Deep Neural Network (DNN) for Outlier Detection in Heavy Metal Voltammetric Data

Alya Kamilah
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya 60111, Indonesia
6025241066@student.its.ac.id

Riyanarto Sarno
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya 60111, Indonesia
riyanarto@its.ac.id

Kelly Rossa Sungkono
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya 60111, Indonesia
kelly@its.ac.id

Abdullah Faqih Septiyanto
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
7025221022@student.its.ac.id

Rizqy Ahsana Putri
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya 60111, Indonesia
6025231069@student.its.ac.id

Taufiq Choirul Amri
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
7025222001@student.its.ac.id

Fadlilatul Taufany
Department of Chemical Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
f_taufany@chem-eng.its.ac.id

Dwi Sunaryono
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya 60111, Indonesia
dwi@its.ac.id

Shoffi Izza Sabilla
Department of Medical Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
shoffi@its.ac.id

***Abstract*— The majority of existing outlier detection methods are unable to adequately address the inherent complexity and variability of voltammetry data, particularly in the context of heavy metal detection. The effectiveness, precision, and adaptability of these methods are constrained, particularly when confronted with complex datasets. To address this gap, this study proposes a novel outlier detection method combining Isolation Forest, One-Class SVM, Local Outlier Factor (LOF), Z-Score, and Deep Neural Network (DNN) techniques. The proposed approach is designed to enhance accuracy by leveraging the unique strengths of each method in identifying outliers within complex data. Evaluation results demonstrate that the DNN method achieves the highest accuracy, reaching 99.87%, significantly outperforming other tested models. This robust performance suggests that the proposed method has substantial potential to improve the reliability of detecting heavy metal anomalies in voltammetric data. Specifically, the DNN method demonstrated an F1-score of 1.00 across multiple evaluation metrics, outperforming other models such as Isolation Forest (F1-score: 0.65) and LOF (F1-score: 0.75).**

Keywords*— *Deep Neural Network (DNN), Heavy Metals detection, Outlier Detection, Voltamettry

I. Introduction

In statistical analysis, a data point or case that displays unusual features in relation to the overall distribution of the dataset is referred to as an outlier. Hawkins' seminal definition [2] defines an outlier as an observation that deviates significantly from the rest of the dataset, prompting the question of whether it was produced by a distinct mechanism. This definition is widely accepted within the statistical community. The four main factors that cause outliers are as follows: (1) data input mistakes, (2) computer program missing values, (3) observations that are not part of the sample population, and (4) extreme values that are not normally distributed, as outlined by [3].

Outliers in data can have a significant impact on the performance and analytical dependability of machine learning models, particularly in the context of heavy metal identification. Basic outlier identification techniques, such as Kalman filtering, have been employed in previous studies on heavy metal detection using voltammetry techniques to reduce data noise. Despite its efficacy in reducing noise, Kalman filtering is constrained in its capacity to detect complex outliers, which frequently arise from physical disturbances in voltammetry data, including temperature variations or surface contamination, particularly when a Glossy Carbon Electrode (GCE) is utilised.

Moreover, previous studies conducted by [6] employed the Moving Average Filter (MAF), Slope Identification, and Z-score to develop an outlier identification technique for photoplethysmograph (PPG) signals. When applied to more complex voltammetry data, this combination of techniques is not without limitations, despite its efficacy in noise reduction and extreme data detection. PPG-based techniques are not optimal for detecting heavy metals due to the unique challenges posed by voltammetry, including its sensitivity to changes in physical conditions.

Furthermore, additional outlier detection methods have been employed in general outlier identification research, including Isolation Forest, One-Class Support Vector Machine (SVM), and Local Outlier Factor (LOF). Isolation Forest [7] may be effective in managing outliers in large datasets, but it is not necessarily the optimal choice for voltammetry data, given that its distributions are often highly diverse. In contrast, the one-class SVM is more effective in scenarios where data exhibits uniform distributions. However, it tends to face challenges in cases where the data is imbalanced, such as in the context of heavy metal analysis. When the parameters are not accurately calibrated, the local outlier factor (LOF), which employs local density to identify outliers, often becomes unstable.

