International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-12, Issue-4; Apr, 2025
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.124.6

Simultaneous determination of Copper, Zinc and Nikel in Electroplating Waste water by UV-VIS Spectroscopy Combined with advanced Machine Learning and Deep Learning Models

Anh Nguyen Thi Lan[1], Hung Nguyen Tran[1], Huong Nguyen Thu[1], Binh Khuat Hoang[1], Hung Khong Manh[1], Huong Pham Thu[2], Thanh Nguyen Chi[2], Son Nguyen Manh[3], Thao Ta Thi[3]

[1]Institute of Chemistry and Materials, Academy of Military Science and Technology
[2] Institute of Information Technology, Academy of Military Science and Technology
[3]Faculty of Chemistry, VNU University of Science, Vietnam National University, Hanoi

Received: 03 Mar 2025,

Receive in revised form: 05 Apr 2025,

Accepted: 12 Apr 2025,

Available online: 25 Apr 2025

***Keywords*— deep learning, heavy metals, machine learning, PAN, simultaneous quantitation, spectrophotometry**

Abstract*— Monitoring to evaluate wastewater quality during the production process requires simple measurements and Realtime analysis as well. Among common methods for heavy metal analysis, the UV-VIS absorption spectroscopy is considered a potential analytical method due to its low cost and simple operation, direct online integration with treatment tanks. However, it faces limitations in simultaneously analyzing multiple metals due to overlapping absorption spectra. This study applied machine learning (ML) algorithms (Decision Tree (DT), Random Forest(RF) and deep learning (DL) models (Multilayer Perceptron - MLP, and 1D Convolutional Neural Network - 1D-CNN) to improve the accuracy of simultaneous quantitative analysis of three metals—Cu, Zn, and Ni in electroplating wastewater—based on VIS absorption spectra data of their colored complexes in aqueous solution with the PAN reagent in the presence of a surfactant. Large datasets were collected from UV-VIS spectra of 500 wastewater spiked samples in the range of 620-500 nm with a 1 nm interval, resulting in a dataset of size 500x121, followed by the application of ML and DL models using the Python programming language. Model performance was evaluated based on the correlation coefficient (R^2) and root mean square error (RMSE). Preprocessing methods such as first-order derivatives and Principal Component Analysis (PCA) were applied to reduce noise in the dataset before training with machine learning algorithms. Results showed that the 1D-CNN model outperformed the others, achieving $R^2 > 0.88$ and RMSE < 0.036 for all three analytes. It is supposed by its ability to directly extract nonlinear features from raw data without the need for dimensionality reduction. In contrast, the DT, RF, and even MLP models, which utilized principal component analysis (PCA) for dimensionality reduction, demonstrated significantly lower accuracy due to information loss during the reduction process. The proposed model was successfully applied for rapid and simple metal concentration

determination in practical samples using a test kit with reagent, a compact spectrophotometer, and an automated PC-based data reading application. These findings demonstrate that combining UV-VIS spectroscopy with machine learning and deep learning algorithms is an effective and feasible approach for the simultaneous detection of multiple heavy metals in specific matrix wastewater samples.

I. INTRODUCTION

Anthropogenic emissions of heavy metals from industrial sectors such as steel production, metallurgy, electroplating, battery and accumulator manufacturing, metal surface treatment, recycling, and the chemical and fertilizer industries contribute to the bioaccumulation of these toxic elements in the human body, posing a significant public health concern [1]. Within the electroplating industry, transition metals including copper (Cu), zinc (Zn), and nickel (Ni) are routinely employed to form thin metallic coatings on substrates, typically composed of metal or plastic, with the objective of enhancing mechanical durability, providing corrosion resistance, and improving surface aesthetics [2]. This process results in the release of heavy metals through activities such as post-plating rinsing, solution spillage, and tank cleaning, with concentrations ranging from a few milligrams per liter (mg/L) to several hundred mg/L [3]. Before this wastewater is discharged into the environment, it must be treated using methods such as coagulation, adsorption, or ion exchange to ensure it complies with the permissible limits for Cu, Zn, and Ni as defined by QCVN 40:2021/BTNMT—2 mg/L, 3 mg/L, and 1 mg/L respectively—prior to release into non-potable water bodies (e.g., rivers and lakes used for irrigation, transportation, etc.) [4]. Exceeding these limits can pose serious health risks: copper, due to its oxidative properties, can damage red blood cells [5]; nickel exposure may lead to kidney toxicity, allergic reactions, contact dermatitis, and even cancer [6]; and although zinc is generally less toxic, excessive intake can still be harmful, increasing the risk of cardiovascular disease, kidney stones, and other health issues [7].

In addition to measuring the concentrations of Zn, Cu, and Ni in treated wastewater prior to environmental discharge using standard analytical techniques such as Atomic Absorption Spectroscopy (AAS) [8], Inductively Coupled Plasma Optical Emission Spectroscopy (ICP/OES) [9], and Inductively Coupled Plasma Mass Spectrometry (ICP/MS) [10], it is practically essential to perform rapid analysis of these metals at various stages of the production process. Real-time monitoring of their concentrations before discharge is crucial for timely and effective adjustments to the wastewater treatment process. A simple method such as UV-VIS molecular absorption spectroscopy, which offers suitable sensitivity, speed, accuracy, and the ability to simultaneously analyze all three metals using a single color-forming organic reagent, requires resolving the overlapping of absorption spectra [11]. To address this, chemometrics, machine learning and deep learning methods have been employed to optimize the advantages of traditional techniques such as UV-VIS and overcome their limitations.

