Biointerface Research in Applied Chemistry
Open-Access Journal (ISSN: 2069-5837)

Article
Volume 16, Issue 1, 2026, 33
https://doi.org/10.33263/BRIAC161.033

Computational Intelligence and K-Nearest Neighbour-Based Approach to Predict Zn(ii) Ions Biosorption from Blood Plasma-Protein; Central Composite Design and Machine Learning Strategies

Abuchi Elebo [1,*], Danzarami Danlami [1,2], Idris Asifau [1,3], Precious Peter Didam [1], Abatyough Terungwa Michael [4], Jibunor Udoka Victor [5], Alex Kynesieh Waitieh-Kabehl II [6], Usman Olayinka Odunola [7], Thomas Aondofa Nyijime [8], Ighodaro Stanley [9]

[1] Department of Chemistry, Ahmadu Bello University, Zaria, Nigeria
[2] Department of Science Laboratory Technology, National Institute of Leather Science and Technology, Nigeria
[3] Federal Ministry of Innovation, Science and Technology, Abuja, Nigeria
[4] Chemical Sciences Department, Bingham University, Karu, Abuja, Nigeria
[5] Department of Chemistry, Federal University, Otueke, Bayelsa State, Nigeria
[6] Department of Anatomy, Faculty of Medical Sciences, Ahmadu Bello University, Zaria, Nigeria
[7] Department of Microbiology, Ahmadu Bello University, Zaria, Nigeria
[8] Department of Chemistry, Joseph Sarwuan Tarka University, Makurdi, Nigeria
[9] Department of Anaesthesia, Faculty of Medical Sciences, Ahmadu Bello University, Zaria, Nigeria
* Correspondence: abuchielebo@yahoo.com;

Received: 3.08.2025; Accepted: 14.11.2026; Published: 15.02.2026

Abstract: Central composite design (CCD) and machine learning (ML) strategies through the aid of response surface methodology (RSM) and k-nearest neighbor (KNN) were employed to forecast the *in-vitro* decontamination of Zn(ii) from blood plasma-protein (BPP) using a decontaminating agent, *Opuntia fragalis* leaf (OFL). The blood plasma-protein (BBP) was characterised by the presence of Zn(ii). After that, spiking with known concentrations of Zn(ii) ions was performed to ensure optimal decontamination efficacy of the biosorbent (OFL). This study employed three responses and generated models to capture the simultaneous interactive effects of the independent process factors as they influence the responses based on the residual Zn(ii) ions concentration in the BPP (Q_1), the concentration of Zn(ii) ion decontaminated by OFL biosorbent-removal efficiency (Q_2), and OFL biosorbent recovery efficacy-desorption (Q_3). The functionality of the KNN models was contrasted with RSM models using the full dataset (KNN1) and partitioned data (KNN2) criteria, with the correlation coefficient (R2) and root mean square error (RMSE) as metrics. Parameter tuning was performed to optimize the performance of the developed KNN models. It was found to be significantly influenced by the nearest neighbour's k-parameter, attributed to the disparity in the two approaches. The KNN1 model showcased better performances characterised by higher R^2 = 0.8190 - 0.9985 and lower RMSE = 0.1052 - 2.1291 against the RSM model of R^2 = 0.7418-0.9564 and RMSE = 0.9105 - 1.250. As per the KNN2 models, although the performance was lower, the decontamination efficiency was higher than that of the RSM models.

Keywords: Algorithm; artificial intelligence; biosorption; blood plasma; optimisation; machine learning.

1. Introduction

Hazardous heavy metals are biochemical elements that accumulate in ecosystems, threatening the environment, waterways, and all living organisms, including plants and animals [1,2]. These toxic substances are released into the atmosphere through natural and human-induced activities, such as burning fossil fuels [3,4]. Once ingested, heavy metals can travel throughout the body, accumulate in cells and tissues, bind to proteins and nucleic acids, and cause significant cellular damage [5]. Toxic metals have far-reaching consequences for human health, causing widespread damage to the central nervous system, lungs, kidneys, and other vital organs [6]. Prolonged exposure can lead to neurological disorders, such as Parkinson's and Alzheimer's diseases, as well as physical and muscular degeneration [7,8]. Furthermore, long-term contact with certain heavy metals can alter nucleic acids, induce mutations, and disrupt hormone function, ultimately increasing the risk of cancer [9]. Exposure to toxic metals can also impair haemoglobin production, gastrointestinal health, and reproductive function, leading to severe nervous system damage [10].

However, conventional methods for removing metals from water include flotation, electrolysis, precipitation, ion exchange, ultrafiltration, and adsorption [11]. However, traditional adsorbents such as activated carbon are often costly, operationally challenging, and limited in their detoxification efficiency, making them less accessible, especially in developing countries [12]. In contrast, abundant and environmentally friendly biological materials, such as agricultural waste, can be repurposed as low-cost adsorbents [13]. Biological approaches, including biosorption, mycoremediation, and phytoremediation, offer a viable and effective alternative to physical and chemical methods for metal detoxification from industrial and agricultural waste [3]. In contrast, biosorption refers to the process by which metal ions are adsorbed from aqueous solutions onto biological materials [14]. This occurs when substances on the biosorbent surface are trapped by biological materials such as algae, bacteria, and fungi [15]. Over the past two decades, research has identified various biomass materials with detoxifying properties, characterised by functional groups such as carboxyl, phosphate, and amino groups [16]. Commonly used biomasses include waste products from microbial production, sugar, crab shells, nut shells, and plant waste [6,17]. Compared to chemical methods, biosorption is a cost-effective and efficient approach to mitigating environmental pollution [18]. One promising biomaterial is *Opuntia fragilis*, a type of cactus with nutritional and medicinal properties, making it a suitable candidate for detoxification studies [19].

Furthermore, the human blood system is comprised of three primary components: red blood cells, white blood cells, and blood plasma. Notably, blood plasma, which accounts for 60 % of the total blood volume, serves as a transport medium for toxic substances [20]. Its composition is largely water-based, supplemented with proteins, hormones, minerals, and other substances. A key function of blood plasma is the conveyance of waste products. However, when non-biodegradable, toxic metals are introduced into the body and transported through the blood plasma, they can accumulate and pose significant health risks. Fortunately, biomass offers a potential solution, as it can be employed to adsorb these toxic metals [9,20].

In recent times, artificial intelligence (AI) modelling has emerged as a powerful tool across various sectors, outperforming response surface methodology (RSM) predictions [21,22]. AI's exceptional learning capabilities, fueled by its ability to analyse historical data, enable it to deliver accurate and reliable results. Unlike RSM, AI can utilise specialised mathematical functions and fitting, making it a more robust approach [23]. Moreover, AI excels

in extrapolating target responses with high accuracy, revolutionising data mining and analysis across all fields. Machine learning, a subset of AI, plays a crucial role in classification, providing valuable insights into population groups and driving data-driven decision-making [24]. Similarly, the use of categorisation techniques offers a significant advantage in understanding the underlying rules governing different data types. Various data mining techniques have been employed, including support vector machines (SVM) [25], artificial neural networks (ANN) [26], ensemble classifiers [27], k-nearest neighbour (KNN) [28], and Bayesian classifiers [26]. Notably, the KNN approach stands out for its exceptional performance, robustness to local noise, low sensitivity, and simplicity. These features make KNN algorithms highly effective in tackling complex classification tasks in real-world scenarios, providing accurate and generalised results [29]. The KNN technique has expanded its applications across various fields, including social science [28], engineering, and medicine [29], with notable contributions in areas such as breast cancer detection [30], power distribution monitoring [28], protein inhibitor classification [31], and environmental monitoring [32]. Given AI's significant impact on multiple sectors, its role in removing toxic metals from human blood plasma has garnered substantial attention [30]. Interestingly, despite KNN's potential for modelling biosorption processes, especially with limited data, there is a notable knowledge gap in comparing RSM and KNN for predicting heavy metal detoxification from blood plasma using untreated green plants, such as *Opuntia fragilis* leaves.

This study seeks to develop AI-powered predictive models using KNN algorithms to forecast the removal of Zn(ii) ions from contaminated blood plasma proteins (BPP) using unmodified *Opuntia fragilis* leaves as an eco-friendly biosorbent. Additionally, it aims to investigate the biochemical mechanisms underlying Zn(ii) ion toxicity, including protein and enzyme binding, activity alteration, and damage. Notably, this research introduces a novel comparison between KNN and RSM models, evaluating their performance and desorption models using full and partitioned datasets, with assessments based on correlation coefficients (R^2) and root mean square errors (RMSE).

