

Machine Learning Methods for Finding the True Owner of a Land in Kumasi (Ghana)

Lebi Jean-Marc Dali

School of Technology, Blue Crest College, Accra, Ghana

Email: lebi.dal@bluecrest.edu.gh

How to cite this paper: Dali, L.J.-M. (2025) Machine Learning Methods for Finding the True Owner of a Land in Kumasi (Ghana). *Advances in Artificial Intelligence and Robotics Research*, 1, 65-74.
<https://doi.org/10.4236/airr.2025.11005>

Received: July 23, 2025

Accepted: August 23, 2025

Published: August 26, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The objective of this work is to determine the true owner of a land—public or private—in the region of Kumasi (Ghana). For this purpose, we applied different machine learning methods to the ZEF (Center for Development Research) dataset on Land plots in Kumasi. Machine learning is a field in artificial intelligence that has been applied to many domains successfully in other continents. The machine learning algorithms considered are Polynomial Regression, Softmax Regression, Naïve Bayes, Bagging, Neural Network, Decision Tree, Random Forest, k-Nearest Neighbor and Support Vector Machines. In addition, we designed a rule-based convolution algorithm from these aforementioned algorithms, which performs excellently on the given dataset. In fact, our proposed model performs better than the cited models on the Kumasi dataset. This research is not trivial given the conflicts generated by land ownership issues in the sub-Saharan Africa region as a whole and in Ghana particularly. Therefore, we hope this work will help buyers be more alert in ascertaining the true ownership of a land before purchasing it.

Keywords

Land Disputes, Land Conflicts, Solution, Machine Learning Algorithms, Land Ownership

1. Introduction

Land disputes and conflicts are a serious issue in Ghana [1] and in West Africa [2] as a whole. People are constantly fighting over land because land is a crucial element of life; no wonder it is called real estate. Indeed, some use the land for shelter; they build houses to live there, while others use the land for agricultural purposes. The latter use the land to grow the food that they will consume along with that which they will sell to generate funds. Yet, another category uses the land to feed their flocks,

which is their daily activity and means of survival. Therefore, the land is a precious resource that everyone wants to acquire for their own purpose. Likewise, the government possesses a lot of land and uses it for its own activities. And how many are shocked when they receive the visit of caterpillars and bulldozers to destroy their beloved houses or farm because the land actually belongs to the government and it wants to use it to build a road, a hospital, a school, or anything else. People are shocked because they always thought the land belonged to them. They have purchased it from a chief, a family member, a friend or a contractor. They were given some papers and now after spending a lot of money not only to acquire but to build their houses, office or livelihood on this land. However, the truth was that the land always belonged to the state and the sellers had no right to sell it in the first place. And these events bring confusion, sorrow, bitterness, conflicts and even death if not handled properly. Some people speak of Land capture [3] by the government, but the land minister of Ghana Hon. Emmanuel completely dismisses those claims, stating the land was always public and had always belonged to the government. Therefore, the buyers were tricked into purchasing state-owned land. As a result, we, as machine learning researchers, cannot stay idle with this issue. Indeed, we have powerful tools we can use to help address this issue.

Henceforth, we decided to use the tools of machine learning in order to discover the true owners of a land (*i.e.*, whether it belongs to the government or is available for purchase). Indeed, machine learning methods have been applied successfully in many other areas including healthcare [4], education [5], finance [6], manufacturing [7], cybersecurity [8] and so on. Then, why not apply it in the land acquisition business so that potential buyers are equipped with modern tools to discover the true owners of a land before investing so much capital in it? We need to remember that there are evil people in this world who prey on the naiveté of people. These make innocent people believe they can acquire a land cheaply when the truth is the land is not even available for sale in the first place and worst the land belongs to the government. So anything we can do to help should be tried. This is the reason for our work.

In this project, we applied different machine learning algorithms to the ZEF dataset on land plots in Kumasi (Ghana) [9]. We even designed our own model, which is a convolution of some learning algorithms. The result is very. We were able to successfully predict (about 90%) the true owners of a land using our algorithm. This is an excellent result given that the Kumasi dataset given by ZEF is very limited in size and attributes. Also, we outperform the other algorithms by a large margin. This means that if we apply our algorithm to a bigger land dataset, we may reach about 100% successful predictions, given that the larger dataset will improve the generalizability and robustness of our algorithm.