In spectrophotometry for heavy metal analysis, PAN (1-(2-Pyridylazo)-2-naphthol) is a widely used chemical reagent due to its ability to form colored complexes with various transition and heavy metals such as Cu^{2+}, Zn^{2+}, Ni^{2+}, Fe^{3+}, Mn^{2+}, Cd^{2+}, Pb^{2+}, and others. These complexes exhibit characteristic absorption spectra in the visible range (typically between 500–580 nm), enabling the use of UV-Vis spectroscopy to determine metal concentrations [12]. However, when analyzing multiple metals simultaneously in a single sample (e.g., industrial wastewater), the absorption spectra of PAN-metal complexes often overlap, making it difficult to quantify individual metals using traditional spectroscopic methods. To overcome this challenge, chemometrics—a field that applies mathematical and statistical methods to chemical data—has been integrated with spectroscopy to allow for the simultaneous analysis of multiple metals in complex mixtures. Several studies have explored the use of PAN in micellar media for the spectrophotometric determination of heavy metals in combination with chemometric techniques. For instance, one study demonstrated the simultaneous determination of iron(II), nickel(II), and cobalt(II) by applying partial least squares (PLS) regression, both with and without a preprocessing step using direct orthogonal signal correction (DOSC) [13]. In another study, Cu(II), Co(II), and Ni(II) were simultaneously quantified using various mother wavelets from the continuous wavelet transform (CWT) family under optimized conditions for multi-component analysis. The method was validated through the analysis of synthetic ternary mixtures of Cu(II), Co(II), and Ni(II) ions and successfully applied to real sample matrices, including multivitamin tablets, tea, barley, spinach, chocolate, milk powder, soil, seawater, and tap water [14].

Machine learning (ML) and deep learning (DL) have become widely adopted in data-driven approaches for processing outputs from traditional analytical methods. They represent a modern, rapid, and cost-effective research direction for the simultaneous determination of multiple components within a single mixture, while minimizing interference from the sample matrix or other analytes. By leveraging ML and DL techniques, it is feasible to perform simultaneous analysis without prior separation, utilizing the full visible (Vis) spectral range to achieve high analytical accuracy.

This study aims to apply selected traditional ML models including Decision Tree (DT) and Random Forest (RF) and DL techniques using Multilayer Perceptron (MLP) and one-dimensional Convolutional Neural Network (1D-CNN) for the simultaneous determination of zinc (Zn), copper (Cu), and nickel (Ni) in wastewater from an electroplating factory, based on visible (Vis) absorption spectra of their complexes with PAN in the presence of the non-ionic surfactant Tween X. The approach involves three main stages: first, the development of a mathematical model; second, validation of the model using simulated data with added random noise; and finally, application to experimental spectrophotometric data obtained from ternary metal ion mixtures in aqueous solutions.

II. MATERIALS AND METHODS

2.1. Instruments and software

Spectroscopic measurements were conducted using the UH 4150 UV-VIS-NIR spectrophotometer (Hitachi), with a glass cuvette having a path length (d) of 1 cm. pH measurements were performed using the HANNA Instrument 211 microprocessor pH meter. All weighing operations were carried out using the HR-200 analytical balance from AND (Japan), with an accuracy of 0.0001 g. All absorption spectra were recorded and exported using the UV-Win PC software into Microsoft Excel for statistical processing. Chemometric-supported spectroscopic measurements were performed using Python software.

2.2. Materials and reagents

Individual heavy metal standard solutions (Cu^{2+}, Ni^{2+}, and Zn^{2+}) were diluted from standard solutions with a concentration of 1000 mg/L, sourced from Merck, Darmstadt, Germany. All chemicals used were of analytical purity, and double-distilled water was employed to prepare all solutions.

The buffer solutions used for optimization studies included Ammonia buffer (pH 10), Sodium Phosphate buffer (pH 7), Acetate buffer (pH 5), and Britton-Robinson buffer (pH 3). These solutions were freshly prepared and adjusted using a pH meter. The 0.004 M PAN reagent solution was prepared by dissolving 0.1 g of PAN indicator in 100 mL of double-distilled water. The resulting solution was stored in a dark, tightly sealed bottle and used within one week of preparation.

2.3. Sample collection and sample preparation

Wastewater samples from electroplating processes were collected from various sites in a factory located in Bac Ninh province, Vietnam. Immediately after collection, each sample was acidified by adding 3 mL of 55% concentrated nitric acid (HNO_3) per 1000 mL of sample to preserve metal ions and prevent precipitation. Before analysis, the sample was filtered to remove insoluble particulates by filter paper (Whatman). The filtrate was then neutralized to pH 7 and diluted to a final volume of 50 mL with deionized water before being analyzed by UV-Vis spectrophotometry.

2.4. Analytical Procedure

Construction of calibration set and model development

The machine learning model was developed using a UV-Vis spectral dataset comprising 500 spiked samples, derived from the wastewater matrix of an electroplating facility. These samples, originally characterized for their native concentrations of copper (Cu), zinc (Zn), and nickel (Ni), were spiked with known quantities of standard metal solutions to obtain specific target concentrations. The spiking design ensured a representative range of concentrations, with Cu ranging from 0.32 to 2.04 mg/L, Zn from 0 to 0.4 mg/L, and Ni from 0.165 to 0.975 mg/L.