2. Materials and Methods

2.1. Experimental design.

The response surface methodology (RSM) was designed to investigate the influence of biosorbent dose (0.5 - 1.0 g), initial concentration (10-50 mg/L), pH (2-11), and agitation time (10-120 mins) on Zn(ii) ion decontamination effectiveness at physiological temperature (37°C). As a consequence, the required thirty (30) batch experimental trials were carried out in triplicate. The values of the relevant experimental variables are presented in Table 1. A known quantity of OFL biomass was put into a solution containing BPP and a spiking concentration of Zn(ii) ions, and the pH was adjusted using HCl, which was stirred by a mechanical shaker to enable biosorption equilibrium in a suitable period. The mixture was then filtered using filter paper, and the filtrate was tested for the presence of Zn(ii) ions using atomic absorption spectroscopy (AAS). The optimized influence of experimental conditions of Zn(ii) ions decontamination in terms of three (3) responses were, residual Zn(ii) ions concentration in the BPP, the concentration of Zn(ii) ions decontaminated into OFL (removal efficiency), and OFL biosorbent recovery efficacy (desorption), which also are designated as Q_1, Q_2, and Q_3 captured in equations 1, 2, and 3, respectively [32].

$$Q_1 = \frac{(C_o - C_e)V}{M} \quad (1)$$

$$Q_2 = \frac{(C_o - C_e)\times 100}{C_o} \quad (2)$$

$$Q_3 = \frac{100(C_o - C_e)}{C_o} \quad (3)$$

where C_o is the initial Zn(ii) ions concentration, C_e is the final Zn(ii) ions concentration after decontamination, V is the volume of BPP, and M is the mass of the decontaminating agent.

Table 1. Process factors design template.

Process factors	Symbol/unit	Levels of factors				
		-α	**-1**	**0**	**+1**	**+α**
Biosorbent dose	A/(g)	0.10	0.30	0.60	0.80	1.00
Initial concentration	B/(mg/L)	10.00	32.50	55.00	77.50	100.00
pH	C	2.00	4.25	6.50	8.75	11.00
Agitation time	D/(min.)	10.00	37.50	65.00	92.50	120.00

2.2. Desorption inquiry of Zn(ii) ions.

Desorption studies were conducted using HCl as the desorbing agent to assess the reusability of OFL biomass. Fresh OFL (0.5 g) was accurately added to 50 mL of BPP containing 10 mg/L of Zn(ii) ions solution and shaken for 1 hour. The Zn(ii) ions adsorbed onto biomass were then separated using centrifugation. The Zn(ii) ions loaded onto OFL biomass were gently rinsed with water to remove any unadsorbed Zn(ii) ions before being dried. The desorption process was carried out by altering the process parameters shown in Table 2. After shaking the mixture for a set amount of time, the concentration of desorbed Zn(ii) ions was measured by atomic absorption spectroscopy (AAS). Finally, the desorption percentage of zinc ions was calculated as given in Equation 3 [33].

Table 2. CCD experimental operational condition design for Q_1, Q_2, and Q_3.

Exp run	A	B	C	D	Experimental			RSM			KNN		
					Q_1 (mg/L)	**Q_2 (%)**	**Q_3 (%)**	**Q_1 (mg/L)**	**Q_2 (%)**	**Q_3 (%)**	**Q_1 (mg/L)**	**Q_2 (%)**	**Q_3 (%)**
1	-α	0	0	0	1.23	97.76	96.93	1.87	96.76	97.49	1.42	96.93	97.76
2	-1	+1	-1	-1	1.12	98.55	99.70	0.97	99.03	100.05	0.82	99.69	98.55
3	1	+1	+1	-1	3.12	95.97	96.50	3.21	96.55	97.10	3.03	96.50	95.97
4	1	-1	-1	-1	1.74	94.65	93.47	1.94	94.15	93.96	1.79	93.47	94.64
5	0	-α	0	0	0.21	97.90	85.19	0.22	96.76	84.83	0.44	85.19	97.83
6	-α	0	0	0	1.21	97.80	98.79	1.57	97.40	98.11	1.41	98.79	97.72
7	-1	+1	-1	+1	1.61	97.92	98.26	1.59	98.20	98.07	1.41	98.26	97.92
8	0	0	-1	0	3.12	94.33	98.77	1.87	96.76	97.49	2.92	98.77	94.33
9	-1	+1	+1	-1	2.42	96.88	98.22	2.53	97.38	98.01	2.02	98.22	96.88
10	-1	-1	+1	-1	1.83	94.37	94.65	1.90	93.87	95.21	1.79	94.66	94.37
11	+1	0	0	0	2.62	95.24	96.54	1.87	96.76	97.49	2.67	96.54	95.24
12	+α	0	0	0	2.01	96.35	98.36	2.17	96.12	96.88	2.07	98.36	96.35
13	+1	+1	0	+1	2.75	96.45	95.96	2.69	95.87	96.17	2.84	95.97	96.45
14	-1	+1	+1	+1	2.93	96.22	96.37	2.76	96.32	96.49	3.03	96.37	96.23
15	-1	-1	+1	+1	1.76	94.58	96.84	1.95	94.26	96.80	1.82	96.85	94.58
16	0	0	0	-α	2.71	95.07	97.44	2.57	95.54	97.49	2.67	97.44	95.07
17	0	0	-α	0	2.12	96.15	98.39	1.87	95.99	98.31	2.07	98.39	96.14
18	+1	+1	-1	-1	2.88	96.28	97.39	2.95	96.69	97.36	2.38	97.38	96.28
19	+1	-1	+1	-1	1.87	94.25	96.83	1.73	94.49	96.67	1.82	96.83	94.25
20	+1	+1	-1	+1	0.14	99.82	98.95	0.21	99.26	99.54	0.33	98.95	99.82
21	-1	-1	-1	+1	1.54	95.26	92.54	1.26	95.67	92.76	1.64	92.54	95.26
22	0	0	+α	0	1.96	96.44	96.76	1.87	95.99	96.68	2.45	96.76	96.44
23	0	0	0	+α	1.87	96.6	97.52	2.04	95.54	97.49	2.38	97.52	96.67
24	0	0	0	1	2.23	95.95	96.04	1.87	96.76	97.49	2.68	96.04	95.95
25	0	+α	0	0	2.65	97.35	99.78	2.66	96.76	99.42	1.93	99.78	97.35
26	+1	-1	+1	+1	2.87	91.17	93.73	2.73	92.35	94.10	2.35	93.72	91.17

Exp run	A	B	C	D	Experimental			RSM			KNN		
					Q_1 (mg/L)	Q_2 (%)	Q_3 (%)	Q_1 (mg/L)	Q_2 (%)	Q_3 (%)	Q_1 (mg/L)	Q_2 (%)	Q_3 (%)
27	+1	-1	-1	+1	0.67	97.94	91.67	0.72	98.18	91.85	0.74	91.67	97.94
28	0	0	-1	0	1.61	97.07	98.17	1.87	96.76	97.49	2.02	98.16	97.07
29	-1	-1	-1	+1	0.81	97.51	94.45	0.81	97.95	94.28	1.78	94.45	97.51
30	0	0	-1	0	0.52	99.05	98.42	1.87	96.76	97.49	0.33	98.42	99.05

2.3. Description of RSM tactics.

RSM is referred to as a multivariate strategy because it leverages fewer data points for analysis and can incorporate numerous factors into consideration while displaying synchronous and single-interacting effects of independent variables [33,34]. Moreover, RSM works by fitting a data set to a second-order model (equation 4), which reveals ANOVA as a major method for displaying 3D and contour graphs to identify relationships between dependent and independent variables [33].

$$Y = \beta_o + \sum_{i=1}^{n} bi Xi + (\sum_{i=1}^{n} bii Xi)^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} bij XiXj + € \quad (4)$$

Where Y denotes the model's prediction, β_o is the model constant, bi is the linear coefficient, bij is the 2-way interaction coefficient, bii is the quadratic coefficient, Xi and Xj are independent variables, n is the number of independent variables, and € is the error term.

2.4. Description of K-nearest neighbour tactics.

K-NN is one of the most user-friendly learning algorithms that is readily understandable. It is acknowledged as a decently priced and effective algorithm for deployment and execution [26]. Furthermore, it is classified as a non-parametric lazy algorithm since it makes no assumptions about the functional form Y = F(X) [24,26]. This method is capable of handling both classification and regression issues. In terms of classification technique, dataset segmentation is performed by considering the k-nearest neighbors with the highest frequency among the training points. Similarly, with the regression strategy, the mean values of k are used. KNN functionality is based on the correlation of supplied training data with the test case [24]. Nevertheless, the training data quantifies the n attributes, and KNN then searches for the training data pattern that appears to be closest to the test data. As a result, the training data are labelled as the test data's k-nearest neighbour. The Euclidean distance is a measure of the proximity between the training and test data [26]. Equation 5 is used to calculate Euclidean distance.

$$dist(X_1X_2) = \sqrt{\sum_{i=1}^{n} (X_{1i} - X_{2i})^2} \quad (5)$$

Where X_1 and X_2 are the two points, $X_1 = (x_{11}, x_{12}, ..., x_{1n})$ and $X_2 = (x_{21}, x_{22}, ..., x_{2n})$.

Sequel to the distinct advantages captured previously, KNN exhibits the capacity to work with complex classes, achieve better accuracy, and ensure consistent implementation of KNN techniques, due to its straightforward simplicity, which has earned it a reputation and popularity in machine learning [35].

