UNIVERSITY OF ZIELONA GÓRA

CIVIL AND ENVIRONMENTAL ENGINEERING REPORTS

E-ISSN 2450-8594

CEER 2025; 35 (3): 0255-0270
DOI: 10.59440/ceer/205778
Original Research Article

AN END-TO-END DEEP LEARNING MODEL FOR PREDICTING TOTAL INCOMBUSTIBLE CONTENTS IN COAL/STONE DUST

Chenhao WANG[1], Shevaune ZENG, Eleonora WIDZYK-CAPEHART, Gareth KENNEDY
Simtars, Resources Safety & Health Queensland, Queensland, Australia

Abstract

Coal dust explosions are a significant threat to underground coal mines. To reduce this risk, the combustible coal dust is mixed with stone dust to increase the total incombustible contents (TIC). Conventional TIC measurement methods rely on time-consuming laboratory analyses, involving extensive sample preparation and chemical testing. In contrast, near-infrared (NIR) spectroscopy has emerged as a rapid, non-destructive alternative for TIC prediction. However, existing machine learning models for analysing high-dimensional spectral data often require extensive preprocessing, increasing the analysis complexity.

In this study, we present a residual - Convolutional Neural Network (CNN) based method for end-to-end analysis of raw near-infrared (NIR) spectral data to reduce preprocessing requirements while accurately classifying the TIC levels. The model was evaluated using 300 coal/stone dust samples, with 100 coal samples sourced from various Australian coal mines. The deep learning model, configured with optimal nine residual blocks, demonstrated high accuracy in predicting high TIC samples (TIC ≥ 85%), achieving misclassification rates of 0.05 on the training set and 0.14 on the testing set, respectively.

Two challenges were identified: class imbalance and spectral overlap. The low TIC samples (TIC < 70%) accounted for only 9% of the total dataset (27 out of 300), resulting in poor prediction for this underrepresented class. Additionally, significant spectral similarity with distinct TIC values reduced the model's generalization ability. Despite these challenges, our study demonstrates the potential for developing a reliable and efficient end-to-end deep learning framework for TIC prediction, which would allow for a significant reduction in preprocessing efforts.

Keywords: coal dust explosions, total incombustible contents, spectroscopy, deep learning

1. INTRODUCTION

Coal dust introduces significant safety and health hazards in underground mining. Beyond its long-term health effects on miners such as lung diseases, explosions can be triggered by uncontrolled coal dust [1]. Over the past century, coal dust explosions have caused some of the worst mining disasters,

[1] Corresponding author: Chenhao WANG, Simtars, Resources Safety & Health Queensland, 2 Robert Smith Street, Redbank, 4301, Queensland, Australia, Chenhao.wang@simtars.qld.gov.au, +61 409 355 866

including the Mount Mulligan Mine disaster in Australia in 1921, which killed 75 miners, and the Pike River Mine disaster in New Zealand in 2010, resulting in 29 fatalities. These disasters highlight the importance of monitoring and controlling coal dust in underground mines to ensure safety. A key preventative measure is increasing the total incombustible content (TIC) of the coal dust in the mine by adding stone dust to the already present coal dust, which reduces its combustibility. In Australia, regulations set the TIC thresholds to 70%, 80% and 85% depending on the zone within the mine [2, 3]. Accurate and timely TIC measurements are therefore critical to adhere to the regulations and, thus, to reduce explosion risks.

Traditional TIC measurement methods, such as Low Temperature Ashing (LTA), rely on chemical testing and extensive manual sample preparation, making them unsuitable for real-time monitoring in dynamic mining environments. While the Coal Dust Explosibility Meter (CDEM), developed by the National Institute for Occupational Safety and Health (NIOSH), was introduced to address the time gap, it suffered from significant limitations, including complex calibration, moisture sensitivity, and reduced accuracy for certain coal types, especially in Australia [4]. As a result, labour-intensive laboratory analysis remains the primary method for TIC measurement, limiting operational efficiency.

Near-Infrared Spectroscopy (NIRS) has emerged as a fast, non-destructive method for material analysis, with proven success across industries, such as agriculture, food processing, and mining. NIRS has enabled reliable prediction of material properties when combined with machine learning models. For example, Begum et al. [5] classified coal species using short-wave infrared spectroscopy, while Zou et al. [6] employed NIRS with enhanced broad learning to identify coal grades. Andres et al. [7] demonstrated the accuracy of NIRS with Partial Least Squares Regression (PLS) for coal property prediction, and Li et al. [8] validated its feasibility for sulphur quantification in Australian coal. Machine learning models, including Support Vector Machines (SVMs) and Extreme Gradient Boosting (XGBoost), have also been successfully applied to NIRS data, capturing relationships between spectra and coal quality metrics. However, applying NIRS to TIC prediction in coal/stone dust samples still faces several challenges.

Traditional machine learning methods require extensive data preprocessing. This includes baseline correction, smoothing, normalization, outlier removal, data augmentation and other, which must be recalibrated when analysing samples from different source. This dependence on predefined preprocessing steps reduces the generalizability of these models. Moreover, such methods are less effective in capturing complex, non-linear patterns in high-dimensional spectral data, resulting in reduced predictive accuracy, especially when faced with noisy or imbalanced datasets. Advances in deep learning, particularly Convolutional Neural Networks (CNNs), have provided significant potential for overcoming these challenges. Unlike traditional methods, CNNs can process raw spectral data directly, eliminating the need for extensive preprocessing. They can automatically identify patterns and relationships within spectral data, making them highly adaptable to diverse coal/stone dust samples. For instance, Chakravartula et al. [9] utilized CNNs for detecting coffee adulteration using Fourier-Transform NIR spectroscopy, and Le et al. [10] employed CNNs for predicting coal quality.

In this paper, we propose an end-to-end deep learning framework that integrates NIRS with a CNN architecture incorporating residual blocks for TIC prediction in coal/stone dust samples. This framework eliminates the need for extensive preprocessing and demonstrates adaptability to small, imbalanced coal/stone dust datasets. The structure of this paper is as follows: Section 2 outlines the methodology, including sample preparation, NIR measurement, baseline correction, and model architecture; Section 3 presents results and discussion, and Section 4 concludes with key findings and future directions.

2. METHODOLOGY

The overall framework of the TIC prediction system is illustrated in Fig. 1. Specifically, this workflow begins with sample preparation and data collection. Coal/stone dust samples were prepared with predefined TIC levels, ensuring they represented diverse conditions and comply with Australian regulatory thresholds. These samples were analysed using a portable NIR spectrometer to capture spectral data across the wavelength range of 900 – 1700 nm. Once the spectral dataset was collected, it was immediately divided into training (80%) and testing (20%) datasets. The training dataset, which reflects the TIC categories' imbalance, was adopted for model development. Rather than employing artificial balancing techniques, such as, data augmentation, the adopted approach in this study depended on the robustness of the deep learning model's architecture to effectively handle the imbalance by learning from the raw spectral features. The test dataset was strictly reserved for final evaluation to prevent data leakage.