To facilitate model development, the dataset was partitioned into training and testing sets—C_train and A_train for the concentration and absorbance training data, and C_test and A_test for the corresponding test data—using an 80:20 split ratio. These data matrices were then used as input for various machine learning algorithms implemented in Python. Each model was evaluated to identify the most suitable algorithm for accurate and robust prediction of metal concentrations based on the spectral data.

Construction of validation set and assessment of model performance

A comprehensive statistical evaluation was conducted to assess the performance of the developed machine learning and deep learning models. The analytical accuracy of the models was characterized in terms of trueness and precision. The trueness was quantified by measuring the degree of agreement between the predicted concentrations generated by the models and the known reference concentrations in the training and testing datasets. This agreement was evaluated using the coefficient of determination (R^2) between the results calculated by the model (yi') and the actual contents of heavy metals in the training or test set (yi), $yi*$ is the mean value of the observations yi:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - yi')^2}{\sum_{i=1}^{n}(yi - yi*)^2} \qquad (1);$$

The precision was assessed through the root mean square error (RMSE), which reflects the dispersion of predicted values around the actual values.

The formula of the root mean square error RMSE:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - yi')^2} \qquad (2);$$

- yi is the actual value at point i ;
- yi' is the predicted value at point i;
- n is the number of data points.

III. RESULTS AND DISCUSSION

3.1. Optimization of Experimental Conditions for Color Complex Formation

To ensure the accuracy, sensitivity, and reproducibility of spectrophotometric determination of metal ions using PAN (1-(2-Pyridylazo)-2-naphthol), a series of experiments were conducted to optimize the key parameters affecting the formation of the colorimetric complexes. The following factors were systematically investigated:

3.1.1. Wavelength Range Selection

The absorption spectra of representative samples—including standard solutions, real wastewater, and spiked wastewater samples—were recorded to identify the wavelength region with the most significant analytical signals. The selected wavelength range allowed for focused analysis, minimizing noise and reducing computational load during model development. The result obtained in figure 1 show that the absorption spectra of the mixture of three standard substances exhibited overlapping maximum absorption peaks, with no distinct absorption maxima characteristic of each individual metal. Therefore, it is not possible to separately analyze each metal in the mixture using the UV-Vis method alone, necessitating the integration of multivariate algorithms to address the quantitative analysis of each metal in this mixture. Furthermore, this allowed the identification of a significant signal wavelength range of 500 nm to 620 nm. Experiments to construct matrices for the development of machine learning programs and algorithms for future software construction will focus solely on this wavelength range to save experimental time.

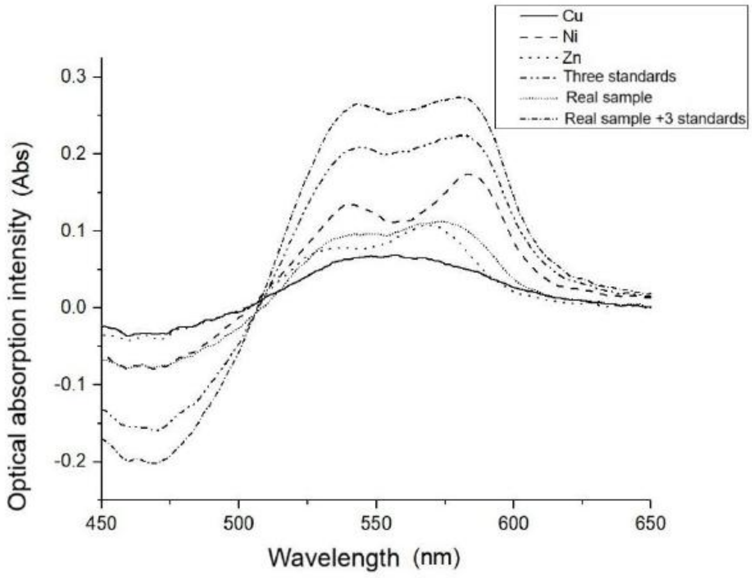


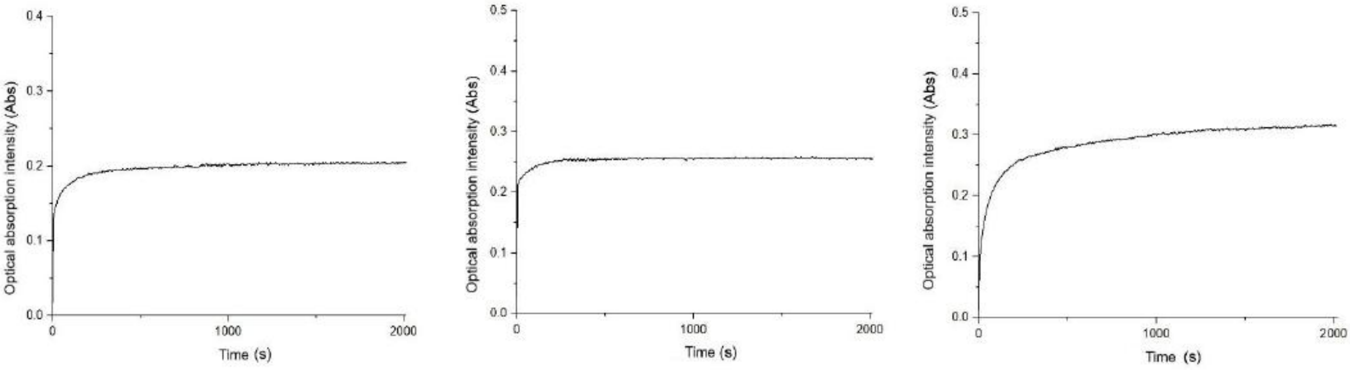
Fig.1. Absorption spectra of standard samples, real samples, and spiked real samples after complexation with PAN