2.5. Appraisal of the design models' performance.

The RSM and KNN models' effectiveness was evaluated using the coefficient of regression and root mean square error (RMSE) from equations 6 and 7, respectively.

$$R^2 = \frac{(Y_i - Y_i^k)2}{(Y_i - Y^c)2} = 1 - \frac{SSR}{SST} \quad (6)$$

$$RMSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i^k)^2 = \frac{1}{n}\sum_{i=1}^{n} SST \quad (7)$$

Where Y_i, Y_i^k and Y^c represent the experimental, predicted, and mean responses, respectively. SST denotes the total sum of squares, and SSR is the estimated error sum of squares.

2.6. WEKA simulation niche.

For the computational intelligence (CI) experiment, the Waikato Environment Knowledge Analysis (WEKA) modelling environment was used to design KNN models in two steps. First designated as KNN1, KNN was used in the same manner as RSM modelling. This encompasses the full data set generated by the laboratory work on Zn(ii) ion removal from BPP, detailed in section 2.1, and it was used for both WEKA niche training and validation. Similarly, the second modelling in the WEKA environment is designated KNN2, and the data set was subdivided into 75 % for training the KNN model and 25 % for validating the trained KNN model. Taking into cognisance the smaller amount of data set employed for the experiment, KNN could only be used since RSM lacks the capability and property of data splitting (modelling). This lends to KNN modelling's benefits over RSM. According to RSM, the data sets were coded before modelling, whereas data normalisation was performed in the KNN models (KNN1 and KNN2). This prompt was introduced to enhance calculations and the model's prediction capabilities [31].

2.7. Optimisation tactics.

In KNN, optimisation involves tuning the k parameter, which specifies the number of neighbours used to predict the best test instances. Moreover, for each k value used, the experiment was repeated to attain the best performance for the KNN models. Although computing KNN is relatively expensive, especially when the dataset is large, it is necessary to compute the distance between the new data and each data point in the training set [31]. The dataset for this research is limited, and such a challenge does not suffice. Additionally, KNN is vulnerable to dimensionality concerns owing to the computation of all dataset characteristics as a function of priority similarity in approaches [25]. However, the dataset attributes in this study are not large, so the challenge does not arise in the suggested model [32].

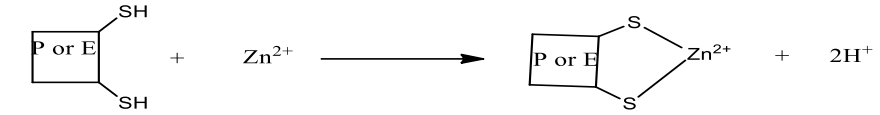
3. Results and Discussion

3.1. FTIR approach to the biochemical mechanistic interaction of OFL and Zn(ii) ions.

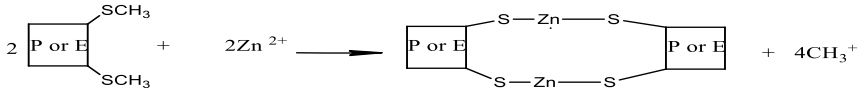
When Zn(ii) ions enter the body through water and food, they are acidified by the stomach's (acidic medium) hydrochloric acid, resulting in a variety of oxidation states (Zn^{2+} or Zn^{+}) that can easily chelate with biological molecules such as proteins and enzymes found in

blood plasma, resulting in the formation of strong and stable biotoxic compounds. Figure 1a and 1b demonstrate the functional groups found in OFL biomass as well as the molecules responsible for Zn(ii) ions cleanup following contact with the BPP. The band frequencies 1340.57-1440.56 cm^{-1}, 1637.62-1650.12 cm^{-1}, and 2550.21-262042 cm^{-1} correspond to symmetric R-N-O stretch, R-NH bend, R-C=C- stretch, and extremely weak R-SH bonds.

Moreover, heavy metals have long been known to bind to the R-SH group, leading to the formation of SCH3 on methionine and SH on cysteine. *In vitro* experiments demonstrated that Zn(ii) ions inhibit the synthesis of SH-transferases such as glutathione reductase, thioreductase, and the pore active sites are responsible for chelating cysteine residues [36]. The interactions between Zn(ii) ions and proteins are represented by Equations 8 and 9. The replacement of hydrogen in the R-SH and SCH_3 groups by oxidised zinc (Zn^{2+}) inhibits protein functions or enzyme activity.



(8)



(9)

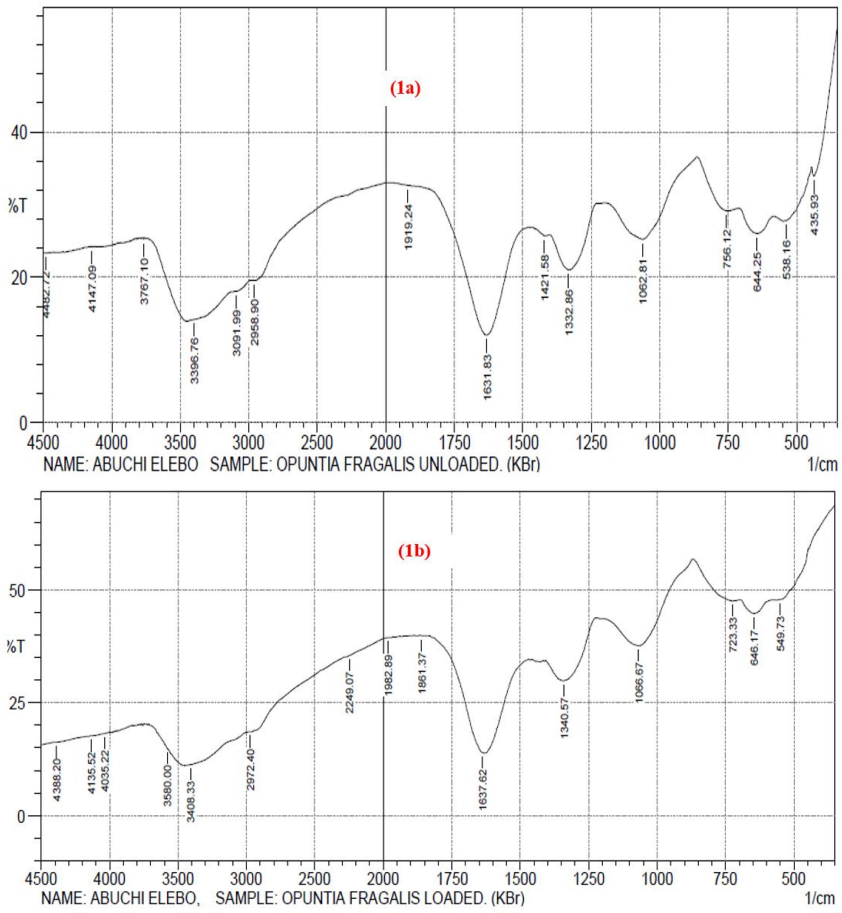


Figure 1. FTIR of **(a)** unloaded OFL biosorbent; **(b)** loaded OFL biosorbent with Zn(ii) ions.

However, a zinc-protein bond can form as a substrate to enzymes, resulting in an enzyme-substrate complex that cannot bind with other substances until it interacts with a substrate that can chelate Zn(ii) ions to its pores and binding sites; otherwise, the substrate's outcome is not produced because the zinc ions are bound to the protein or enzymes [37]. As a result, the Zn(ii) ions are chelated to the tissue (see equations 8 and 9), causing organ

dysfunctions, abnormalities, and damage. Additionally, the addition of a decontaminating agent, such as OFL biomass, which contains several functional groups (Figure 1b), will serve as a chelating agent with active pore sites capable of binding zinc ions.