The suggestion of applying Deep Neural Network (DNN)s (DNN), which have the capacity to recognise intricate patterns and nonlinear correlations in voltammetry data, is intended to overcome these shortcomings. A method that has not yet been extensively used in heavy metal detection, DNN allows the model to uncover hidden patterns that are essential to locating outliers in voltammetry data.

The proposed method of this study to assess the efficacy of diverse outlier detection techniques in voltammetry data for the detection of heavy metals, with a particular emphasis on their reliability and accuracy. The findings of this evaluation are anticipated to provide insights into the most effective methods and their potential applications in enhancing the quality of future data analysis.

II. Research Method

A. Voltammetric Data

An electrochemical method known as cyclic voltammetry (CV) is often employed to examine intricate complexation processes, particularly in the field of environmental chemistry to gain insight into the binding of heavy metals to organic ligands [10][11]. In this study, voltammograms were generated to elucidate the relationship between current and potential, thereby revealing information about the electrochemical reactions occurring within the sample. These were produced using a Glossy Carbon Electrode (GCE) as the working electrode and a sintered Reference Electrode (RE) to maintain potential stability.

B. Isolation Forest

An anomaly detection technique, designated as Isolation Forest, is capable of expeditiously and efficaciously differentiating anomalous data points from those that conform to the norm [12]. It accomplishes this by constructing Isolation Trees (iTrees), which partition the data randomly until each point is isolated. Anomalies, which are generally sparse and significantly different, have a shorter path length in the tree, making them easier to detect. The anomaly score equation $s(x,n)$ in Isolation Forest shown in Eq. (1) :

$$s(x,n) = 2 - \frac{e(h(x))}{c(n)} \tag{1}$$

Where :

$s(x,n)$: Score for anomaly in data point x
e(h(x)) : Path length expected to isolate data point x
c(n) : Normalising constant

This Eq. (1). is employed to evaluate the anomalous nature of data points within the Isolation Forest, on the assumption that outliers are isolated with greater alacrity than other data points.

C. One-Class Support Vector Mechine (SVM)

A variant of the Support Vector Machine (SVM) methodology, specifically designed for the identification of outliers or anomalies within datasets, is known as the One-Class Support Vector Machine (One-Class SVM) [13]. As labelled outlier data is not required for training, this approach is classified as unsupervised machine learning. Rather, the objective of the One-Class SVM is to learn a decision boundary that distinguishes normal data from atypical data. The One-Class SVM is trained only on normal data (inliers). The objective of the one-class support vector machine (SVM) is to identify a hyperplane that separates the data points in the feature space from the origin. The decision shown in Eq.(2) :

$$f(d) = \omega \cdot \varphi(d) + b \tag{2}$$

Where:

$f(d)$: decision equation
ω : weight vector, determines hyperplane orientation.
$\varphi(d)$: maps input space to feature space via kernel.
b : bias, shifts the hyperplane.

The decision Eq. (2). in One-Class SVM is used to separate normal data from anomalies. Through the use of mapping data to a higher-dimensional space, the model learns accurate boundaries to distinguish normal data from outliers. In this study, data from three categories (Cadmium, Non-Heavy Metal, and Lead) were processed and normalized. The One-Class SVM with an RBF kernel was trained using cross-validation,treating Cadmium as an outlier labeled as -1 and the other categories as inliers labeled as 1. Model accuracy in detecting outliers was assessed using a confusion matrix and accuracy score.