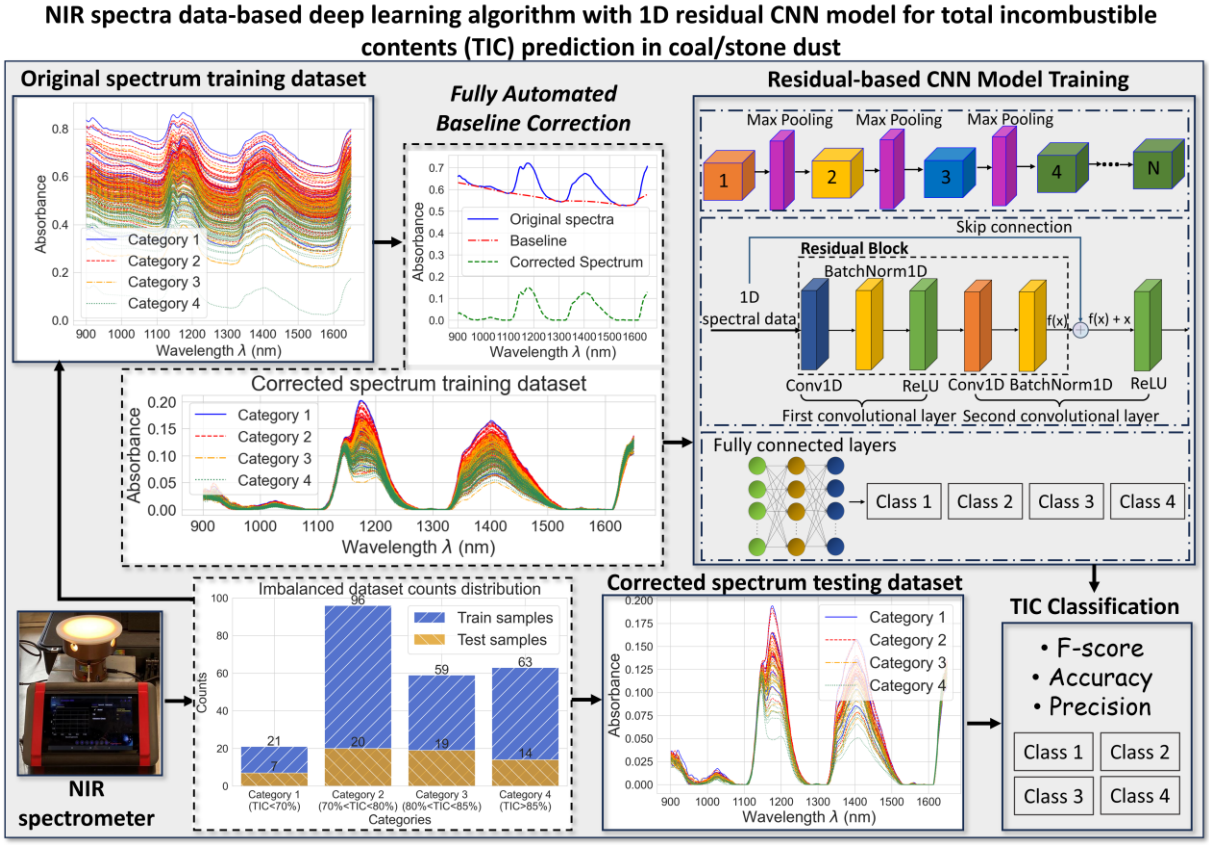

Fig. 1. Overview of the NIR spectral data-based deep learning model for TIC classification in coal/stone dust

The second step was an essential preprocessing. The raw spectral data typically contains noise, baseline drift and other variabilities caused by environmental conditions or the instrument itself. To ensure data quality, an automated baseline correction algorithm that dynamically adjusts the spectral baseline to enhance signal quality was applied. The fully automated baseline correction was applied consistently across training and testing datasets.

The core of the system was the CNN with residual blocks, specifically designed to process high-dimensional spectral data directly. The CNN architecture includes multiple modules, such as, fully

connected layer, max pooling layer, adaptive average pooling layer and residual blocks, for extracting both local and global features from the input spectra. These learned spectral features were adopted in another fully connected layer to classify the samples into four TIC categories: class 1 (TIC < 70%), class 2 (70% ≤ TIC < 80%), class 3 (80% ≤TIC < 85%), and class 4 (TIC ≥ 85%).

The final stage of the presented work involved model evaluation and classification. After training the CNN model on the training dataset, its performance was validated using the test dataset. The evaluation focused on key metrics, including F-score, accuracy, precision and misclassification rate, to provide a comprehensive evaluation of the model's predictive capabilities.

2.1. Samples preparation and spectral measurements

A total of 100 coal samples were collected from mines in Queensland and New South Wales, Australia, to ensure diversity and representativeness in the dataset for TIC prediction. These samples included both fresh and aged coal stored for up to two decades, showing a range of ash and moisture levels. Proximate analyses were performed on all samples following ISO 11722, ISO 1171, and ISO 562 Standards to characterize their properties prior to further preparation. Statistical tests, including the Levene Test and One-Way ANOVA, as listed in Table 1, indicated that, while most parameters i.e. ash, volatile matter and fixed carbon, showed no significant difference ($p > 0.05$), moisture content exhibited a significant regional variation ($p < 0.05$), which was likely due to difference in aging and drying during storage.

Table 1. Statistical Analysis of Proximate Analysis by Region

Coal Quality	Levene Test	One-Way ANOVA
Ash	0.21, p=0.65	0.19, p=0.67
Moisture	7.20, p=0.009	4.60, p=0.04
Volatile Matter	0.01, p=0.93	1.41, p=0.24
Fixed Carbon	2.32, p=0.13	0.16, p=0.69

To prepare the coal/stone dust samples for analysis, the coal was first crushed to form particles smaller than 4 mm using a Jaqus laboratory jaw crusher and then milled to less than 250 μm with a Retsch Cross Beater mill. The processed coal dust was mixed with stone dust to produce a total of 300 coal/stone dust samples. Each of the 100 coal samples was used to prepare three mixed samples with TIC levels targeted at 70%, 80%, and 85%, respectively.

The TIC includes both stone dust and the ash content of coal. For example, 70% TIC means that 70% of the total weight of the coal/stone dust sample is composed of incombustible components (stone dust + coal ash), while the remaining 30% consists of combustible coal contents, such as, Fixed Carbon and Volatile Matter. The stone dust, primarily composed of calcium carbonate (CaCO3) and supplied by Sibelco and SEQ lime, was milled to a similar particle size as the coal dust. Its minimal absorption of the NIR wavelengths ensured that it did not significantly affect the spectral data.