3.2. Build-up of regression model equations and statistical analysis.

The final empirical regression equations for the Q1, Q2, and Q3 responses are presented in equations 10, 11, and 12, respectively. The positive and negative signs preceding the coefficients represent the target factors' synergistic and antagonistic influence. The existence of a single factor in the term infers a univariate impact, two variables infer a bivariate effect, three variables infer a trivariate effect, and a second-order term of the variable present infers a quadratic effect [25]. Therefore, an analysis of variance (ANOVA) was performed to ensure a thorough examination of personalised interaction as a result of quadratic effects of the process factors influencing Q_1, Q_2, and Q_3 for Zn(ii) ions removal from BPP. Tables 3, 4, and 5 show the mean square and sum of squares of each component, as well as the F-value and p-values, respectively. Significant model terms are those with p-values less than 0.05 [38]. As a reference to Q_1, Q_2, and Q_3 data, see Tables 3, 4, and 5. For Q_1, the significant model terms are B, AD, ACD, A^2C; for Q_2, they are ACD, A^2B, A^2C; and for Q_3, they are A, B, BC, B^2, ABD, ACD, A^2B, A^2D. Additionally, the statistical findings indicated that the models successfully predicted the detoxification of Zn(II) ions from BPP for Q_1, Q_2, and Q_3.

$$Q_1 = +1.87 + 0.1508A + 0.61B - 0.1333D - 0.3025AD + 0.1187BC - 0.1688BD + 0.2288CD - 0.1061B^2 + 0.1089D^2 - 0.2112ABD + 0.3263ACD - 0.367A^2B + 0.5650A^2C \quad (10)$$

$$Q_2 = +96.76 - 0.3199A + 0.4718AD + 0.2454BC - 0.4370CD - 0.1918C^2 + 0.3055D^2 - 0.7414ACD + 1.15A^2B - 1.13A^2C + 0.3624ABCD \quad (11)$$

$$Q_3 = +97.49 - 0.3079A + 3.65B - 0.4089C - 1.07BC + 0.1470BD - 1.34B^2 + 0.5934ABD - 0.4449ACD - 0.3313BCD - 1.95A^2B + 0.5760A^2C - 0.4302A^2D \quad (12)$$

3.3. Interaction factors' effects on Q_1, Q_2, and Q_3.

The experimental data used to forecast and create RSM and KNN models are summarised in Table 2. These lab results represent the average values for each of the 30 samples, which were each tested three times. Compared with RSM predictions, KNN predictions are more in line with experimental or laboratory data, with negligible residual errors for Q_1, Q_2, and Q_3. Figure 2a–c depicts various 3D, cubic, and contour plots to show the correlation between the interaction factors and the relevant responses [28]. The interaction condition modification induced by the biosorption of Zn(ii) ions onto the biosorbent under investigation resulted in responses with varying degrees of impact [31].

3.4. Effects of statistics and models on interaction parameters in Q_1, Q_2, and Q_3.

As it relates to the responses (Q_1, Q_2, and Q_3), analysis of variance (ANOVA) was utilised to determine the significance or influence of the process factors on the biosorption of Zn(ii) ions from BBP, as shown in Tables 3, 4, and 5, respectively. However, independent factors with p-values less than 0.05 are termed significant with high F-values. Hence, it can be established that the lower the p-value, the higher the F-value, which insinuates a higher impact on the responses. The model's p-values of 0.0061, 0.0008, and < 0.0001 appear to be significant

with corresponding F-values of 3.88, 5.46, and 31.06 for Q_1, Q_2, and Q_3, respectively [26,32]. It can be observed that the lack of fit (LOF) for all ANOVA models (Tables 2, 3, and 4) is non-significant, as the p-values are greater than 0.05, indicating that the models adequately fit the biosorption process and can navigate the design space by predicting the experimental data. The quadratic model equations expressed in equations 10, 11, and 12 were used to determine the single, double, and quadratic interaction effects of the process factors on the responses. The negative and positive signs in the RSM equations indicate the antagonistic and synergistic effects of the process factors on the responses, respectively [37].

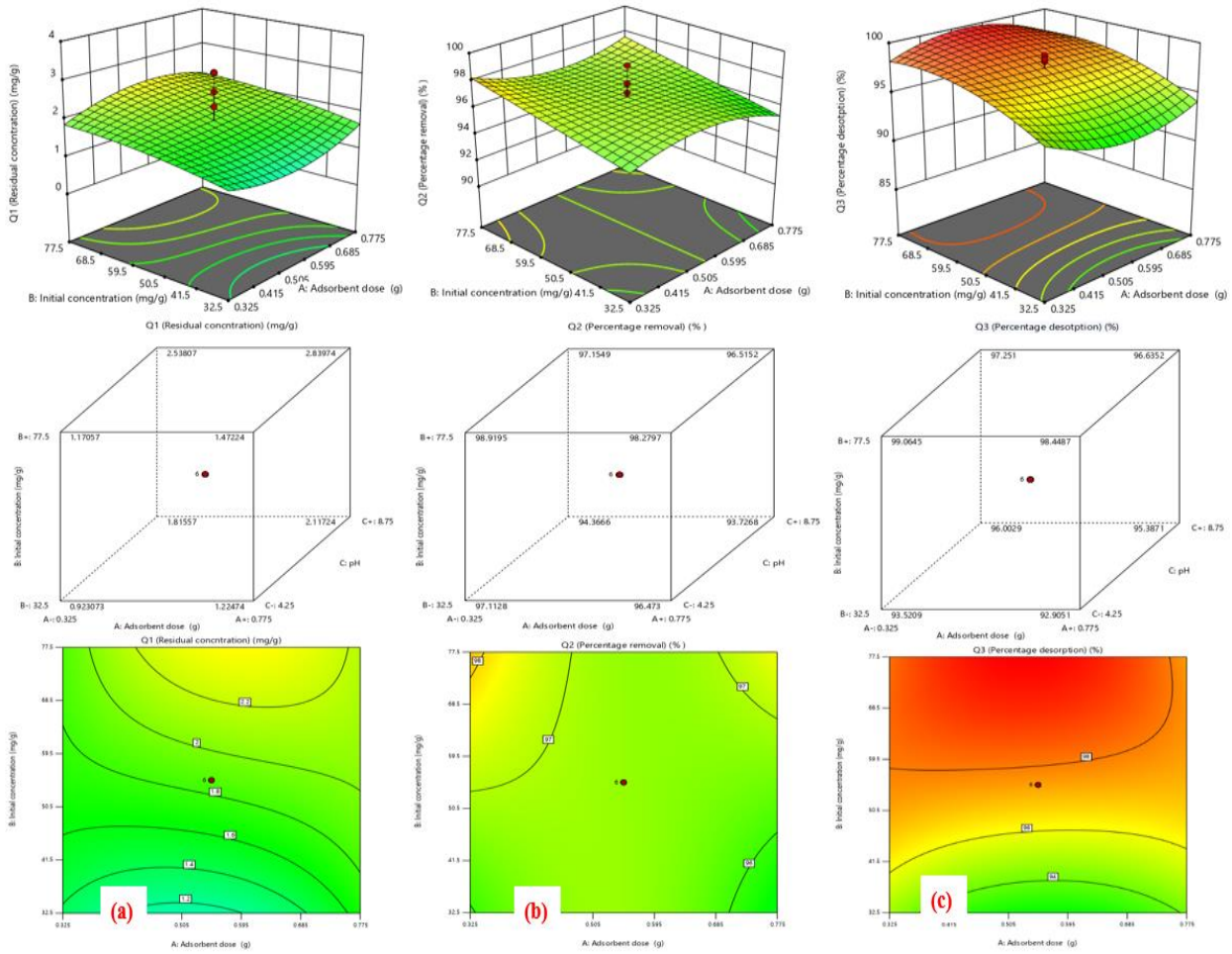


Figure 2. 3D, cube, and contour plots showing combined effects of **(a)** Factor A vs B; **(b)** Factor A vs B vs C; **(c)** Factor A vs B on Q_1, Q_2, and Q_3.

Table 3. ANOVA for the response surface quadratic model for Q_1.

Source	Sum of squares	df	Mean square	F-value	p-value	Verdict
Model	16.14	13	1.24	3.88	0.0061	significant
A-Biosorbent dose	0.5460	1	0.5460	1.71	0.2100	
B-Initial concentration	2.98	1	2.98	9.30	0.0076	
D-Agitation time	0.4267	1	0.4267	1.33	0.2652	
AD	1.46	1	1.46	4.57	0.0482	
BC	0.2256	1	0.2256	0.7049	0.4135	
BD	0.4556	1	0.4556	1.42	0.2502	
CD	0.8372	1	0.8372	2.62	0.1254	
B^2	0.3202	1	0.3202	1.00	0.3321	
D^2	0.3374	1	0.3374	1.05	0.3199	
ABD	0.7140	1	0.7140	2.23	0.1548	
ACD	1.70	1	1.70	5.32	0.0348	
A^2B	0.7203	1	0.7203	2.25	0.1531	
A^2C	5.11	1	5.11	15.96	0.0010	
Residual	5.12	16	0.3201			

Source	Sum of squares	df	Mean square	F-value	p-value	Verdict
Lack of Fit	0.5693	11	0.0518	0.0568	0.9999	not significant
Pure Error	4.55	5	0.9105			
Cor Total	21.26	29				

Table 4. ANOVA for the response surface quadratic model for Q_2.

Source	Sum of squares	df	Mean square	F-value	p-value	Verdict
Model	65.77	10	6.58	5.46	0.0008	significant
A-Biosorbent dose	2.46	1	2.46	2.04	0.1696	
AD	3.56	1	3.56	2.96	0.1018	
BC	0.9636	1	0.9636	0.8000	0.3823	
CD	3.06	1	3.06	2.54	0.1278	
C^2	1.05	1	1.05	0.8690	0.3629	
D^2	2.65	1	2.65	2.20	0.1541	
ACD	8.80	1	8.80	7.30	0.0141	
A^2B	21.11	1	21.11	17.53	0.0005	
A^2C	20.35	1	20.35	16.89	0.0006	
ABCD	2.10	1	2.10	1.74	0.2023	
Residual	22.89	19	1.20			
Lack of Fit	7.84	14	0.5599	0.1860	0.9942	not significant
Pure Error	15.05	5	3.01			
Cor Total	88.65	29				

Table 5. ANOVA for the response surface quadratic model for Q_3.