3.1.2. Complexation Time Optimization

The time-dependent behavior of absorbance intensity for each metal-PAN complex was studied to determine the optimal complexation time for stable and reproducible measurements. The absorbance of the PAN–Cu complex increased rapidly during the initial 10–100 seconds, followed by a slower rise from 100–200 seconds, and reached a plateau after approximately 250 seconds, indicating the formation of a stable complex. The PAN–Ni complex exhibited a rapid increase in absorbance within the first 200 seconds, after which the rate of change gradually slowed. Stability was achieved only after approximately

1500 seconds (~25 minutes), suggesting slower complexation kinetics compared to Cu. The PAN–Zn complex showed a rapid increase in absorbance initially, but the signal remained unstable even after 250 seconds, indicating more variable or slower kinetics in complex stabilization.

Based on these observations, a resting time of 25 minutes (1500 seconds) was selected as the optimal complexation period for all three analytes. This duration ensured that all metal–PAN complexes, particularly the slower-stabilizing Ni–PAN and Zn–PAN complexes, had sufficient time to reach full complexation and stable absorbance. This standardized waiting period enhanced the reliability and consistency of spectrophotometric measurements across all samples (Figure 2).



PAN-Zn PAN- Cu PAN-Ni

Fig.2. Effect of time on the formation of colored complexes

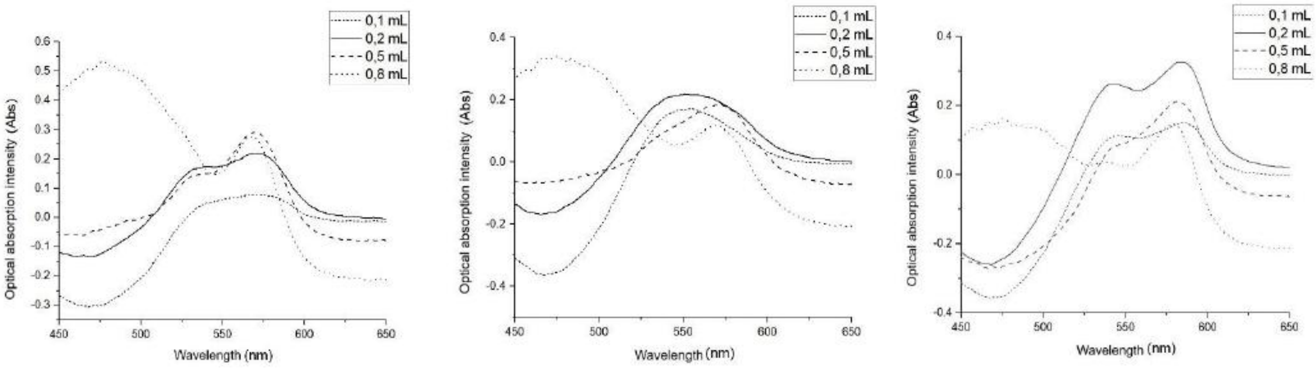
3.1.3. Reagent Quantity Optimization

The influence of reagent volume on the formation and stability of the metal–PAN complexes was systematically investigated. Standard solutions containing 0.5 ppm of each metal ion (Cu^{2+}, Ni^{2+}, Zn^{2+}) were reacted with varying volumes of 0.004 M PAN in a final mixture volume of 10 mL to determine the optimal reagent concentration.

A consistent trend was observed across all analytes:

As the volume of PAN reagent increased, the absorbance intensity of the resulting complexes also increased, indicating enhanced complex formation due to greater availability of the ligand. However, when the PAN volume exceeded 0.5 mL, the absorbance signals began to exhibit distortions and signs of spectral interference, likely due to excess reagent interacting with the spectrophotometric background or forming non-specific complexes.

Based on these findings, 0.5 mL of 0.004 M PAN in a 10 mL final volume was identified as the optimal reagent quantity for forming stable and measurable metal–PAN complexes. This volume ensured sufficient reagent availability for complete complexation without introducing absorbance interference or baseline instability (Figure 3).



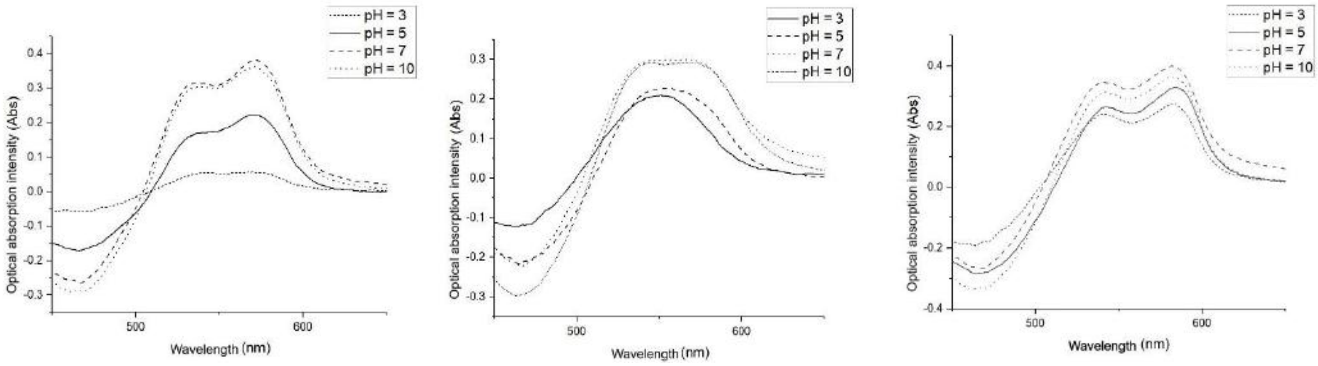
PAN-Zn PAN-Cu PAN-Ni

Fig.3. Effect of reagent volume on the absorbance of complexes within the studied spectral range