D. Local Outlier Factor (LOF)

An unsupervised machine learning approach that employs density-based techniques is known as the Local Outlier Factor (LOF). By measuring the density of data points in their immediate neighbourhoods, as opposed to analysing the complete dataset, the LOF is able to identify outliers [14]. The LOF value for a data point is defined as Eq. (3) :

$$\mathrm{LOF}_{(xi)} = \frac{\sum xj \in Nk(xi)\,\frac{lrd(xj)}{lrd(xi)}}{|Nk(xi)|} \tag{3}$$

where:

Nk : Nearest neighbors
lrd : Local Reachability Density (LRD)

LOF equation (3) uses the parameter n_neighbors to detect outliers, with data classified as inliers (1) or outliers (0). Evaluation is performed using accuracy score and confusion matrix. In this study, LOF is used to detect outliers in three data categories: Cadmium, Non-Heavy Metal, and Lead. The data is loaded, normalized using StandardScaler, and LOF is applied with various values of n_neighbors to determine the neighbors in calculating local density. This model identifies the Cadmium category as an outlier and the other categories as inliers.

E. Z-Score

A statistical technique, known as the Z-score, is employed to ascertain the extent to which a given value deviates from the mean of a given data set. The Z-score is expressed in units of standard deviation from the mean, shown in Eq. (4) :

$$\mathrm{Z} = \frac{x - M}{s} \tag{4}$$

where:

Z : Z-score
x : observed value
M : Mean value
s : standard deviation

The Z-score may be obtained by utilising the decision equation (4). The Z-score is calculated by dividing the difference between a given value and the population mean by the population standard deviation. The degree of divergence is determined by the variations within the data. As data variability increases, the resulting variance also grows [15].

F. Deep Neural Network (DNN)s

A Deep Neural Network (DNN) (DNN) is an artificial neural network comprising multiple hidden layers, enabling the model to discern intricate patterns within data sets. This paper utilises a DNN technique with a model that employs the binary cross-entropy loss equation for the identification of outliers in binary classification. This loss equation calculates the difference between predictions and actual labels shown in Eq (5) :

$$BCE = -\frac{1}{n}\sum(yi\ \log(\hat{y}_l) + (1 - yi)\log(1 - \hat{y}i) \quad (5)$$

where:

y : true label
$\hat{y}$: model's prediction,
N : number of samples.

This Eq. (5). measures prediction error, where a lower loss indicates higher accuracy. The sigmoid activation in the output layer maps the output to 0 or 1, suitable for distinguishing between inliers and outliers. The data is standardized using StandardScaler for stability, and training is conducted with K-Fold Cross-Validation to avoid overfitting.

III. Proposed Method

The recommended methodology for this study is to categorise outliers as either legitimate or invalid data as shown in Fig. 1.