Source	Sum of squares	df	Mean square	F-value	p-value	Verdict
Model	240.96	12	20.08	31.06	< 0.0001	Significant
A-Biosorbent dose	2.28	1	2.28	3.52	0.0779	
B-Initial concentration	106.51	1	106.51	164.77	< 0.0001	
C-pH	1.34	1	1.34	2.07	0.1685	
BC	18.45	1	18.45	28.54	< 0.0001	
BD	0.3455	1	0.3455	0.5345	0.4747	
B^2	51.94	1	51.94	80.35	< 0.0001	
ABD	5.63	1	5.63	8.72	0.0089	
ACD	3.17	1	3.17	4.90	0.0408	
BCD	1.76	1	1.76	2.72	0.1176	
A^2B	20.30	1	20.30	31.40	< 0.0001	
A^2C	1.77	1	1.77	2.74	0.1164	
A^2D	2.96	1	2.96	4.58	0.0471	
Residual	10.99	17	0.6464			
Lack of Fit	4.73	12	0.3940	0.3146	0.9532	not significant
Pure Error	6.26	5	1.25			
Cor Total	251.94	29				

3.5. Implications of the generated KNN-based model's performance on the k value.

In Figure 3, for the Q_1, Q_2, and Q_3 models, respectively, the effects of KNN parameters on the performance of generated KNN-based models (KNN1 and KNN2) are captured. As shown in Table 5, similar findings were used in other papers, primarily using different data for training and validation. To enable comparison with RSM models, which are often constructed on the latter, this research compares a well-known technique with one that uses the same training data for validation [25,31]. The patterns shown in Figure 3 highlight the utility of the data, and the nature of the experiment performed on the data set significantly impacts the value of the k-parameter. Moreover, Figure 3 combines and compares the best performances of the RSM, KNN1, and KNN2, based on the experimental findings that were confirmed versus the anticipated outcomes, as well as their associated R^2 and RMSE for the Q_1, Q_2, and Q_3 models [23,39].

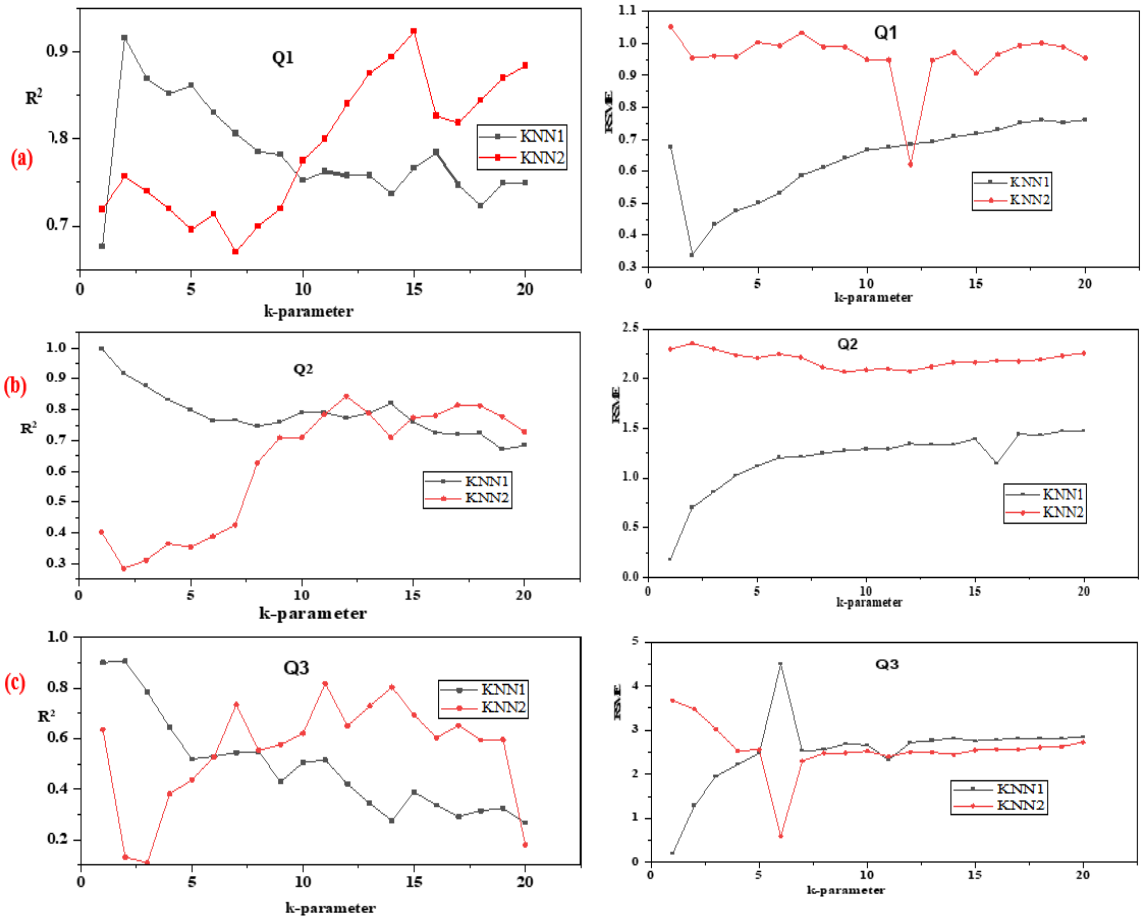


Figure 3. Tuning optimal k-value for KNN1 and KNN2 models for **(a)** Q_1; **(b)** Q_2; **(c)** Q_3.

3.6. Effect of k-value on Q_1, Q_2 and Q_3 KNN based predictive models.

The patterns in Figure 3 highlight how the nature of the experiment conducted on the dataset and the substantial influence of data type on the k-parameter value. Also, KNN2 for Q_1 (Figure 2a and Figure 3c) shows lower model performance than KNN1; this might be due to the strict methodology, which divided the data into training (75 %) and testing (25 %) groups [40]. In this case, the model's performance was assessed by optimising the k parameter, resulting in R2 = 0.8696 and RMSE = 0.9061 at k = 15. By raising the k value, this ideal value was gradually attained; however, any further increases beyond 15 had a detrimental impact on the model's performance [41].

Similarly, the KNN1-developed model performed better than the KNN2 models across the tested k values within the designated range (Figure 2b-c and 3 b- c) for experiments investigating the impact of experimental conditions on the removal and desorption efficiencies of the OFL biosorbent [42]. For both Q_2 and Q_3, the optimal KNN1 models' performance was also arrived at when k = 1, which yielded R^2 = 0.9985; RMSE = 0.1052 and R^2 = 0.9964; RMSE = 0.1130, respectively. Therefore, both models' performances decreased as k-values were further increased from 2 to 20. Nevertheless, based on the same situation, R^2 values for Q_2 ranged from 0.9985 to 0.6713, while RMSE values for Q_2 ranged from 0.1052 to 1.4735, and Q_3 values ranged from 0.113 to 4.498. The prediction is consistent with the findings of [42] and [43], which revealed improved accuracy at k = 13 (RMSE = 0.518) and k = 10 of KNN models identified as the optimal, respectively, which have minimal influence as the k-value was raised throughout a broad range of k-values 14 to 40 (Table 6). The best model

performance for Q_2 and Q_3 for KNN2 was also found at k = 12 (R^2 = 0.8435 and RMSE = 2.0759) and k = 7 (R^2 = 0.7346 and RMSE = 2.3021), respectively.

Table 6. Recent findings on employing k-nearest neighbour in diverse niches of applications.

Application area	Experimented k-value	Data processing	Optimal k-value	Ref.
Identification of protein kinase inhibitors	1–16	Split data	7	[30]
Predicting soil cation exchange capacity	1–50	Split data	10	[39]
Energy generation from biomass	1–4	Split data	3	[28]
Scent classification	3, 5, and 7	Split data	No significant influence	[24]
Zn ions from BPP	1-20	Unsplit data	1	This study
Zn ions from BPP	1-20	Split data	15	This study

According to Figures 3b-c, Q_1's RMSE values were noticeably considerably lower than Q_2's and Q_3's [44]. According to all of the findings for Q_1 and Q_3 in Figures 3a–c and Table 6 for Zn ions removal utilising OFL from BPP, the lowest k-values (especially for k = 1) produced better classification with the KNN1 approach for Q_2 and Q_3, whereas k = 2 with KNN1 for Q_1. In contrast, they showed that greater k-values are required for optimality in the circumstances of the KNN2 trials. This might be explained by differences between the two methodologies, as there are fewer training and testing samples in KNN2 than in KNN1 [44]. Conversely, the impact of noise on data classification is reduced by high k-values, but this reduces the distinctness of class boundaries [37], which affects model performance, as demonstrated by the study's findings. The k-value that is chosen throughout the testing is obviously what determines the KNN location. The importance of choosing a high k-value for KNN classification cannot be overstated, as it provides rich probabilistic information while remaining sensitive to findings in the decision region [35,41].