The TIC levels of the mixed coal/stone dust samples were designed to follow three normal distributions centred at 70%, 80% and 85%, respectively, each generating 100 random TIC values with a standard deviation of 5%. These random TIC values ensured realistic variability reflective of regulatory thresholds, providing a diverse and representative dataset for model training.

Each coal/stone dust sample was prepared using a fixed 15g of coal dust, and the required stone dust weight was calculated using Eq. 2.1:

$$w_{coal\_dust} = \frac{1 - TIC * 100\%}{1 - Ash\_Content * 100\%} * w_{total} \tag{2.1}$$

where: w_{coal_dust} is the coal dust weight (15g in this case), w_{total} is the total weight of coal/stone dust sample, $Ash\_Content$ is the percentage of the ash in the original coal. The required stone dust weight was then calculated as per Eq. 2.2:

$$w_{stone\_dust} = w_{total} - w_{coal\_dust} \tag{2.2}$$

where: w_{stone_dust} is the weight of the stone dust needed to achieve the target TIC value. This systematic approach ensured accurate and consistent mixing of the coal and stone dust for each target TIC level.

The prepared samples were placed in clear glass petri dishes and spread into a thin, uniform layer approximately 2–3 mm thick to ensure consistent spectral measurements. Measurements were conducted on a neutral-coloured laboratory countertop to avoid interference from the surface or container. NIR spectra were collected using a StellarNet NIR ADK portable spectrometer with a resolution of <5 nm and accuracy of <0.25 nm (Fig. 1), capturing data in the wavelength spectrum between 900 – 1700 nm. The instrument recorded absorbance units and counts, and all spectral data were tagged with unique sample identifiers. The global dataset was then divided into training (80%) and testing (20%) subsets.

As a total of 300 coal/stone dust samples were prepared, 240 samples were used for training and the remaining 60 samples were reserved for testing. It should be noted that the dataset size is relatively limited for training deep neural networks, which typically require large-scale data to achieve robust generalization. However, the aim of this study was to conduct an initial investigation into the feasibility of applying a deep learning model to predict TIC directly from raw or minimally pre-processed NIR spectral data. This exploratory approach intentionally avoided extensive pre-processing and data augmentation to evaluate the baseline capability of CNN architectures under realistic constraints that commonly occur in industrial environments. To reduce the risk of overfitting on small datasets, a lightweight one-dimensional CNN (1D CNN) was adopted in this study. This architecture is well-suited for sequential spectral data and contains significantly fewer parameters compared to standard image-based 2D CNNs.

In parallel research, a comprehensive analysis was done using classical machine learning algorithms, including Partial Least Squares Regression (PLS), Support Vector Machines (SVM) and Random Forests (RF), in combination with various advanced spectral data pre-processing and synthetic data generation techniques, such as Synthetic Minority Over-sampling Technique (SMOTE) and Gaussian Process Regression (GPR). These combinations of models and synthetic data generation techniques demonstrated improved accuracy and robustness on the same imbalanced and small sample dataset. In addition, similar synthetic data generation techniques will be incorporated into the deep learning framework in future work, with the aim of developing a more scalable, accurate, and generalizable TIC prediction system suitable for practical field deployment.

2.2. Fully automated baseline correction algorithm

The baseline correction process is an essential step in preparing NIR spectral data for analysis as it ensures that noise and baseline shifts, introduced by environmental or instrumental variability, do not significantly impact the model's ability to extract meaningful features. Numerous studies have proposed

various approaches to baseline correction for spectral data [11-14]. After a comprehensive evaluation of these methods, a small-window moving average-based fully automated baseline estimation method was implemented [12]. This approach is model-free, adaptive to a wide range of spectral baseline for integration into the deep learning modelling framework for fully automated correction. The flowchart of this algorithm is illustrated in Fig. 2.

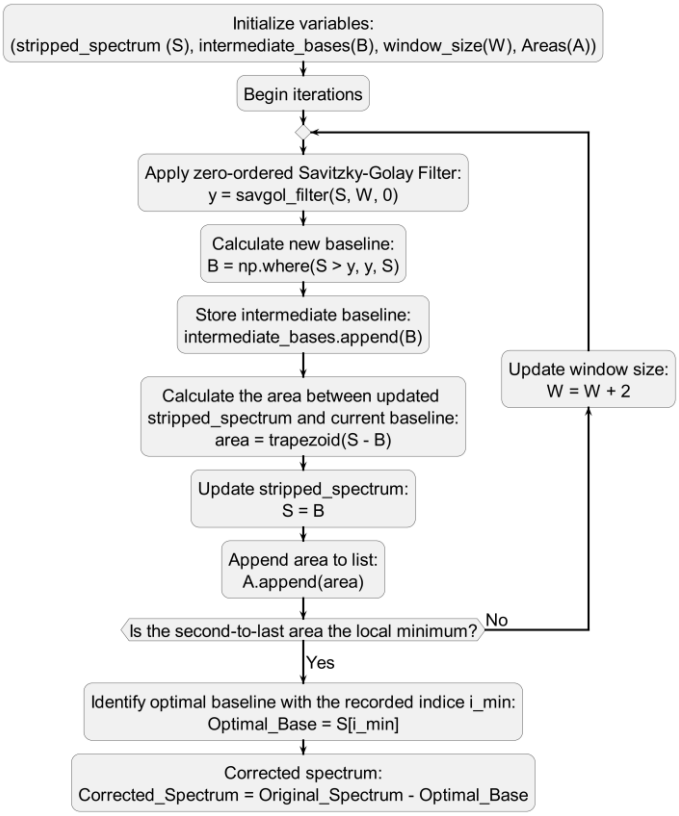

Fig. 2. Flowchart of the baseline correction algorithm

The correction process begins by initializing the stripped spectrum S, intermediate baseline shapes B, the window size W, and area A. A zero-ordered Savitzky-Golay filter is applied iteratively to calculate a smoothed baseline for the raw spectral data. This filter preserves the spectral signal's shape and key features while smoothing out noise. During each iteration, the baseline B is updated relative to the stripped spectrum S, and the intermediate baselines are stored for further analysis. The area between the stripped spectrum and the baseline is recalculated using the trapezoidal integration method, and the window size W is incrementally increased to refine the baseline calculation. A critical step in this algorithm is identifying the optimal baseline. This is achieved by determining the local minimum of the second-to-last area value in the list of areas calculated during the iterations. The baseline corresponding to this local minimum is selected as the optimal baseline, and the corrected spectrum is computed by subtracting the optimal baseline from the original spectrum. This iterative and adaptive process allows the algorithm to handle a wide range of baseline variations without requiring manual adjustments.

