enabling continuous coal stream monitoring and automated alarms when significant deviations are detected. [22,27, 30]
- **Adaptive and transferable models:** Implementing transfer learning or domain adaptation techniques to port models across instruments, plants or coalfields with minimal additional calibration data. [23, 29, 32]

7. Conclusion

This paper has presented a methodology and system design for non-destructive coal type classification using FT-NIR spectroscopy and machine learning. By combining spectral preprocessing, PCA-based dimensionality reduction and supervised classifiers with a deployable software pipeline, the proposed approach offers a viable path towards integrating rapid coal classification into industrial workflows.

While quantitative performance metrics depend on specific datasets and are left for future experimental work, existing literature and the structure of the presented system indicate that such approaches can achieve high accuracy and substantial operational benefits when properly calibrated and maintained.[10, 20, 22, 23, 27] As industries continue to seek more efficient and environmentally responsible use of coal, non-destructive, data-driven methods like the one described here can play an important role in enabling smarter fuel management and process optimization

8. References

1. J. N. Lester, et al., "Analysis of coal by diffuse reflectance near-infrared spectroscopy," Analytica Chimica Acta, 2005.
2. Y. Zhang, et al., "Enhancing multi-type coal quality prediction accuracy with fusion spectra and classification models using NIRS and XRF techniques," Journal of Analytical Atomic Spectrometry, 2024.
3. "Using AI with NIRS-XRF Fusion to Improve Coal Quality Analysis," Spectroscopy Online, 2025.
4. Metrohm AG, "Analysis of moisture, ash, carbon, and volatile content in coal by NIRS," Application Note AN-NIR-117, 2023.
5. X. Le, et al., "Coal classification method based on visible-infrared spectroscopy and improved multilayer extreme learning machine," *Measurement*, 2020.
6. Q. Wang, et al., "Advanced Multi-Parameter Prediction of Coal Quality Using LIBS and Ensemble Machine Learning Techniques," *ACS Omega*, 2025.
7. Kentucky Geological Survey, "Volatile Matter (Part of Proximate Analysis)," University of Kentucky, online article, accessed 2026.
8. M. Li, et al., "Near-Infrared Spectrum of Coal Origin Identification Based on LVQ with SVM Algorithm," *Guang Pu Xue Yu Guang Pu Fen Xi*, 2016.
9. X. Chen, et al., "Non-destructive spectroscopy assisted by machine learning for coal industrial analysis," preprint, 2024.
10. Metrohm AG, "Analysis of moisture, ash, carbon, and volatile content in coal by NIRS (full PDF)," 2023.
11. H. Liu, et al., "A near-infrared spectroscopy dataset of coal and coal-measure rock," *Scientific Data*, 2024.[1]
12. Y. Sun, et al., "Non-Destructive Spectroscopy Assisted by Machine Learning for Coal Industrial Analysis," SSRN preprint, 2024.
13. "Proximate Analysis of Coal | Coal Analysis Procedure," Coal Handling Plants, technical article, 2024.
14. Z. Zhao, et al., "Accurate Coal Classification Using PAIPSO-ELM with Near-Infrared Reflectance Spectra," *Applied Spectroscopy*, 2024.