However, excessively large k-values might have an adverse effect by distorting the estimated location and preventing decision-makers from accounting for additional data sets [5]. The test set is typically used to evaluate classification accuracy by applying various k-values in the experiment, although the results indicate that the k-value is an odd number (Table 6). Moreover, the results estimated from this study are consistent with those of [1,2], who used partition data for training and searching using the KNN method. Because less data were used in this study to split the data into training and validation (as in KNN2), the validation findings are more likely to contain noise [31], which could give the appearance of higher k-values.

Nevertheless, if the search is carried out using a set of data used for both training and validation (as in KNN1), then k = 1 is a suitable choice. It is crucial to note that in the KNN2 scenario, where the k-value was used at the lowest value of 1, the performance of the KNN classifier is improved at higher k-values. In contrast, the KNN1 scenario depicts a different but better performance when the k-value was used at the value of 1. This statement is consistent with the research mentioned by [24,26]. To avoid weakening class boundaries, a large k-value is used to assess a classifier during reduction [25]. This would imply, nevertheless, that the KNN classification performed less well. This explains why there were a few k-values below 13 in the literature, whereas there were fewer reports of k-values above 20 [41].

3.7. RSM models vs k-nearest neighbour-based models: effect of operating parameters on Q_1, Q_2, and Q_3.

In contrast, Figure 4 shows the performance of RSM, KNN1, and KNN2 at their optimal levels for Q1, Q2, and Q3, respectively, represented as experimental vs. expected, R^2, and RMSE. In general, the models (KKN1 and KNN2) demonstrated outstanding prediction

capacity for the produced optimal k-values through parameter adjustment. This supports the feasibility of using the CI model's closest neighbour in the context of the current investigation [41].

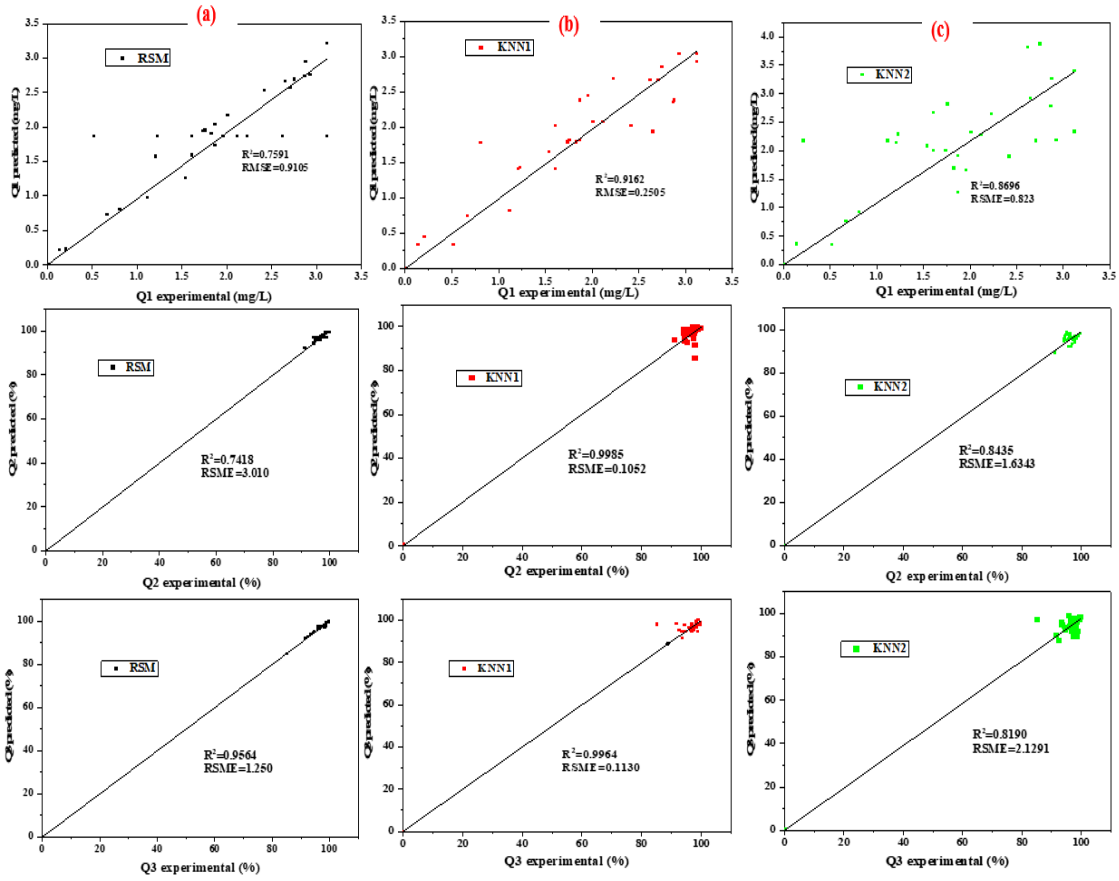


Figure 4. Optimal performance results for RSM, KNN1, and KNN2 models for **(a)** Q_1; **(b)** Q_2; **(c)** Q_3 in terms of experimental vs predicted, R^2, and RMSE.

3.8. RSM models vs KNN1-based models: effect of operating parameters on Q_1, Q_2, and Q_3.

To address the data set, KNN1 was also used for RSM modelling. Figure 4b shows the model's performance and predictability for Q_1, Q_2, and Q_3 [25]. As previously stated, the literature analysis noted that many CI algorithms have been compared with RSM across various application sectors, as shown in Table 7. This is a blend of back-propagation and fuzzy logic [35,41], error back-propagation learning method [42], artificial neural network-particle swarm optimisation (ANN-PSO) [38], and many more. Furthermore, no acknowledged research has described KNN using these two different methods for transposing and processing the dataset, despite the uniqueness of KNN in this field of study and RSM in terms of the transposing data methodology adopted (Table 7) [44].

Furthermore, according to the data shown in Figure 4b, the KNN model performs better than the RSM models shown in Figure 4a. This feature may be related to much lower RMSE values and stronger correlation coefficients. On the other hand, for Q_1, the RSM model had R^2 = 0.7591 and RSM of 0.9105 compared to the KNN1-based model's R^2 = 0.9162 and RMSE of 0.2505. For responses Q_2 and Q_3, there was a difference between the RSM and KNN1-based models (Figure 4b). While for Q_2 and Q_3, KNN1-based models achieved high R2 of 0.9985 and 0.9964, with corresponding RMSE of 0.1052 and 0.1130, respectively, the RSM models for Q2 and Q3 underperformed, with R2 of 0.7418 and 0.9564, and corresponding RMSE of 3.010 and 1.25, respectively. It is conspicuous that KNN1 models learn and fit better into the

experimental data than RSM models [35,44]. This is captured in the corresponding experimentally optimised response plots for the KNN1 model in Figure 4b, which collapse towards the lines of best fit [41].

3.9. RSM models vs KNN2-based models: effect of operating parameters on Q_1, Q_2, and Q_3.

The KNN2 model's predictions are shown in Figure 4c for Q1, Q2, and Q3, as a follow-up to studies that previously cited and compared the performances of RSM and CI models (Table 7). This study utilised k-nearest neighbour experimentation and data partitioning for KNN2 (75 % for training and 25 % for validation). For instance, testing based on this methodology appears to disregard the performance of k-nearest neighbours in KNN1 models, particularly for Q_3, where it performs worse than RSM models. This can be attributed to the KNN2-based models for Q1, Q2, and Q3 having lower R2 values (0.8696-0.8435 and 0.819) and higher RMSE values (0.823-1.6343 and 2.1291) than the other models, respectively [43,45]. It is important to keep in mind that Q_3 (desorption efficiency) is also a useful parameter, given that Q1's performance in removing zinc ions from the BPP has outpaced that of the RSM models. Yet, when the disparity between the experimental methods is taken into account, it is found that KNN2's model performance is better than RSM's. Several studies have compared the effectiveness of RSM and ANN models, as shown in Table 7 [39]. More specifically, Q_3 is a helpful indicator of the desorption effectiveness of adsorbents (Q3a) and the adsorptive removal of zinc ions from BPP using a biosorbent (Q3b), both of which are scantily discussed in the CI literature [28,32]. Moreover, k-nearest neighbour and other CI modelling strategies are excellent suggestions for providing an alternative modelling niche for severe zinc ion detoxification within the BPP experimental niche, functioning as a practical method for reducing polluted biofluids [38,41].

Table 7. Recent findings contrasting CI and RSM models on diverse niches of applications.