3.1.4. pH Optimization

The effect of pH on the formation and stability of the metal–PAN complexes was evaluated by preparing standard solutions of Cu^{2+}, Ni^{2+}, and Zn^{2+} (each at 0.5 ppm) and reacting them with 0.1% PAN in buffer solutions at different pH levels (pH 3, 5, 7, and 10). The results revealed that at pH 3 and pH 5, the absorbance signals were present but relatively low and less stable, likely due to incomplete complexation or protonation of functional groups on the PAN ligand that hinder metal binding. At pH 10, although complexation occurred, the absorbance was inconsistent, possibly due to hydrolysis or precipitation of metal ions at high alkalinity. At pH 7, the metal–PAN complexes demonstrated the highest absorbance intensity and excellent signal stability, indicating optimal complex formation under neutral conditions.

Therefore, pH 7 was selected as the optimal pH condition for the spectrophotometric analysis, as it promotes maximum complex stability and absorbance response across all three target metals.



PAN- Zn PAN- Cu PAN- Ni

Fig.4. Effect of pH on the complexation process of Cu^{2+}, Ni^{2+}, and Zn^{2+} with the PAN Reagent

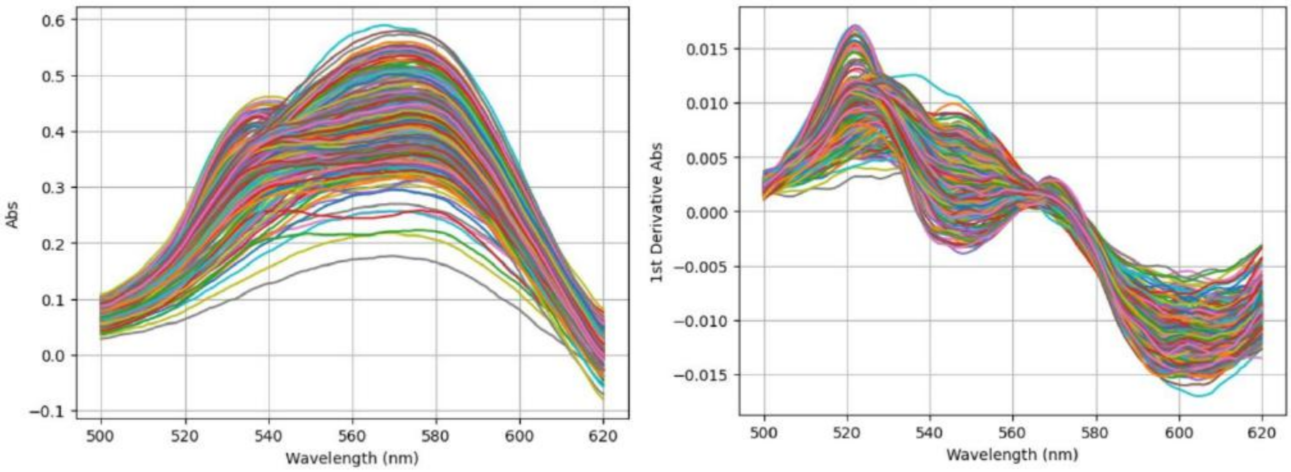


Fig.5: raw spectra containing raw data of 500 colored complex solutions (a) *and 1st derivative spectra (b)*

3.2. Preprocessing before using ML and DL

The preliminary analysis using machine learning (ML) and deep learning (DL) models on the unprocessed UV-Vis spectral data demonstrated limited predictive performance. Specifically, the models using raw spectra (figure 5a) yielded relatively low accuracy, with coefficient of determination (R^2) values ranging from 0.80 to 0.85, and poor precision, as indicated by root mean square error (RMSE) values exceeding 1. In addition, the computational time required for training and prediction was substantially prolonged due to the high dimensionality and noise inherent in the raw spectral data. To address these limitations, data preprocessing was deemed essential for noise reduction of interferences in matric of samples through spectral transformation methods such as first or second derivatives, which enhance signal quality by eliminating baseline drift and smoothing fluctuations. Besides that, dimensionality reduction, for example using Principal Component Analysis (PCA), to extract the most informative features from the spectral data while minimizing redundancy and computational load**.**

3.2.1. Derivative of entire spectra

Noise reduction through spectral transformation methods such as first or second derivatives, which enhance signal quality by eliminating baseline drift and smoothing fluctuations.

3.2.2. Reducing the Size of Input Data Using PCA

A Given the high dimensionality of the original input data matrix (121 spectral variables across 500 samples), Principal Component Analysis (PCA) was employed as a dimensionality reduction technique to transform the dataset into a lower-dimensional space while preserving its essential information.