The effectiveness of the baseline correction algorithm is illustrated in Fig. 3(a) and (b). Fig. 3(a) shows the original spectral samples from the training dataset, highlighting significant baseline shifts across different classes, which could introduce noise and bias into downstream processing. After

applying the correction algorithm, the corrected spectra in Fig. 3(b) are well-aligned, with baseline variations effectively removed while preserving the spectral features necessary for TIC classification. This automated and iterative correction method eliminates the need for manual calibration or external adjustments, ensuring a standardized preprocessing pipeline. The corrected spectral data served as new inputs for the TIC classification model.

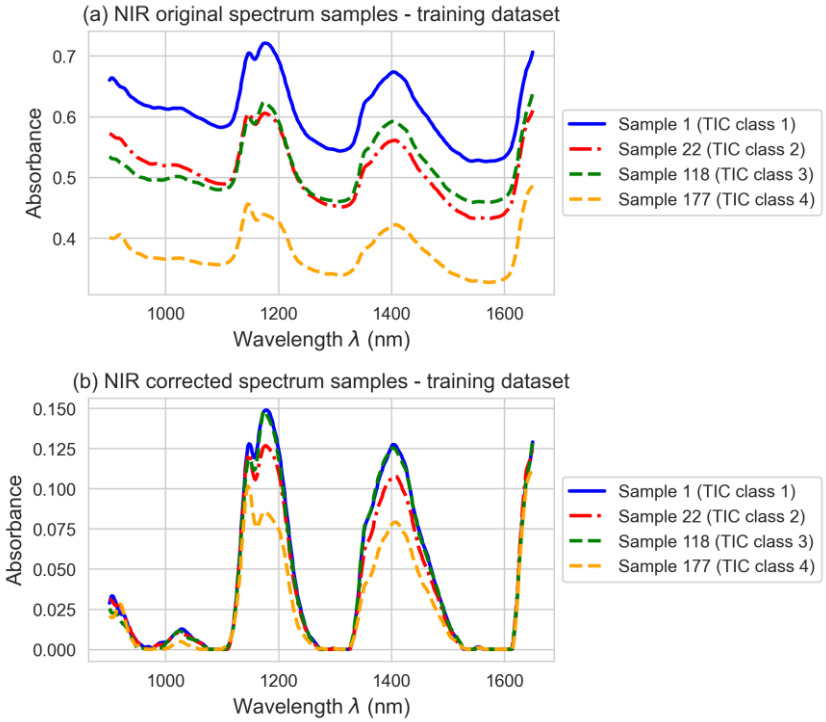

Fig. 3. The original and corrected spectrum samples in the training set: (a) original samples in different classes, (b) corrected samples in different classes

2.3. Model architecture

This section details the model architecture, as illustrated in Fig. 4. The architecture is based on 1D-CNN and incorporates residual blocks to efficiently process the high-dimensional spectral data. This design allows the model to capture both low-level and high-level features from the NIR spectra, which are essential for accurate TIC classification. At the input stage, the corrected spectral data are fed into the model, starting with a series of convolutional layers organized into residual blocks. These residual blocks utilize skip connections to preserve information from earlier layers and address the vanishing gradient problem during training.

The earlier residual blocks are designed with fewer channels to focus on low-level local features, such as, the peaks and valleys of individual spectra. These local features capture subtle variations in the data that are indicative of TIC differences. As the data moves deeper into the network, subsequent residual blocks with an increased number of channels extract high-level global features, such as, the overall spectral shape, which are critical for differentiating between TIC categories. To further enhance computational efficiency and avoid overfitting, max pooling layers are applied between groups of residual blocks to reduce data dimensionality while retaining the most important spectral features. An adaptive average pooling layer then transforms the output into a fixed-length feature vector, ensuring

compatibility with the final fully connected layers. Finally, the fully connected layers map the extracted features to the output layer, where the data is classified into four TIC categories: class 1 (TIC < 70%), class 2 (70% ≤ TIC < 80%), class 3 (80% ≤TIC < 85%), and class 4 (TIC ≥ 85%).

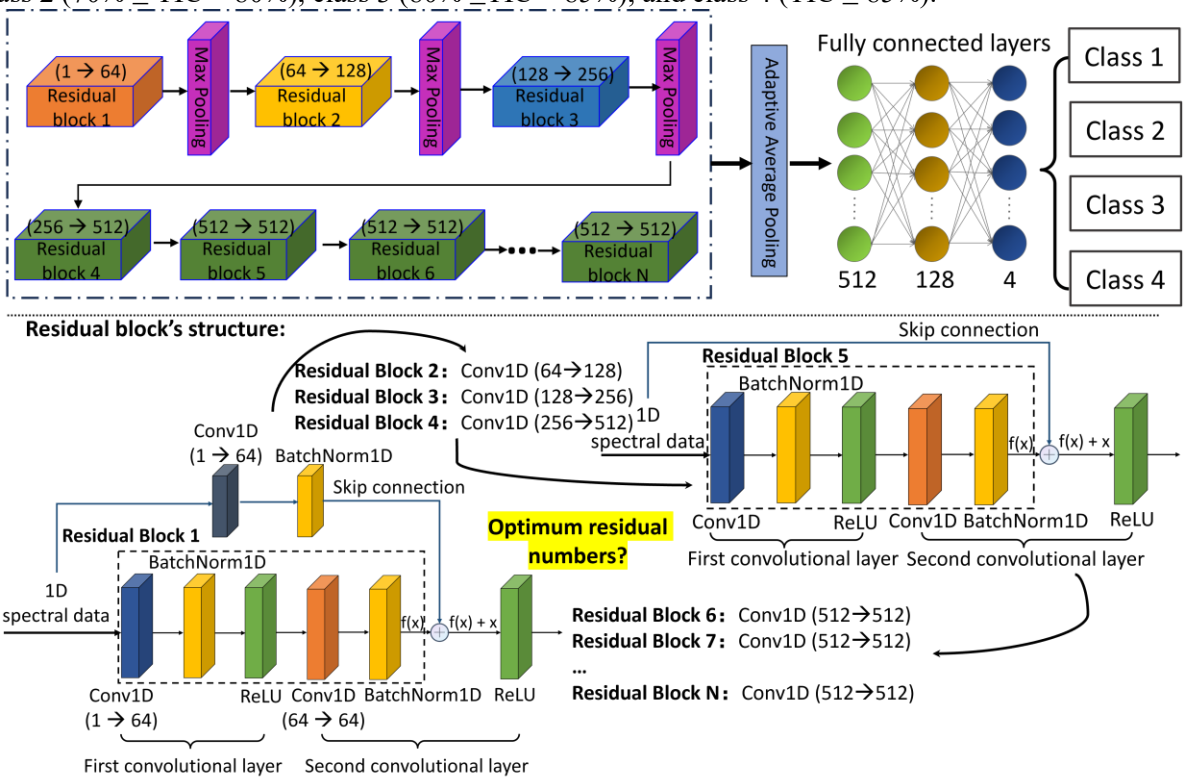

Fig. 4. Model architecture of the 1D CNN with residual blocks

A key aspect of this model is the role of residual blocks in improving feature extraction and learning efficiency. The skip connections within these blocks allow the network to learn both local and global patterns effectively, even as the depth of the network increases. However, the number of residual blocks significantly influences the model's performance. While deeper networks can extract more complex features, excessive depth may lead to overfitting in accuracy. The effect of the number of blocks on model performance was investigated by varying the number of residual blocks from 0 to 15. This aimed to identify the optimal number of residual blocks within the current model architecture that yield the highest classification accuracy.