Types of CI	Application area	Data processing	Model compared	Best model	Ref.
KNN	Biosorption of Zn ions from BBP	Split	RSM vs KNN2	KNN2	**This study**
KNN	Biosorption of Zn ions from BBP	Unsplit	RSM vs KNN1	KNN1	**This study**
ANN	Adsorption of dyes from water	Split data	RSM vs ANN	ANN	**[21]**
Particle swarm optimisation (ANN-PSO)	Efficient extraction of antioxidants from Vernonia cinerea leaves	Split data	RSM, ANN	ANN	**[46]**
Back propagation (ANN)	Extraction of Oleonolic Acid from Ocimum sanctum	Split data	RSM, ANN, and Fuzzy Logic	ANN	**[39]**
Error back-propagation learning algorithm	Production of bioethanol from pumpkin peel wastes	Split data	RSM ANN	ANN	**[6]**

4. Conclusion

In this study, a central composite design was utilised to build descriptive optimisation and computational intelligence prediction models for the percentage removal and desorption of zinc ions from BPP utilising an ecologically friendly biosorbent from *Opuntia fragalis* leaf. The answers were based on the concentration of residual zinc ions (Q_1), the percentage removal efficiency of zinc ions (Q_2), and the percentage desorption efficiency of zinc ions (Q_3). By using the entire data (KNN1) and split data (KNN2) strategies, the performance and capacity

of models to learn from experimental data were evaluated between RSM and KNN models. The correlation coefficient (R^2) and root mean square error (RMSE) were used to evaluate the models' performances. The created KNN-based model's best results were shown to be highly influenced by the closest neighbour k-parameter, with ideal Q_1 values of k = 2 and 15; Q_2 values of k = 1 and 12; and Q_3 values of k = 1 and 11 for KNN1 and KNN2, respectively. The difference between the two methods' k-values may be related to the KNN2 models' use of less training and validation data than the KNN1 models, compared with the RSM model, which had R^2 = 0.7418-0.9564 and 0.9110-3.010, respectively. The KNN1 model performed better with R^2 = 0.9162-0.9985 and a lower RMSE of 0.1052-0.2505. In addition, the KNN2 models perform better than the RSM models for Q1 and Q2, but worse for Q_3. KNN models have so far demonstrated a clear dominance and advantage over the RSM for predictive modelling strategy.

Author Contributions

Conceptualization, A.B.; methodology, A.B. and I.A.; software, A.T.M.; validation, U.O.O.; formal analysis, I.S.; investigation, D.D.; resources, P.P.D.; data curation, J.U.V.; writing—original draft preparation, A.B.; writing—review and editing, A.B.; visualization, A.K.W.; supervision, A.B.; project administration, T.A.N. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

No new data were created or analysed in this study. Data sharing is not applicable.

Funding

This research received no external funding.

Acknowledgments

I sincerely express my appreciation to my supervisors, who provided valuable input, and to the entire staff of the Department of Chemistry, Ahmadu Bello University, Nigeria, for their valuable advice and suggestions.

Conflicts of Interest

The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Definition
AI	Artificial Intelligence
CI	Computational Intelligence
KNN	K-Nearest Neighbor
RSM	Response Surface Methodology
CCD	Central Composite Design
OFL	*Opuntia fragalis* Leaf
BPP	Blood Plasma Protein
RMSE	Root Mean Square Error
R^2	Correlation Regression
ML	Machine Learning
AI	Artificial Intelligence
SEM	Scanning Electron Microscopy
SVM	Support Vector Machine
ANN	Artificial Neural Network
AAS	Atomic Adsorption Spectroscopy
Q_1	Residual Zn (ii) conc. in BPP
Q_2	Conc. Of Zn (ii) decontaminated into BPP
Q_3	OFL biosorbent recovery efficacy-desorption
PSO	Particle Swarm Optimisation

References

1. Briffa, J.; Sinagra, E.; Blundell, R. Heavy metal pollution in the environment and their toxicological effects on humans. *Heliyon* **2020**, *6*, e04691, https://doi.org/10.1016/j.heliyon.2020.e04691.
2. Yadav, H.; Kumar, R.; Sankhla, M.S. Residues of pesticides and heavy metals in crops resulting in toxic effects on living organisms. *J. Seybold Rep* **2020**, *1533*, 9211. https://doi.org/10.13140/RG.2.2.24806.65609.
3. Margarida Alves, M.; González Beça, C.G.; de Carvalho, R.G.; Castanheira, J.M.; Sol Pereira, M.C.; Vasconcelos, L.A.T. Chromium removal in tannery wastewaters "polishing" by *Pinus sylvestris* bark. *Water Res.* **1993**, *27*, 1333-1338, https://doi.org/10.1016/0043-1354(93)90220-C.
4. Siddiqua, A.; Hahladakis, J.N.; Al-Attiya, W.A.K.A. An overview of the environmental pollution and health effects associated with waste landfilling and open dumping. *Environ. Sci. Pollut. Res.* **2022**, *29*, 58514-58536, https://doi.org/10.1007/s11356-022-21578-z.
5. Zhang, H.; Yuan, X.; Xiong, T.; Wang, H.; Jiang, L. Bioremediation of co-contaminated soil with heavy metals and pesticides: Influence factors, mechanisms and evaluation methods. *Chem. Eng. J.* **2020**, *398*, 125657, https://doi.org/10.1016/j.cej.2020.125657.
6. Volesky, B. Sorption and Biosorption. BV-Sorbex, Inc., St. Lambert (Montreal). *Quebec, Canada* **2003**. http://doi.org/10.4236/jwarp.2012.44023.
7. Karri, R.R.; Ravindran, G.; Dehghani, M.H. Chapter 1 - Wastewater—Sources, Toxicity, and Their Consequences to Human Health. In Soft Computing Techniques in Solid Waste and Wastewater Management; Karri, R.R., Ravindran, G., Dehghani, M.H., Eds.; Elsevier: **2021**; pp. 3-33, https://doi.org/10.1016/B978-0-12-824463-0.00001-X.
8. Shang, Z.; Zhang, L.; Zhao, X.; Liu, S.; Li, D. Removal of Pb(II), Cd(II) and Hg(II) from aqueous solution by mercapto-modified coal gangue. *J. Environ. Manag.* **2019**, *231*, 391-396, https://doi.org/10.1016/j.jenvman.2018.10.072.
9. Calaf, G.M.; Ponce-Cusi, R.; Aguayo, F.; Muñoz, J.P.; Bleak, T.C. Endocrine disruptors from the environment affecting breast cancer. *Oncol Lett.* **2020**, *20*, 19-32, https://doi.org/10.3892/ol.2020.11566.
10. Muhammad, A.; Shabbir, H.; Kashif, J.; Abdul Rehman, K.; Muhammad, S. The Sources, Toxicity, Determination of Heavy Metals and Their Removal Techniques from Drinking Water. *World J. Appl. Chem.* **2020**, *5*, 34-40, https://doi.org/10.11648/j.wjac.20200502.14.