PCA works by identifying new orthogonal variables—principal components (PCs)—which represent directions of maximum variance in the data. These PCs are linear combinations of the original variables and are ranked according to the amount of variance they explain.

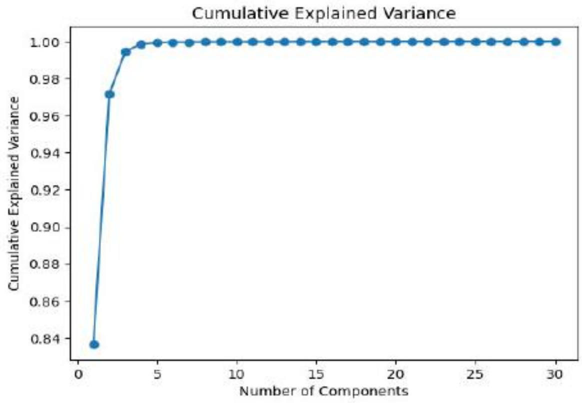


Fig.6. Percentage of Variance Explained vs. Number of Principal Components (PCs)

Analysis of the explained variance ratio (as illustrated in Figure 6) revealed that over 99% of the total variance is retained within the first five principal components. Therefore, the dataset was reduced to a 5-dimensional feature space, significantly simplifying the computational burden while maintaining the integrity of the original spectral information.

This step is crucial for enhancing model performance by removing redundant and noisy variables, reducing overfitting, particularly in machine learning models trained on high-dimensional data, and decreasing computational time in both model training and prediction phases.

3.2. Selection of Suitable Algorithms and Evaluation of Models

3.2.1. Machine learning model validation

The Decision Tree model was optimized using a set of hyperparameters designed to balance model complexity with generalization performance. Specifically, a max_depth of 80 was selected to constrain the tree's growth, thereby mitigating overfitting and enhancing model interpretability. The random_state was fixed at 24 to ensure reproducibility of results through consistent random number generation. The parameter min_samples_split was set to 10, defining the minimum number of samples required to split an internal node, which helps prevent the model from learning noise in the training data. Additionally, min_samples_leaf was set to 5 to ensure a minimum number of samples at each terminal node, promoting smoother decision boundaries and reducing model variance. This hyperparameter configuration was determined based on cross-validation performance to achieve an optimal balance between predictive accuracy and computational efficiency.

The transformed dataset after using was then used to train a Random Forest regression model optimized with the following hyperparameters: 100 estimators to ensure robust aggregation and minimize variance, a maximum tree depth of 40 to control model complexity and prevent overfitting, a fixed random state of 12 for reproducibility, a minimum sample split of 2 to allow for fine-grained pattern learning, and a minimum sample per leaf of 1 to maximize flexibility in capturing subtle variations. This configuration was selected based on cross-validation outcomes to strike a balance between predictive accuracy, generalization ability, and computational performance.

Among the evaluated models, PCA-DT demonstrated the weakest performance, with R^2 values of 0.269 (Zn), 0.3 (Cu), and 0.276 (Ni), indicating that the model explains less than 30% of the data variance and fails to capture essential patterns. Correspondingly, RMSE values ranged from 0.359 (Ni) to 0.471 (Zn), reflecting substantial prediction errors. The poor performance is attributed to PCA's removal of low-variance components, which may contain nonlinear or compound-specific information critical for Decision Trees—an algorithm inherently limited in modeling complex relationships. Additionally, PCA can distribute noise across components, further degrading performance in noise-sensitive models like Decision Trees.

Although the PCA-Random Forest (PCA-RF) model demonstrates improved performance over the PCA-Decision Tree (PCA-DT)—with R^2 values reaching approximately 0.6 for Zn and Ni—it still shows limited predictive ability for Cu, with an R² of only 0.2. This disparity suggests that while the ensemble structure of Random Forest enhances the model's ability to capture nonlinear relationships and reduce the effect of noise, the dimensionality reduction step using PCA may have eliminated essential features specific to Cu prediction. Consequently, despite outperforming PCA-DT, the PCA-RF model remains constrained by the information loss introduced by PCA, particularly affecting the accurate quantification of certain analytes.

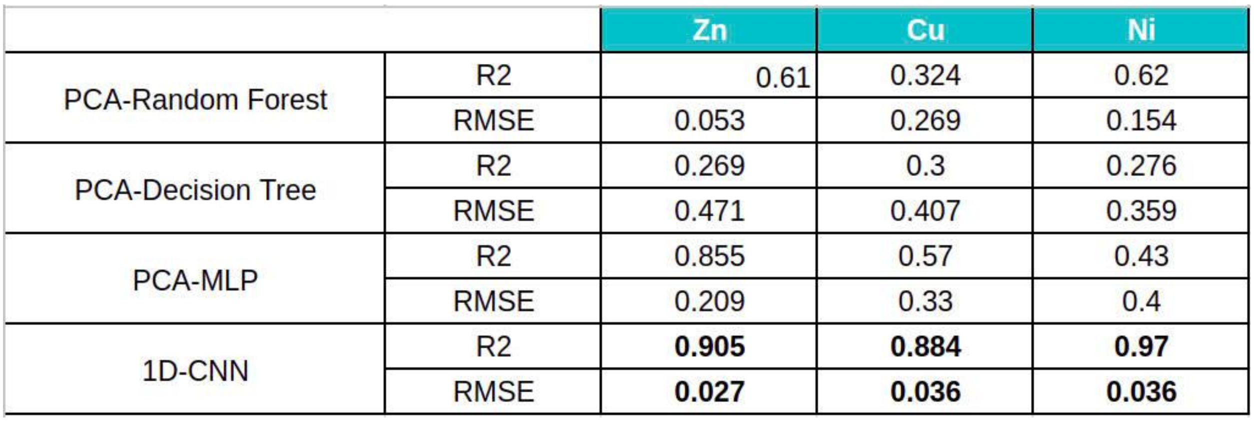
3.2.2. Validation of Deep learning models