2.4. Modelling environment and parameters

All modelling was conducted using Python 3.11.7 and PyTorch 2.2.0 and processed on a 13th Gen Intel® Core™ i7-13850HX CPU with 64 GB RAM, running Windows 10. All training and inference were performed on the CPU without GPU acceleration. The corrected spectral data were used as single channel 1D input to the model.

The residual block structure configuration followed a progressive channel expansion strategy: from 1 to 64 channels in the first block, then doubling in subsequent blocks up to 512 channels. All subsequent residual blocks maintained 512 input and output channels. Max pooling layers with a kernel size and stride of 2 were applied between major stages to reduce dimensionality. The final fully

connected layer reduced the extracted features from 512 to 128 and then to 4 output units, representing the four TIC classes. Model training used the Cross Entropy Loss Function and the Adam optimizer (learning rate 0.001), with a batch size of 32 and 50 training epochs. These parameters were kept consistent for all tests evaluating different numbers of residual blocks.

3. RESULTS AND DISCUSSION

This section presents the results and analysis of the proposed CNN model with residual blocks for TIC based classification of coal/stone dust. Section 3.1 presents the influence of the number of residual blocks on model performance, identifying the optimal configuration. Section 3.2 demonstrates the training results with the optimal residual block configuration, highlighting the model's learning capability through misclassification rates and performance metrics. Section 3.3 evaluates the model's generalization ability on the testing dataset, providing insights into its reliability for unseen data. Finally, Section 3.4 explores the spectral data similarity, discussing the challenges of overlapping spectral patterns across TIC classes.

3.1 Influence of residual blocks

The results of the investigation into the influence of the number of residual blocks on model performance are summarized in Fig. 5. These results show the training and testing classification accuracies for models using selected numbers of residual blocks (0, 3, 6, 9, 12 and 15) to evaluate performance trends across increasing model depth.

As demonstrated, the classification accuracy generally improves as the number of residual blocks increases, reaching a peak at 9 residual blocks. At this point, the training accuracy is 0.58, and the testing accuracy is 0.6, indicating a good balance between the model's ability to learn from the training data and its generalization to unseen data. The improved performance with up to 9 residual blocks can be attributed to the increased capacity of the model to extract meaningful features from the high-dimensional spectral data. The residual connections in these blocks allow the model to effectively learn both low-level and high-level features, enhancing its overall predictive ability.

However, beyond 9 residual blocks, the model's performance begins to decrease, particularly on the test dataset. For instance, when the number of residual blocks reaches 12, the testing accuracy drops significantly from 0.6 to 0.38 by over 36.7%. This decrease suggests that excessive network depth introduces overfitting, where the model becomes overly specialized to the training data and loses its ability to generalize. The training accuracy remains relatively stable after 9 blocks, further indicating that the additional blocks contribute little to meaningful learning and may instead introduce noise or redundant features.

The performance at 0 residual blocks (a simple CNN without residual connections) is notably poor, with a testing accuracy of 0.35. This shows the importance of residual connections in enabling the network to transfer information effectively across layers and eliminate vanishing gradient issues. The gradual improvement from 0 to 9 residual blocks demonstrates the value of residual learning in improving feature extraction and classification accuracy for spectral data. While adding more blocks increases computational cost and the risk of overfitting, too few blocks limit the model's capacity to capture essential patterns in the current datasets (Fig. 4). These results indicate that 9 residual blocks represent the optimal configuration for this dataset, balancing model complexity and performance.

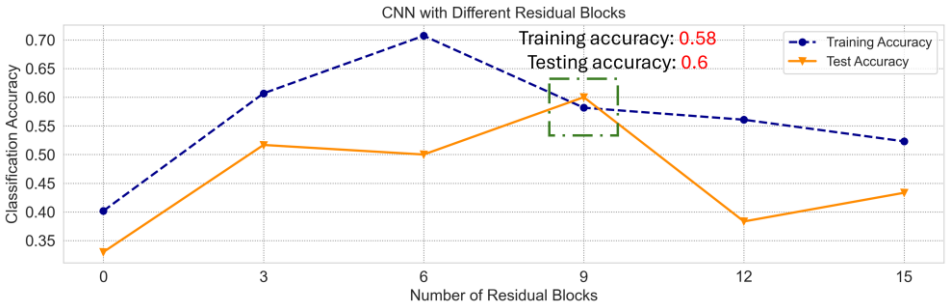

Fig. 5. Training and testing classification accuracy using CNN model with different residual blocks

3.2 Training results with optimal residual blocks

The training results with the optimal number of residual blocks (9 blocks) are presented using multiple metrics, as shown in Fig. 6 and Fig. 7. These metrics include the misclassification rate, confusion matrix, and performance scores (precision, recall, and F1-score) across the four TIC classes.

The misclassification rates for each TIC class, as shown in Fig. 6, exhibit significant differences in the model's performance across the classes. Class 1 (TIC < 70%) exhibits the highest misclassification rate at 1.0, indicating that all class 1 samples are incorrectly classified. This poor performance is further supported by the confusion matrix, where none of the true class 1 samples are correctly predicted. This may be attributed to the limited representation of class 1 in the training dataset or spectral similarities between class 1 and neighbouring classes.

Classes 2 (70% ≤ TIC < 80%) and 3 (80% ≤ TIC < 85%) also show relatively high misclassification rates of 0.47 and 0.53, respectively. The confusion matrix indicates that a significant number of samples from these classes are misclassified as each other, likely due to the overlapping spectral features in their TIC ranges. In contrast, class 4 (TIC ≥ 85%) demonstrates a very low misclassification rate of 0.05, with most samples correctly classified. The confusion matrix in Fig. 6 confirms this result, showing 60 correctly predicted samples for class 4 out of 63 total samples.