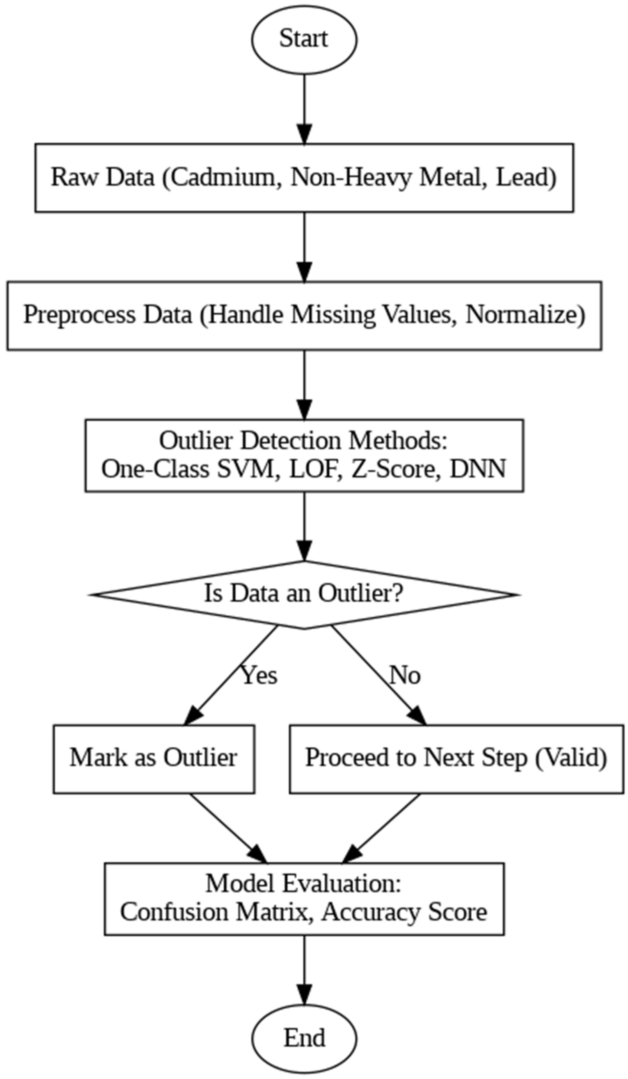

Fig. 1. Proposed Method

A. Preprocess Data

In the initial phase of data processing, the raw data was obtained from three primary categories: Cadmium, Non-Heavy Metal, and Lead. Each category was stored in a separate folder. Each data file within the aforementioned folders was accessed and read using a bespoke function that extracts the initial 802 rows and the first two columns from each Excel file. In the event that a file contained a number of rows that fell below the threshold of 802, the remaining entries were populated with NaN values in order to maintain consistency in size and facilitate further processing. Once all the data had been accessed, the DataFrames from each category were merged horizontally into a single, comprehensive DataFrame encompassing the entire dataset across the three categories. At this juncture, any instances of missing values were identified and all rows containing NaN were removed, thus ensuring that only data of the highest quality would be utilized for analysis.

Subsequently, the read data was transformed into a one-dimensional array to ensure consistency in shape. This data was then stored in a structured list, while the category labels were stored as numeric codes: 2 for Cadmium, 1 for Non-Heavy Metal, and 0 for Lead. To guarantee the absence of any remaining missing values, any remaining NaN positions were filled with zeroes.

The preprocessing stage proceeded with the application of standardisation using the StandardScaler function from the scikit-learn library. This process resulted in the data being normalised, whereby each feature exhibited a mean of zero and a standard deviation of one. Standardisation is of great importance in maintaining scale consistency across features, thus enabling outlier detection methods to operate with greater reliability and accuracy. The original data labels (2, 1, and 0) were then converted into a binary format for the purpose of inlier and outlier classification. In this format, label 1 was assigned to inliers (Non-Heavy Metal and Cadmium), while label 0 was assigned to outliers (Lead).

The preprocessing stage was thus successful in transforming the raw data into a form that was both clean and standardised, thereby rendering it fit for use in the outlier detection methods proposed in the subsequent analysis stage.