The PCA-MLP (Principal Component Analysis–Multilayer Perceptron) model was constructed using a neural network architecture consisting of four layers with 5 input neurons (corresponding to the number of retained principal components), followed by two hidden layers containing 50 and 25 neurons, respectively, and a single output neuron for regression prediction. The activation functions used were ReLU for the hidden layers to introduce nonlinearity and enable learning of complex relationships, and a linear activation function at the output layer to facilitate continuous output for regression tasks.

The PCA-MLP model utilizes input data reduced via Principal Component Analysis (PCA) to enhance computational efficiency and reduce redundancy. The architecture comprises multiple fully connected hidden layers, where each neuron is linked to all neurons in adjacent layers. Nonlinear activation functions, such as ReLU, are applied within these layers to enable the model to learn complex nonlinear relationships in the spectral data. While the model achieves strong predictive performance for Zn (R² = 0.855, RMSE = 0.209), its effectiveness is notably lower for Cu and Ni, indicating element-specific limitations. Moreover, the model's complexity requires extensive hyperparameter tuning and longer training times, reflecting higher computational demands.

The 1D-CNN (One-Dimensional Convolutional Neural Network) model was designed to process the full spectral data (121 input features) using a sequential convolutional architecture. The network includes four Conv1D layers with increasing filter sizes: 32, 64, 128, and 100, all utilizing a kernel size of 3 to capture local spectral patterns. ReLU was employed as the activation function throughout the convolutional layers to introduce nonlinearity. A final dense output layer with a single neuron was used for regression prediction. This architecture enables the model to learn hierarchical features from the spectral data while maintaining robustness and flexibility in capturing complex relationships.

Table 1. Statistical parameters obtained by the Machine learning and deep learning method



		Zn	Cu	Ni
PCA-Random Forest	R2	0.61	0.324	0.62
	RMSE	0.053	0.269	0.154
PCA-Decision Tree	R2	0.269	0.3	0.276
	RMSE	0.471	0.407	0.359
PCA-MLP	R2	0.855	0.57	0.43
	RMSE	0.209	0.33	0.4
1D-CNN	R2	**0.905**	**0.884**	**0.97**
	RMSE	**0.027**	**0.036**	**0.036**

The 1D-CNN model demonstrates superior predictive performance, with R² values ranging from 0.884 to 0.97 and RMSE values between 0.027 and 0.036, owing to its capacity to extract complex spatial and nonlinear features from spectroscopic data. Unlike PCA-based models, 1D-CNN operates directly on the full spectral input, preserving all intrinsic patterns and avoiding potential information loss associated with dimensionality reduction. Its convolutional and pooling layers enable effective automatic feature extraction and noise suppression by emphasizing local signal patterns, making it highly robust in complex matrices. This advantage is particularly evident in the accurate prediction of Cu, where earlier models underperformed.

It can be said that Models incorporating PCA—such as PCA-DT, PCA-RF, and PCA-MLP—are notably impacted by the loss of critical spectral features during dimensionality reduction, particularly for analytes like Cu that exhibit weak or overlapping absorption bands. Among them, PCA-DT performs the worst, with R² values around 0.27–0.3, highlighting its limited capacity to model nonlinear relationships in noisy environments. Although PCA-RF and PCA-MLP show marginal improvements due to ensemble learning and nonlinear activation functions, they still fall short in achieving reliable predictive accuracy.

Conversely, the 1D-CNN model not only delivers superior accuracy but also inherently reduces data dimensionality via pooling layers, eliminating the need for external preprocessing techniques like PCA. This end-to-end learning capability allows 1D-CNN to retain and extract essential spatial and nonlinear features from UV-Vis spectra, even in complex sample matrices. Its robustness to noise and overlapping signals makes it particularly suitable for real-world applications. As a result, integrating UV-Vis spectroscopy with 1D-CNN offers a powerful, efficient, and scalable solution for simultaneous multi-metal quantification in environmental monitoring and industrial wastewater analysis.

3.3 Analysis of real samples and comparison with AAS techniques

To evaluate the effectiveness and accuracy of the newly developed rapid analysis program, the concentrations of Zn, Cu, and Ni were determined in 5 samples with arbitrary concentrations within the range the program was trained on. The obtained results were then compared with those obtained using the AAS method (Table 2).

Table 2. Compare the results obtained by UV/CNN method with AAS method

	Zn^{2+}			Cu^{2+}			Ni^{2+}		
No	AAS method	UV/CNN method	% error	AAS method	UV/CNN method	% error	AAS method	UV/CNN method	% error
1	0,035	0,034	2,29	0,318	0,321	1,12	0,165	0,176	6,67
2	0,05	0,054	7,40	0,660	0,691	4,67	0,320	0,341	6,56
3	0,071	0,076	7,04	0,790	0,813	2,87	0,051	0,055	7,84
4	0,73	0,754	3,29	1,040	1,111	6,87	0,420	0,440	4,76
5	0,12	0,110	8,33	0,379	0,351	7,40	0,850	0,808	4,92
R^2	99.9			99.7			99.6		

Based on the comparison of Zn, Cu, and Ni concentrations obtained from the AAS method and the UV/CNN method under investigation, it is observed that the correlation between the two methods is excellent, with $R^2 > 99$. Additionally, the deviation between the results from the UV/CNN method and the reference AAS method is consistently below 10%. These findings suggest that the UV/CNN method is a promising approach for the rapid, easy, and accurate determination of metal concentrations in wastewater from electroplating processes.