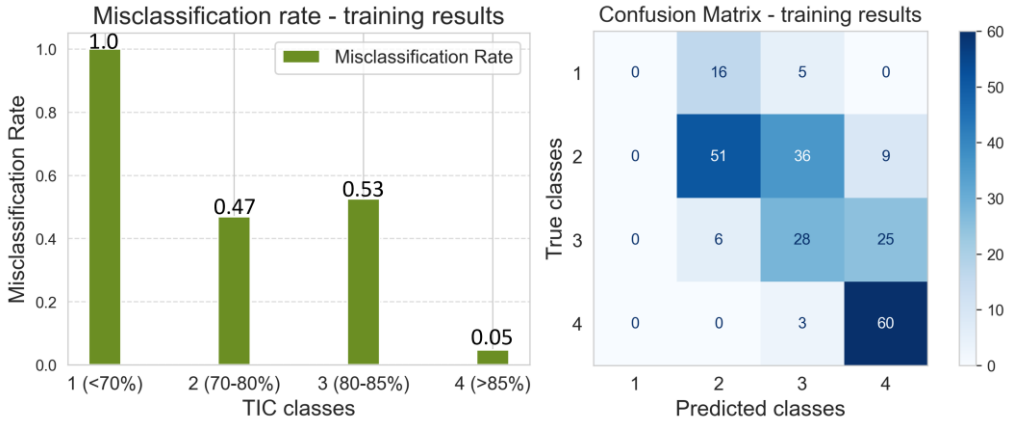

Fig. 6. Misclassification rate and confusion matrix using the model with optimized residual blocks on the training set

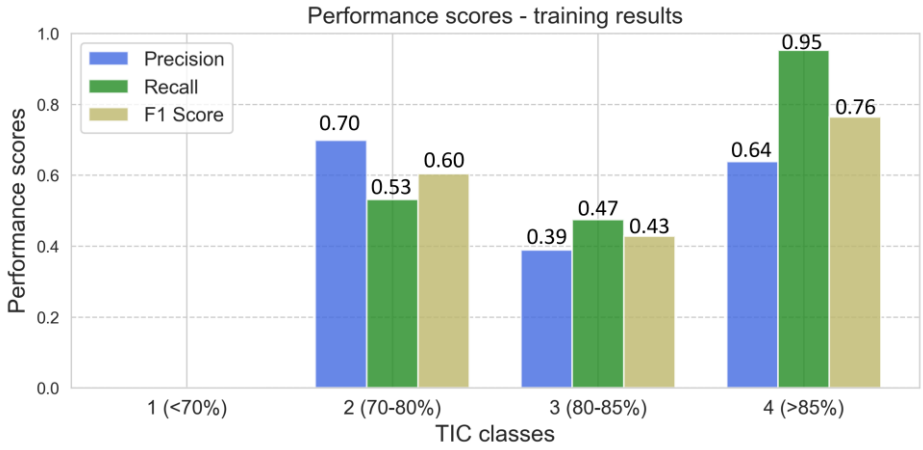

Fig. 7. Performance scores of the model with optimized residual blocks on the training dataset

The model's further performance evaluation using precision, recall, and F1-score for each class is shown in Fig. 7. Precision measures the proportion of correctly predicted samples among all samples predicted for a given class, recall measures the proportion of correctly predicted samples among the total true samples for a class, and the F1-score provides a harmonic mean of precision and recall providing a balanced evaluation [15].

Specifically, class 1 (TIC < 70%) has the lowest scores across all metrics, with precision, recall, and F1-score all being zero. This indicates that the model struggles to identify samples from this class accurately. Class 2 (70% ≤ TIC < 80%) achieves moderate precision (0.70) but lower recall (0.53) and F1-score (0.60). This shows that while the model is relatively precise in predicting class 2, it fails to capture a significant portion of the true samples, leading to moderate performance. Class 3 (80% ≤ TIC < 85%) has similar performance to class 2, with slightly lower precision (0.39), recall (0.47), and F1-score (0.43). This suggests that spectral overlaps with other classes impact the model's ability to distinguish this range effectively. In comparison, class 4 (TIC ≥ 85%) exhibits the highest performance in terms of precision (0.95), recall (0.64), and F1-score (0.76). These results indicate that the deep learning model excels in identifying high TIC samples, which likely have distinct spectral features, like stronger and more consistent absorption peaks in specific wavelength ranges, that the model can recognize.

These results highlight the varying performance of the model across TIC classes, revealing specific strengths and limitations. The high performance for class 4 (TIC ≥ 85%) suggests that the model effectively captures clear spectral patterns in high TIC samples. However, the poor performance for class 1 (TIC < 70%) points to challenges, such as, minority of Class 1 in the dataset or overlapping spectral features with Classes 2 and 3, which make it difficult for the model to differentiate this range accurately. Similarly, the moderate performance for Classes 2 and 3 reflects the complexity of distinguishing between intermediate TIC ranges.

3.3 Testing results with optimal residual blocks

The testing results for the model with the optimal configuration of 9 residual blocks are presented in Fig. 8 and Fig. 9. These results evaluate the model's generalization performance for unseen data using accuracy metrics, such as the misclassification rate, confusion matrix, and performance scores (precision, recall, and F1-score) for the four TIC classes.

The misclassification rates across TIC classes, as shown in Fig. 8, demonstrate notable differences in the model's ability to correctly classify samples in each class. Similarly to the training results, class 1 (TIC < 70%) has the highest misclassification rate at 1.0, indicating that none of the class 1 samples were correctly predicted. The confusion matrix further supports this observation, as all seven samples from class 1 in the testing dataset were misclassified, mostly as class 2. This issue may be attributed to the minority of class 1 in the training data, as well as the spectral similarities between classes.

In contrast, class 2 (70% ≤ TIC < 80%) and class 3 (80% ≤ TIC < 85%) show moderate misclassification rates of 0.35 and 0.42, respectively. The confusion matrix indicates that some class 2 samples are misclassified as class 3, and vice versa, reflecting the challenge of distinguishing between intermediate TIC ranges. However, the misclassification rates for these classes are significantly lower than for class 1, suggesting that the model is better at identifying intermediate classes, although with some confusion.

For class 4 (TIC ≥ 85%) the misclassification rate is the lowest at 0.14, with most samples correctly classified. The confusion matrix shows that 12 out of 14 samples from class 4 were correctly predicted, highlighting the model's strong performance for this class. The distinct spectral features of high TIC samples likely contributed to this higher accuracy.