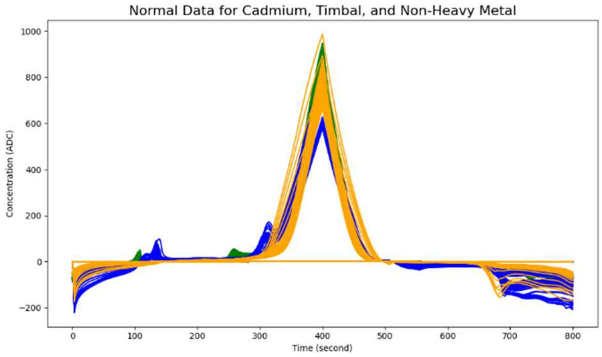

Fig. 2. Normal Data

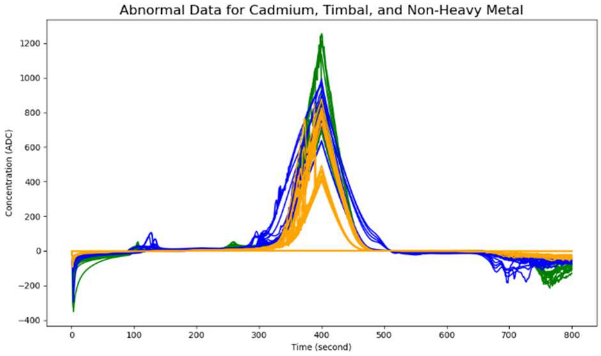

Fig. 3. Abnormal Data

B. *Outlier Detection*

The present study, several techniques for detecting outliers were applied to process data from three categories: Cadmium, Non-Heavy Metal, and Lead. Five approaches were used to effectively detect outliers: One-Class SVM, Isolation Forest, LOF, DNN, and Z-score. One-Class SVM demonstrated good results in terms of efficiency and speed, while Isolation Forest and LOF were more effective in detecting outliers based on the local density of the data. DNN, with its ability to handle more complex data, also showed solid performance, while Z-score offered a simple yet effective method for detecting outliers based on standard deviation from the data.

Each method was tested on a preprocessed and normalized dataset to minimize the influence of different variable scales. The application of these methods allowed for more accurate identification of outliers, with each method having its own strengths and weaknesses. Cross-validation was used to novelty the models more objectively, and the results from the confusion matrix and classification report showed that these models could effectively detect anomalies in complex datasets. Overall, the combination of these methods provides a comprehensive solution for outlier detection in large and diverse datasets.

Fig. 2. shows a graph of normal data for three categories: Cadmium, Timbal, and Non-Heavy Metal. The graph illustrates concentration (ADC) over time (seconds), with a significant peak around 400 seconds. The blue line represents data for Cadmium, orange for Non-Heavy Metals, and green for Lead. These variations highlight the differences in behavior across categories, crucial for identifying outliers. Each data category is marked with a different color (blue, orange, green) to distinguish between them. The sharp peak indicates a concentration spike, while the flat regions on the left and right of the graph represent stable concentration values. This graph demonstrates the variation in concentration across the three categories over time.

Fig. 3. displays abnormal data for three categories: Cadmium, Timbal, and Non-Heavy Metal, where concentration (ADC) is plotted against time (seconds). The data exhibits significant variation, especially around 400 seconds, where a sharp peak is observed, indicative of abnormal s pikes observed around 400 seconds indicate sensor irregularities, underscoring the need for robust outlier detection methods. Each color corresponds to a specific data category. Each data category isrepresented by a different color (blue, orange, green) to differentiate between the types. The abnormality is evident in the fluctuations around the peak and the unusual behavior observed in the tail end of the data. This graph highlights the presence of abnormal patterns or outlier s in the dataset, contrasting with the normal data plot.

IV. Experimental Result

A. *Data Material*

The material data used in this study consists of voltammetric data obtained through sampling with a three-electrode configuration. The data is divided into three main 750 data points. Of these, 500 data points are classified as normal data, while 250 are considered abnormal or outliers, caused by external disturbances such as sensor vibrations. Each test sample in each class represents concentration levels in ppm, with increments of 50, namely 50 ppm, 100 ppm, 150 ppm, 200 ppm, and 250 ppm.

Abnormal data was obtained by moving the sensor, causing unstable peak fluctuations that affected data quality. Various outlier detection methods were applied to compare model performance, aiming to identify the most effective model in predicting the number of inliers and outliers and to assess detection accuracy on the processed dataset. TABLE I displays a comparison of the outlier detection model accuracy using three evaluation metrics: Precision, Recall, and F1-Score. The models tested include Isolation Forest, One-Class SVM, LOF, Z-Score, and Deep Neural

TABLE I OUTLIER DETECTION MODEL EVALUATION

Model	Accuracy	Precision	Recall	F1-Score
Isolation Forest	53.20%	0.65	0.65	0.65
One-Class SVM	62%	0.75	0.65	0.7
Local Outlier Factor	60.40%	0.65	0.87	0.75
Z-Score	59.73%	0.66	0.8	0.73
Deep Neral Network	99.87%	1.0	1.0	1.0