IV. CONCLUSION

This study demonstrates the viability of integrating ultraviolet-visible (UV-VIS) absorption spectroscopy with machine learning (ML) and deep learning (DL) algorithms for the simultaneous quantification of heavy metals—namely zinc (Zn), copper (Cu), and nickel (Ni)—in complex wastewater matrices. Conventional UV-VIS analytical approaches are often constrained by spectral overlap and matrix interferences arising from coexisting contaminant ions, which can significantly hinder the accuracy and reliability of quantification. By leveraging data-driven ML techniques, it is possible to extract latent information from overlapping spectral profiles, thereby enhancing the robustness and precision of the analytical process. Among the models evaluated, the one-dimensional convolutional neural network (1D-CNN) exhibited superior performance, attributable to its capacity to autonomously capture nonlinear and spatial features directly from raw spectral data without requiring preliminary dimensionality reduction procedures such as principal component analysis (PCA). These findings affirm the feasibility and high potential of ML-based approaches—particularly deep learning models like 1D-CNN—for application in environmental analytical chemistry. Future research should focus on validating the generalizability of the proposed model across diverse sample matrices and under varying physicochemical conditions (e.g., natural environmental samples, elevated contaminant concentrations, fluctuations in pH and temperature). Moreover, the development of user-friendly software platforms to integrate UV-VIS spectral data processing with ML algorithms represents a promising avenue for facilitating rapid, cost-effective, and scalable deployment of this technology in real-world environmental monitoring.

ACKOWLEDGEMENTS

This work was financially supported by the Institute of Chemistry and Materials.

REFERENCES

[1] J Ogwuegbu Duruibe, MOC Ogwuegbu, and JN Egwurugwu (2007), Heavy metal pollution and human biotoxic effects. *International Journal of physical sciences* 2007, 2(5): 112-118.

[2] Amitesh, Devendra Dohare, Chhaya Rekhate, Abhinesh Kumar Prajapati, Removal of heavy metal from electroplating wastewater using electrocoagulation: a review, Desalination and Water Treatment 2022, 267: 121-137,

[3] Ahman ML, Wong ZJ, Sarjadi MS, Soloi S, Arshad SE, Bidin K, Musta B. Heavy Metals Removal from Electroplating Wastewater by Waste Fiber-Based Poly(amidoxime) Ligand. *Water*. 2021; 13(9):1260

[4] Ministry of Natural Resources and Environment. *National Technical Regulation on Industrial Wastewater*. QCVN 40:2021/BTNMT, 2021, Ministry of Natural Resources and Environment.

[5] Xiya Ren, Limei Zhao, et al. (2025), Copper-instigated modulatory cell mortality mechanisms and progress in kidney diseases. *Renal Failure*, 47(1), p. 2431142.

[6] Samuel Buxton, Emily Garman, et al. (2019), Concise review of nickel human health toxicology and ecotoxicology. *Inorganics*, 7(7), p. 89.

[7] Ananda S Prasad (2007), Zinc: mechanisms of host defense1. *The Journal of nutrition*, 137(5), p. 1345-1349.

[8] U.S. Environmental Protection Agency. *Method 7000B: Flame Atomic Absorption Spectrophotometry*. SW-846, Revision 2, 2007, U.S. Environmental Protection Agency.

[9] U.S. Environmental Protection Agency. *Method 6010D: Inductively Coupled Plasma-Optical Emission Spectrometry (ICP-OES)*. SW-846, Revision 4, 2018, U.S. Environmental Protection Agency.

[10] U.S. Environmental Protection Agency. *Method 6020B: Inductively Coupled Plasma-Mass Spectrometry (ICP-MS)*. SW-846, Revision 2, 2014, U.S. Environmental Protection Agency.

[11] Hu NaiLiang Hu NaiLiang, Gao HongWen Gao HongWen, et al. (2005), Simultaneous determination of cobalt and nickel in wastewater with 2-hydroxyl-5-benzeneazoformoamithiozone by spectral correction technique. *Journal of the Chinese Chemical Society*, 52(6), p. 1145-1152.

[12] Ashok Rao, K., Rangamannar, B. PAN as a reagent for the substoichiometric determination of nickel and manganese. *Journal of Radioanalytical and Nuclear Chemistry Letters* **93**, 319–326 (1985).

[13] K. Zarei, M. Atabati, Z. Malekshabani, Simultaneous spectrophotometric determination of iron, nickel and cobalt in micellar media by using direct orthogonal signal correction-partial least squares method, Analytica Chimica Acta,Volume 556, Issue 1,2006,Pages 247-254,

[14] Maryam Abbasi Tarighat,a and Abbas Afkhamib J. Simultaneous Spectrophotometric Determination of Cu(II), Co(II) and Ni(II) using Ratio Spectra-Continuous Wavelet Transformation in some Food and Environmental Samples. Braz. Chem. Soc., Vol. 23, No. 7, 1312-1319, 2012