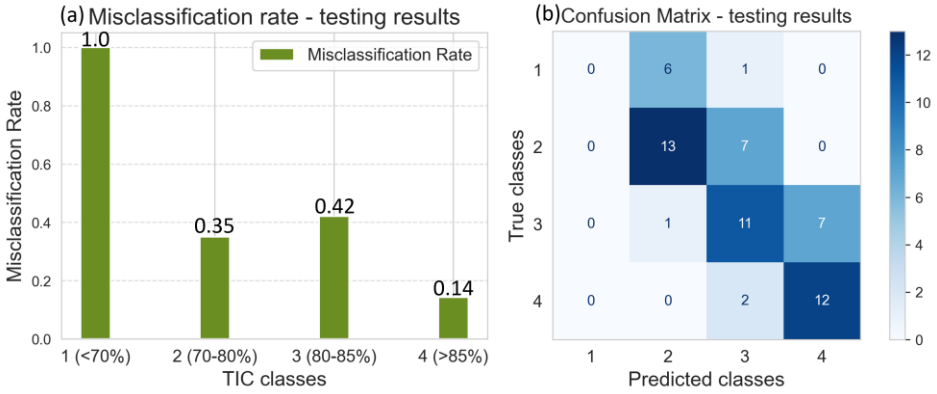

Fig. 8. Misclassification rate and confusion matrix on the testing dataset

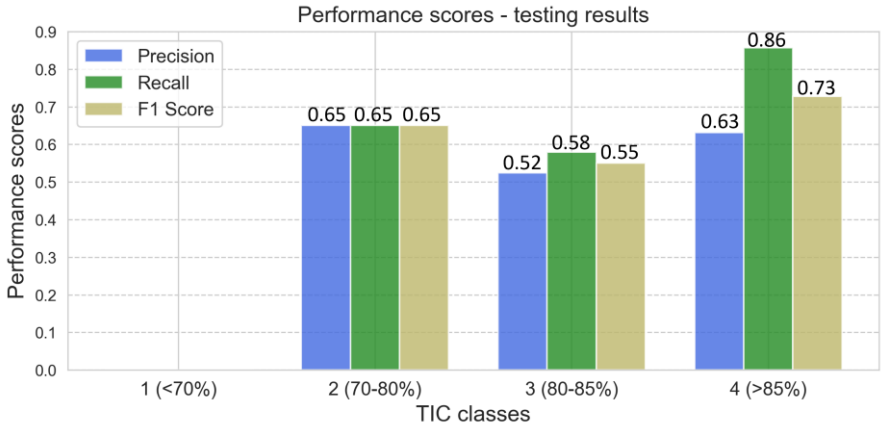

Fig. 9. Performance scores of the model on the testing dataset

These performance scores provide a detailed evaluation of the model's generalization performance for each TIC class, as shown in Fig. 9. For class 1 (TIC < 70%), all performance metrics

are zero, reflecting the model's complete failure to classify this class correctly. This indicates that class 1 samples either share significant overlap with neighbouring classes or are too underrepresented in the training data for the model to learn their features effectively. Class 2 (70% ≤ TIC < 80%) achieved balanced performance with precision, recall, and F1-scores all at 0.65. This suggests that while the model can identify a majority of the class 2 samples, some are still misclassified as class 3 due to overlapping spectral characteristics. Class 3 (80% ≤ TIC < 85%) demonstrates slightly lower performance, with precision at 0.52, recall at 0.58, and an F1-score of 0.55. This indicates that class 3 samples are more frequently confused with neighbouring classes, which impacts the model's overall performance for this range. For class 4 (TIC ≥ 85%), the model achieves its highest performance, with precision at 0.86, recall at 0.63, and an F1-score of 0.73. These results indicate that the model effectively identifies high TIC samples, likely due to their distinct and separable spectral features.

The testing results highlight significant variability in the model's ability to classify TIC classes, with strong performance for class 4 and moderate success for classes 2 and 3. However, the failure to classify class 1 accurately shows the limitations of the current dataset and model configuration. The minority of class 1 in the training data, combined with the spectral similarities between low TIC classes, presents a major challenge for the model. These findings suggest that improving data balance may enhance the model's performance for minority classes. Additionally, the moderate performance for intermediate TIC classes (2 and 3) indicates that further optimization of the model architecture may be necessary to better capture minor differences in spectral features.

3.4 Spectral data similarity

The selected spectral curves shown in Fig. 10 represent samples from different TIC classes. These curves reveal a significant challenge in the classification task, i.e. the high degree of spectral similarity between samples with different TIC values. These plots show the absorbance spectra for several samples, each with a unique TIC value. Despite belonging to different TIC classes, the spectral curves show significantly similar patterns, particularly in their overall shape, peaks, and valleys.

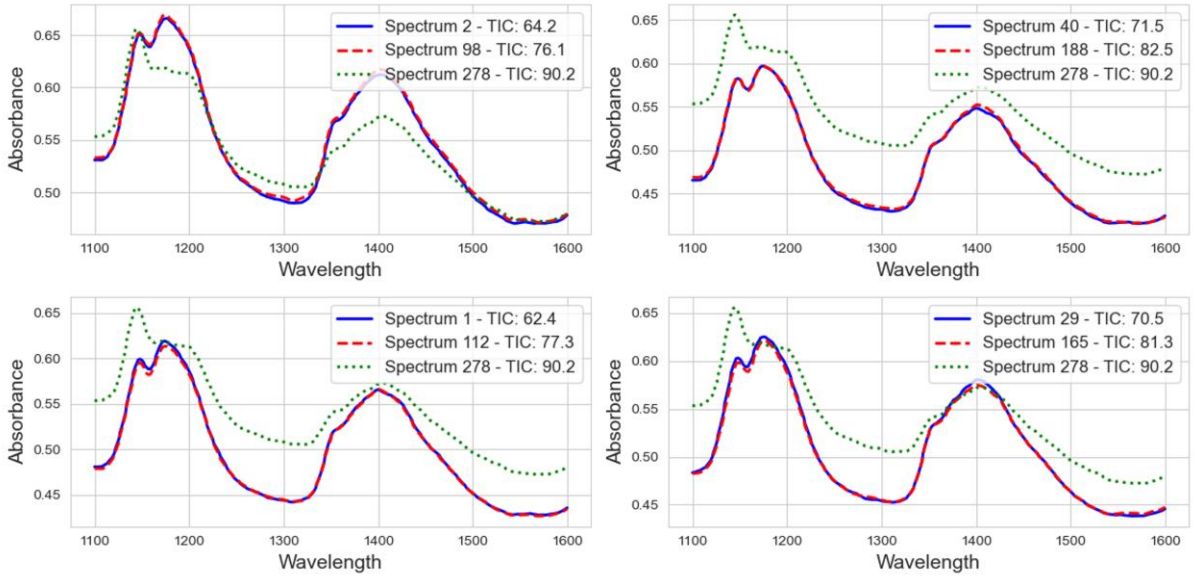

Fig. 10. Spectral similarity in neighbouring classes with different TIC values

For instance, spectrum 2 (TIC: 64.2) from class 1 (TIC < 70%) and spectrum 98 (TIC: 76.1) from class 2 (70% ≤ TIC < 80%) are visually almost indistinguishable, with only minor differences in the intensity around specific wavelengths near 1250 nm and 1400 nm. A similar pattern is observed in the upper-right subplot, where spectrum 40 (TIC: 71.5) from class 2 and spectrum 188 (TIC: 82.5) from class 3 (80% ≤ TIC < 85%) exhibit highly overlapping spectral profiles. This overlap is consistent across the other subplots, such as the bottom-left comparison between spectrum 1 (TIC: 62.4) from class 1 and spectrum 112 (TIC: 77.3) from class 2, as well as the bottom-right comparison between spectrum 29 (TIC: 70.5) from class 2 and spectrum 165 (TIC: 81.3) from class 3.