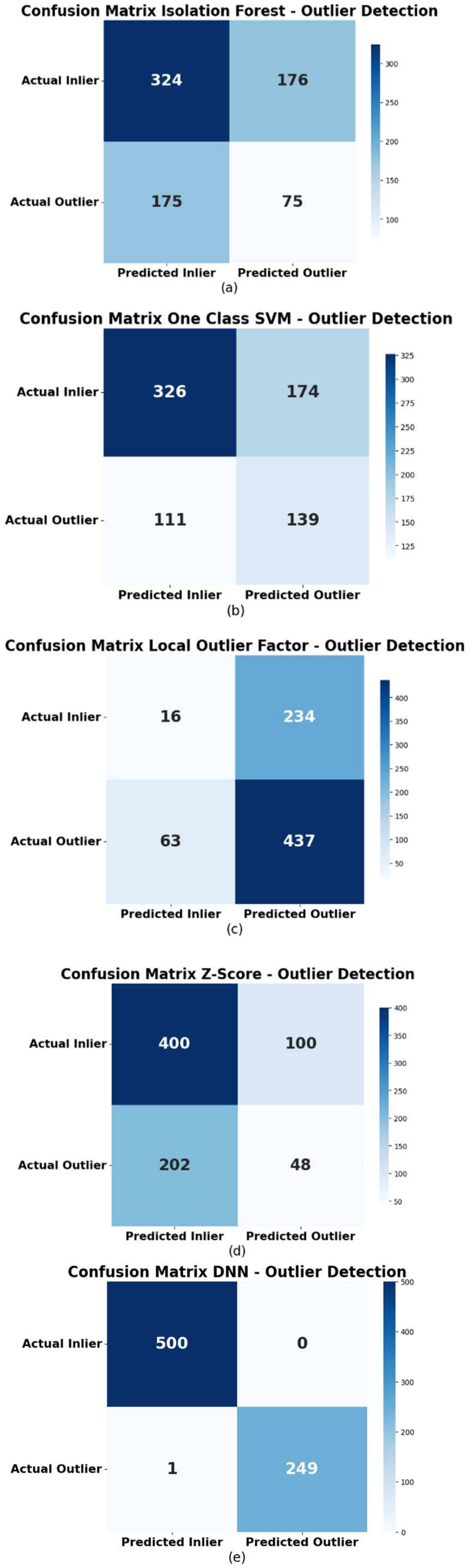

Fig.4 (a) Isolation Forest (b) One-Class SVM (c) LOF (d) Z Score (e) DNN

The objective is to establish a connection between the recall and precision metrics. The recall is defined as the percentage of actual positive data accurately detected out of all positive data in the dataset. In contrast, precision measures the percentage of genuine positive forecasts out of all positive predictions generated by the model. The F1-score represents a balanced indicator of the model's capacity to identify genuine positives while simultaneously reducing false

B. Confusion Matrix Analysis

To gain a deeper understanding of the model's performance, confusion matrices are novelty for each tested outlier detection model. The combined confusion matrix shows the distribution of correct and incorrect predictions for each category, highlighting a high frequency of correctly predicted inliers (data accurately predicted as inliers) and correctly predicted outliers (data accurately predicted as outliers) in certain categories. However, the presence of false positives and false negatives indicates areas where the model may lack precision, especially for rarer or more challenging categories. Overall, the confusion matrix reflects the model's ability to identify high-frequency categories and highlights areas that may require additional data or further model refinement to improve accuracy.

The confusion matrix in Fig. 4 show a comparison of the prediction results from various outlier detection models. Isolation Forest successfully classified 324 inliers and 75 outliers with reasonable accuracy, though there were some prediction errors. One-Class SVM had more misclassifications, especially for outliers, with 326 correctly predicted inliers and errors in predicting outlier data. LOF displayed several misclassifications of both inliers and outliers, while Z-Score showed better accuracy for inliers but still had significant errors in predicting outliers. DNN-demonstrated the best performance, correctly predicting 500 inliers with only 1 misclassified outlier, achieving perfect results compared to the other models. Overall, DNN excels in outlier detection accuracy, while the other models showed varying performance.