11. Izah, S.C.; Chakrabarty, N.; Srivastav, A.L. A Review on Heavy Metal Concentration in Potable Water Sources in Nigeria: Human Health Effects and Mitigating Measures. *Expo. Health* **2016**, *8*, 285-304, https://doi.org/10.1007/s12403-016-0195-9.
12. Crini, G.; Lichtfouse, E.; Wilson, L.D.; Morin-Crini, N. Conventional and non-conventional adsorbents for wastewater treatment. *Environ. Chem. Lett.* **2019**, *17*, 195-213, https://doi.org/10.1007/s10311-018-0786-8.
13. Topare, N.S.; Wadgaonkar, V.S. A review on application of low-cost adsorbents for heavy metals removal from wastewater. *Mater. Today Proc.* **2023**, *77*, 8-18, https://doi.org/10.1016/j.matpr.2022.08.450.
14. Castanho, N.R.C.M.; de Oliveira, R.A.; Batista, B.L.; Freire, B.M.; Lange, C.; Lopes, A.M.; Jozala, A.F.; Grotto, D. Comparative Study on Lead and Copper Biosorption Using Three Bioproducts from Edible Mushrooms Residues. *J. Fungi* **2021**, *7*, 441, https://doi.org/10.3390/jof7060441.
15. Pokethitiyook, P.; Poolpak, T. Biosorption of Heavy Metal from Aqueous Solutions. In Phytoremediation: Management of Environmental Contaminants, Volume 3, Ansari, A.A., Gill, S.S., Gill, R., Lanza, G.R., Newman, L., Eds.; Springer International Publishing: Cham, **2016**; pp. 113-141, https://doi.org/10.1007/978-3-319-40148-5_4.
16. Priya, A.K.; Gnanasekaran, L.; Dutta, K.; Rajendran, S.; Balakrishnan, D.; Soto-Moscoso, M. Biosorption of heavy metals by microorganisms: Evaluation of different underlying mechanisms. *Chemosphere* **2022**, *307*, 135957, https://doi.org/10.1016/j.chemosphere.2022.135957.
17. Yildiz, S. Artificial neural network (ANN) approach for modeling Zn(II) adsorption in batch process. *Korean J. Chem. Eng.* **2017**, *34*, 2423-2434, https://doi.org/10.1007/s11814-017-0157-3.
18. Ahluwalia, S.S.; Goyal, D. Microbial and plant-derived biomass for removal of heavy metals from wastewater. *Bioresour. Technol.* **2007**, *98*, 2243-2257, https://doi.org/10.1016/j.biortech.2005.12.006.
19. Isaac, A.A. Overview of Cactus (*Opuntia Ficus-Indica* (L): a Myriad of Alternatives. *Ethno-Med.* **2016**, *10*, 195-205.
20. Guo, W.; Hansson, J.; van der Wijngaart, W. Synthetic Paper Separates Plasma from Whole Blood with Low Protein Loss. *Anal. Chem.* **2020**, *92*, 6194-6199, https://doi.org/10.1021/acs.analchem.0c01474.
21. Mu'azu, N.D. Insight into ANN and RSM models' predictive performance for mechanistic aspects of Cr (VI) uptake by layered double hydroxide nanocomposites from water. *Water* **2022**, *14*, 1644. https://doi.org/10.3390/w14101644.
22. Manzar, M.S.; Benaafi, M.; Costache, R.; Alagha, O.; Mu'azu, N.D.; Zubair, M.; Abdullahi, J.; Abba, S.I. New generation neurocomputing learning coupled with a hybrid neuro-fuzzy model for quantifying water quality index variable: A case study from Saudi Arabia. *Ecological Informatics* **2022**, *70*, 101696. https://doi.org/10.1016/j.ecoinf.2022.101696.
23. Mu'azu, N.D.; Abubakar, I.R.; Blaisi, N.I. Public acceptability of treated wastewater reuse in Saudi Arabia: Implications for water management policy. *Science of the Total Environment* **2020**, *721*, 137659. https://doi.org/10.1016/j.scitotenv.2020.137659.
24. Olatunji, O.O.; Akinlabi, S.; Madushele, N.; Adedeji, P.A. Property-based biomass feedstock grading using *k*-Nearest Neighbour technique. *Energy* **2020**, *190*, 116346, https://doi.org/10.1016/j.energy.2019.116346.
25. Müller, P.; Salminen, K.; Nieminen, V.; Kontunen, A.; Karjalainen, M.; Isokoski, P.; Rantala, J.; Savia, M.; Väliaho, J.; Kallio, P.; Lekkala, J.; Surakka, V. Scent classification by K nearest neighbors using ion-mobility spectrometry measurements. *Expert Syst. Appl.* **2019**, *115*, 593-606, https://doi.org/10.1016/j.eswa.2018.08.042.
26. Abu Alfeilat, H.A.; Hassanat, A.B.A.; Lasassmeh, O.; Tarawneh, A.S.; Alhasanat, M.B.; Eyal Salman, H.S.; Prasath, V.B.S. Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. *Big Data* **2019**, *7*, 221-248, https://doi.org/10.1089/big.2018.0175.
27. Ameer, K.; Chun, B.-S.; Kwon, J.-H. Optimization of supercritical fluid extraction of steviol glycosides and total phenolic content from *Stevia rebaudiana* (Bertoni) leaves using response surface methodology and artificial neural network modeling. *Ind. Crop. Prod.* **2017**, *109*, 672-685, https://doi.org/10.1016/j.indcrop.2017.09.023.
28. Cherif, W. Optimization of K-NN algorithm by clustering and reliability coefficients: application to breast-cancer diagnosis. *Procedia Comput. Sci.* **2018**, *127*, 293-299, https://doi.org/10.1016/j.procs.2018.01.125.
29. Peng, X.; Cai, Y.; Li, Q.; Wang, K. Control rod position reconstruction based on *K*-Nearest Neighbor Method. *Ann. Nucl. Energy* **2017**, *102*, 231-235, https://doi.org/10.1016/j.anucene.2016.12.026.
30. Arian, R.; Hariri, A.; Mehridehnavi, A.; Fassihi, A.; Ghasemi, F. Protein kinase inhibitors' classification using K-Nearest neighbor algorithm. *Comput. Biol. Chem.* **2020**, *86*, 107269, https://doi.org/10.1016/j.compbiolchem.2020.107269.

31. Shen, T.; Kong, W.; Liu, F.; Chen, Z.; Yao, J.; Wang, W.; Peng, J.; Chen, H.; He, Y. Rapid Determination of Cadmium Contamination in Lettuce Using Laser-Induced Breakdown Spectroscopy. *Molecules* **2018**, *23*, 2930, https://doi.org/10.3390/molecules23112930.
32. Jensen, W.A. Response Surface Methodology: Process and Product Optimization Using Designed Experiments 4th edition. *J. Qual. Technol.* **2017**, *49*, 186-188, https://doi.org/10.1080/00224065.2017.11917988.
33. Jarrah, N.; Mu'azu, N.D.; Essa, M.H.; Zubair, M. Response surface modeling and optimization of sludge activated carbon production conditions for phenolic compounds removal from water. *Desalin. Water Treat.* **2017**, *100*, 320-332, https://doi.org/10.5004/dwt.2017.21777.
34. Halder, R.K.; Uddin, M.N.; Uddin, M.A.; Aryal, S.; Khraisat, A. Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *J. Big Data* **2024**, *11*, 113, https://doi.org/10.1186/s40537-024-00973-y.
35. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter* **2009**, *11*, 10–18, https://doi.org/10.1145/1656274.1656278.
36. Chrestensen, C.A.; Starke, D.W.; Mieyal, J.J. Acute Cadmium Exposure Inactivates Thioltransferase (Glutaredoxin), Inhibits Intracellular Reduction of Protein-glutathionyl-mixed Disulfides, and Initiates Apoptosis*. *J. Biol. Chem.* **2000**, *275*, 26556-26565, https://doi.org/10.1074/jbc.M004097200.
37. Naganuma, A.; Miura, N.; Kaneko, S.; Mishina, T.; Hosoya, S.; Miyairi, S.; Furuchi, T.; Kuge, S. GFAT as a target molecule of methylmercury toxicity in Saccharomyces cerevisiae. *FASEB J.* **2000**, *14*, 968-972, https://doi.org/10.1096/fasebj.14.7.968.
38. Ahmad, M.A.; Alrozi, R. Optimization of preparation conditions for mangosteen peel-based activated carbons for the removal of Remazol Brilliant Blue R using response surface methodology. *Chem. Eng. J.* **2010**, *165*, 883-890, https://doi.org/10.1016/j.cej.2010.10.049.
39. Zolfaghari, A.A.; Taghizadeh-Mehrjardi, R.; Moshki, A.R.; Malone, B.P.; Weldeyohannes, A.O.; Sarmadian, F.; Yazdani, M.R. Using the nonparametric *k*-nearest neighbor approach for predicting cation exchange capacity. *Geoderma* **2016**, *265*, 111-119, https://doi.org/10.1016/j.geoderma.2015.11.012.
40. Mailagaha Kumbure, M.; Luukka, P.; Collan, M. A new fuzzy *k*-nearest neighbor classifier based on the Bonferroni mean. *Pattern Recognit. Lett.* **2020**, *140*, 172-178, https://doi.org/10.1016/j.patrec.2020.10.005.
41. Joaquín, D.; Francisco, C.; Salvador, G.; Francisco, H. An Interval Valued K-Nearest Neighbors Classifier. In Proceedings of the Proceedings of the 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology, 2015/06, **2015**; pp. 378-384, https://doi.org/10.2991/ifsa-eusflat-15.2015.55.
42. Joshi, A.; Ashish, M. Analysis of k-nearest neighbor technique for breast cancer disease classification. *Int. J. Recent Sci. Res* **2017**, *8*, 1005-19008. https://dx.doi.org/10.24327/ijrsr.2018.0904.1997.
43. Khajeh, M.; Kaykhaii, M.; Sharafi, A. Application of PSO-artificial neural network and response surface methodology for removal of methylene blue using silver nanoparticles from water samples. *J. Ind. Eng. Chem.* **2013**, *19*, 1624-1630, https://doi.org/10.1016/j.jiec.2013.01.033.
44. Alara, O.R.; Abdurahman, N.H.; Afolabi, H.K.; Olalere, O.A. Efficient extraction of antioxidants from *Vernonia cinerea* leaves: Comparing response surface methodology and artificial neural network. *Beni-Suef Univ. J. Basic Appl. Sci.* **2018**, *7*, 276-285, https://doi.org/10.1016/j.bjbas.2018.03.007.
45. Joshi, S.; Bajpai, S.; Jana, S. Application of ANN and RSM on fluoride removal using chemically activated D. sissoo sawdust. *Environ. Sci. Pollut. Res. Int.* **2020**, *27*, 17717-17729, https://doi.org/10.1007/s11356-020-08153-0.
46. Sadhukhan, B.; Mondal, N.K.; Chattoraj, S. Optimisation using central composite design (CCD) and the desirability function for sorption of methylene blue from aqueous solution onto *Lemna major*. *Karbala Int. J. Mod. Sci.* **2016**, *2*, 145-155, https://doi.org/10.1016/j.kijoms.2016.03.005.

Publisher's Note & Disclaimer

instructions, or products mentioned in the content. Readers are encouraged to independently verify any information before relying on it, and the publisher assumes no responsibility for any consequences arising from the use of materials contained in this publication.