A high TIC sample (spectrum 278, TIC: 90.2) is included as a reference curve across all subplots in Fig. 10. This sample belongs to TIC class 4 (TIC ≥85%), showing spectral characteristics that are different from its neighbouring lower-TIC classes. Particularly in the first absorption peak (1100 nm - 1200 nm) and the broader absorption region between 1300 nm and 1500 nm. These different features are more easily separable compared to the minor variations observed among samples from adjacent lower TIC classes, where the spectral curves are largely overlapping.

The spectral similarity between samples from neighbouring classes suggests that the spectral data contains features that are difficult to distinguish using traditional methods. This overlap likely arises because the differences in TIC values may not produce sufficiently unique spectral signatures across the wavelengths measured using the current NIR spectrometer. Additionally, the presence of noise, variations in coal/stone dust particle sizes and environmental uncertainty can also affect spectral differences between classes.

4. CONCLUSIONS AND FUTURE WORK

This study presented an end-to-end convolutional neural network (CNN) model with residual blocks for total incombustible content (TIC) based classification of coal/stone dust samples using Near-Infrared (NIR) spectral data. A total of 300 samples were used in the study, with 240 samples allocated for training and 60 samples for testing. Despite the relatively limited dataset size, the tested residual CNN model demonstrated encouraging results, particularly in predicting high TIC samples (TIC ≥ 85%), where distinct spectral patterns enabled high precision and recall. This indicates the potential of the residual 1D CNN architecture to process high-dimensional spectral data and capture both local and global features for TIC predictions in coal/stone dust samples.

However, the study also identified key limitations that constrain the model's overall performance. Data imbalance, particularly the minority of low TIC samples (TIC < 70%), and the significant spectral similarity between neighbouring TIC classes, 70–80% and 80–85%, severely impact the model's ability to distinguish the minor differences between classes. These factors lead to high misclassification rates in certain TIC ranges, especially for low TIC samples. Furthermore, the current CNN model has not been fully optimized; the architecture and hyperparameters were kept constant to focus on evaluating the impact of residual blocks. As a result, there remains significant potential for improving the model's performance through systematic optimization.

Future work will focus on overcoming these challenges by:

1) Addressing data imbalance; Improving class representation by collecting additional data from samples belonging to the minority TIC ranges to balance the dataset.
2) Enhancing feature separability; Applying advanced spectral preprocessing methods, such as derivatives or wavelength selection, to amplify minor differences between similar spectra and reduce spectral overlap.

3) Optimizing the deep learning model; Utilizing automated hyperparameter tuning techniques, such as Bayesian optimization, to refine key parameters for improved classification accuracy and generalization.

ACKNOWLEDGEMENT

This study was supported by the New South Wales Coal Services Trust, through the Project ID 20633 "System Demonstrator of a Portable NIR Spectrometer for Rapid Stone Dust Compliance Testing".

REFERENCES

1. Cao, W, Qin, Q, Cao, W, Lan, Y, Chen, T, Xu, S and Cao, X 2017. Experimental and numerical studies on the explosion severities of coal dust/air mixtures in a 20-L spherical vessel. *Powder Technology* **310**, 17-23.
2. Ji, Y, Ren, T, Wynne, P, Wan, Z, Ma, Z and Wang, Z 2016. A comparative study of dust control practices in Chinese and Australian longwall coal mines. *International Journal of Mining Science and Technology* **26(2)**, 199-208.
3. Luo, Y, Wang, D and Cheng, J 2017. Effects of rock dusting in preventing and reducing intensity of coal mine explosions. *International Journal of Coal Science & Technology* **4(2)**, 102-109.
4. Wedel, DJ, Belle, B and Kizil, MS 2015. *The Effectiveness of Rapid Stone Dust Compliance Testing in Underground Coal*. 2015 Coal Operators' Conference, The University of Wollongong, Wollongong, NSW, Australia, February, 11-13, 120-150.
5. Begum, N, Maiti, A, Chakravarty, D and Das, BS 2021. Diffuse reflectance spectroscopy based rapid coal rank estimation: A machine learning enabled framework. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **263**, 120-150.
6. Zou, L, Yu, X, Li, M, Lei, M and Yu, H 2020. Non-destructive identification of coal and gangue via near-infrared spectroscopy based on improved broad learning. *IEEE Transactions on Instrumentation and Measurement* **69(10)**, 8043-8052.
7. Andrés, J and Bona, MT 2006. ASTM clustering for improving coal analysis by near-infrared spectroscopy. *Talanta* **70(4)**, 711-719.
8. Li, Z, Fredericks, PM, Ward, CR and Rintoul, L 2010. Chemical functionalities of high and low sulfur Australian coals: A case study using micro attenuated total reflectance–Fourier transform infrared (ATR–FTIR) spectrometry. *Organic Geochemistry* **41(6)**, 554-558.
9. Chakravartula, SSN, Moscetti, R, Bedini, G, Nardella, M and Massantini, R 2022. Use of convolutional neural network (CNN) combined with FT-NIR spectroscopy to predict food adulteration: A case study on coffee. *Food Control* **135**, 108816.
10. Le, BT, Xiao, D, Mao, Y and He, D 2018. Coal analysis based on visible-infrared spectroscopy and a deep neural network. *Infrared Physics & Technology* **93**, 34-40.
11. Cobas, JC, Bernstein, MA, Martín-Pastor, M and Tahoces, PG 2006. A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data. *Journal of Magnetic Resonance* **183(1)**, 145-151.
12. Schulze, HG, Foist, RB, Okuda, K, Ivanov, A and Turner, RFB 2011. A model-free, fully automated baseline-removal method for Raman spectra. *Applied Spectroscopy* **65(1)**, 75-84.
13. Cao, A, Pandya, AK, Serhatkulu, GK, Weber, RE, Dai, H, Thakur, JS, Naik, VM, Naik, R, Auner, GW and Rabah, R 2007. A robust method for automated background subtraction of tissue fluorescence. *Journal of Raman Spectroscopy* **38(9)**, 1199-1205.

14. Prakash, BD and Wei, YC 2011. A fully automated iterative moving averaging (AIMA) technique for baseline correction. *Analyst* **136(15)**, 3130-3135.
15. Powers, D 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies* **2(1)**, 37-63.