C. Evaluation

The evaluation using various applied methods shows that DNN achieves the best performance with perfect scores (1.00) across all three metrics, indicating its high effectiveness in detecting outliers. In contrast, One-Class SVM demonstrates moderate performance with a Precision of 0.56, Recall of 0.44, and F1-Score of 0.49, suggesting that this model performs reasonably well, although not as effectively as DNN. On the other hand, both Isolation Forest and LOF show lower perfmance, with LOF in particular exhibiting very low values on several metrics. Z-Score also displays moderate performance with a Precision of 0.32, Recall of 0.19, and F1-Score of 0.24, indicating its limited effectiveness compared to other models.

The DNN model achieved an accuracy of 99.87% using K-Fold Cross-Validation, significantly outperforming other models such as Isolation Forest (53.20%) and LOF (60.40%). The success of the DNN is attributed to its ability to model complex, nonlinear patterns in voltammetry data. Key factors include the use of a sigmoid activation function in the output layer, which maps predictions to a binary classification, and the optimization of hyperparameters such as learning rate and number of hidden layers.

Tables and confusion matrices were used to validate performance consistency. For example, TABLE 1 shows the superiority of DNN in precision (1.00) compared to Isolation Forest (0.65) and Z-Score (0.66). This result indicates the model's robustness in handling both inliers and outliers effectively.

D. Dicussion

While the DNN model demonstrated remarkable efficacy, certain constraints merit further examination. Passive critics, such as dependence on balanced datasets and sensitivity to data normalization, were identified. For instance, variations in preprocessing techniques (e.g., scaling) markedly influenced model predictions. Future research should prioritize the development of adaptive normalization strategies to mitigate these dependencies.

V. Conclusions

This study proposes a novel approach to outlier detection in heavy metal analysis, specifically tailored to voltammetry data and based on a comparison of five distinct methods. The methods that were tested include Isolation Forest, One-Class SVM, Local Outlier Factor (LOF), Z-Score, and Deep Neural Network (DNN)s (DNN). Among these, DNN is introduced as the primary method, given its potential to model complex patterns and improve accuracy in identifying outliers within high-dimensional datasets.

The results of the evaluation demonstrate that the DNN model exhibits superior performance compared to the other models, achieving an accuracy of 99.87%. This performance is complemented by an F1-score of 1.00, precision of 1.00, and recall of 1.00, making it the most reliable method for detecting outliers in voltammetry data. This makes it the most reliable method for detecting outliers in this context. In comparison, alternative methods, including One-Class SVM, LOF, and Z-Score, exhibit comparatively moderate performance, with One-Class SVM achieving an accuracy of 62%, LOF 60.40%, and Z-Score 59.73%. Isolation Forest exhibited the lowest performance among the tested models, with an accuracy of 53.20%. This comparison highlights the robustness of DNN in handling complex voltammetry data for heavy metal detection, whereas simpler models demonstrate limitations in accuracy.

In future work, further improvements could be made to optimise methods such as Z-Score and LOF in order to enhance their accuracy in detecting subtle or less frequent outlier patterns. Additionally, continued evaluation with a confusion matrix could provide insights into areas where the DNN model can be further refined, particularly in improving detection across diverse outlier categories.

Acknowledgment

This research was funded by Institut Teknologi Sepuluh Nopember (ITS) under Penelitian Kolaborasi Pusat and under the Project Scheme of the Publication Writing and Intellectual Property Rights (IPR) Incentive (Penulisan Publikasi dan Hak Kekayaan Intelektual/PPHKI) Program 2024

References

[1] Purbawa, D. P., Sarno, R., Ardani, M. S. H., Sabilla, S. I., Sungkono, K. R., Fatichah, C.,and Bakhtiar, A. (2022). Adaptive filter for detection outlier data on electronic nose signal. *Sensing and Bio-Sensing Research*, *36*, 100492.D. M. Hawkins, Identification Of Outliers, vol. 11. London, U.K.: Chapman & Hall, 1980.

[2] Sihombing, P. R., Suryadiningrat, S., Sunarjo, D. A., & Yuda, Y. P. A. C. (2022). Identifikasi Data Outlier (Pencilan) Dan Kenormalan Data Pada Data Univariat Serta Alternatif Penyelesaiannya. *Jurnal Ekonomi Dan Statistik Indonesia*, *2*(3), 307-316.

[3] Smiti, A. (2020). A critical overview of outlier detection methods. *Computer Science Review*, *38*, 100306.

[4] Ardani, M. S. H., Sarno, R., Malikhah, M., Purbawa, D. P., Sabilla, S. I., Sungkono, K. R., and Susilo, R. I. (2022). A new approach to signal filtering method using K-means clustering and distance-based Kalman filtering. *Sensing and Bio-Sensing Research*, *38*, 100539.

[5] Amri, T. C., Sarno, R., Abdillah, R., Haq, F. A., Septiyanto, A. F., and Sunaryono, D. (2023, February). Filter Validation for Detecting Outliers of Photoplethysmograph Data. In *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)* (pp. 012-017). IEEE.

[6] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation forest,” in Proc. 8th IEEE Int. Conf. Data Mining. IEEE, 2008, pp. 413–422.

[7] Hejazi M, Singh YP. One-class support vector machines approach to anomaly detection. *Appl Artif Intell*. 2013;27(5):351-366. doi:10.1080/08839514.2013.785791.

[8] Xu, H. Liu, L. Duan and W. Wu, "An Improved LOF Outlier Detection Algorithm," *2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, Dalian, China, 2021, pp. 113-117, doi: 10.1109/ICAICA52286.2021.9498181.

[9] AbouElleef, E. M., Gomaa, E. A., Salem, M. A., Soud, M. R., & El-Ghobashy, M. A. (2024). Cyclic voltammetry analysis of mercuric chloride redox reactions with orange G dye. *Journal of Molecular Liquids*, *414*, 126171.

[10] P.A. Kempler, S.W. Boettcher, S. Ardo, Reinvigorating electrochemistry education, Iscience 24 (5) (2021).

[11] Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. "Isolation-based anomaly detection." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 6.1 (2012): 1-39.

[12] Amer, M., Goldstein, M., & Abdennadher, S. (2013, August). Enhancing one-class support vector machines for unsupervised anomaly detection. In *Proceedings of the ACM SIGKDD workshop on outlier detection and description* (pp. 8-15).

[13] Belyadi, H., & Haghighat, A. (2021). *Machine learning guide for oil and gas using Python: A step-by-step breakdown with data, algorithms, codes, and applications*. Gulf Professional Publishing.

[14] Ramkumar, S., Vijayalakshmi, S., Kanagarajan, P., Patil, R. A. J. K. U. M. A. R., & Lokeshmaran, A. (2018). Z-score and CIAF-a descriptive measure to determine prevalence of under-nutrition in rural school children, Puducherry, India. *Journal of Clinical and Diagnostic Research*, *12*(5), LC24-LC27.

[15] Subasi, A. (2020). *Practical machine learning for data analysis using python*. Academic Press.

[16] Ardani, M., Sarno, R., Khosasih, M. M., Malikhah, M., Purbawa, D. P., Fatichah, C., ... & Sungkono, K. R. (2022). Electronic Nose Signals for Analysing Similarity of Male and Female Axillary Odour to Food Material Aroma. *International Journal of Intelligent Engineering & Systems*, *15*(5).

[17] Sarno, R., Inoue, S., Ardani, M. S. H., Purbawa, D. P., Sabilla, S. I., Sungkono, K. R., ... & Hernaningsih, Y. (2022). Detection of infectious respiratory disease through sweat from axillary using an E-nose with stacked Deep Neural Network (DNN). *IEEE Access*, *10*, 51285-51298.