2. Literature Review

2.1. Machine Learning Algorithms

In this section, we describe various machine learning algorithms; we focus on

those used in this research. Actually, we will try to cover as many algorithms as possible. First, we need to know what we mean by machine learning. By machine learning, we simply mean learning using the machine, i.e., the computer or more precisely computer algorithms. What are we learning? We are learning some model, some function, a pattern that allows us to make predictions. Normally, we will be presented with a dataset (the bigger the better). And using this dataset, we will apply some mathematical formulas to it to discover the gold or the missing element. This is indeed good news because the objectives of many have always been to exploit the ubiquitous data available, for instance, online (big data), to discover things we didn't know before. This prediction can be in the form of classification, where the output is a discrete value or regression where the output is a continuous value.

In this regard, some algorithms that fall under classification are, for instance, SVM (Support Vector Machine) [10], ANN (Artificial Neural Network) [11], KNN (K-Nearest Neighbor) [12] and some important regression algorithms are for instance Linear Regression [13] logistic and softmax regression algorithms [14]. We also have genetic algorithms [15], reinforcement learning algorithms [16]. We must not forget Ensemble algorithms like stacking, committee machine algorithms and bagging [17].

In ANN, we try to emulate the biological neural network. We will have the input layer the output layer (the prediction) and intermediate (hidden) layers. Each layer consists of units that transform the input into the desired output. As for SVM, it tries to find key data instances that will be used to make the classification. We have a decision boundary and a margin separating both classes. KNN, on the other hand, tries to find data instances close to our target element. And using majority vote on these neighbors, we predict the class of our target data. Regression algorithms use the dataset to fit the parameters of the chosen model. Then, they use the model (with the found parameters) to make the prediction for the new value. The genetic algorithm aims to mimic the biological evolutionary process to predict the target attribute. Reinforcement learning is a trial-and-error process where we learn from our mistakes and repeat the process till we get to the solution. Ensemble learning algorithms try to aggregate the solutions from different algorithms. For example, Random Forest aggregates the solution from different decision trees. In Baggings, we run different bootstrap datasets and keep a score of each run. Then, the solution could be a majority vote.

Machine learning algorithms can also be categorized with respect to whether each data instance has a target value. Then, the algorithms are classified as supervised, semisupervised and unsupervised. In supervised machine learning algorithms, the target value of each data instance is provided. Examples of such algorithms include linear or polynomial regression, artificial neural network, and support vector machine. On the other hand, unsupervised machine learning algorithms work with situations where the target attribute is not provided or is unknown for each entity of the data set. Examples of unsupervised learning algo-

rithms include clustering methods and reinforcement learning techniques. Semi-supervised learning techniques are a trade-off between supervised and unsupervised learning algorithms. With semisupervised problems, some dataset instances have their target attribute provided while others do not. Also, the data set may be itself sparse, i.e., some predictor values will not be provided or even when provided they may include errors. For such problems, first a preprocessing stage is warranted, where we will try to approximate values for the missing data. Techniques such as collaborative filtering (user or item based) can be used to first infer the missing elements [18], then some supervised learning algorithms can be applied.

Besides the algorithms already cited, one could add the Rosenblatt Perceptron algorithm, the Naïve Bayes, the Bayesian classifier, the Ensemble learning like committee machine, Bagging.

2.2. Related Work

In this section, we describe some similar research work where machine learning algorithms were successfully used to detect some hidden features in a given dataset.

In Yaokumah's paper [8], the authors applied machine learning algorithms to the problem of detecting malware over the internet. They used algorithms like (Linear Regression) LR, NB (Naïve Bayes), DT (Decision Trees), KNN (K-Nearest Neighbor), SVM (Support Vector Machine), NN (Neural Network), RF (Random Forest), BG (Bagging) and stacking applied on the Bot-IoT dataset. Surprisingly, their model generates a 100% accuracy rate of success.

Hafsah Habehh and friends [4] describe the application of machine learning algorithms in the health system. Indeed, they discover that machine learning allows for the discovery of irregularities in blood samples, bones and organs using just medical imaging. They also find that machine learning has helped physicians in diagnosing diseases, identifying trends in healthcare and developing disease prediction models.

As for Zara [5], they study the use of machine learning in education. They find out that machine learning can be used to predict students' performance, grade students fairly, organize content effectively and suggest a learning path for students' future careers.

Likewise, Bryan Kelly and colleagues [6] study the impact of machine learning in the financial markets. They discovered that machine learning is paramount in predicting product prices and market trends.

In [7], Vimanyu Chopra discusses the impact of machine learning algorithms in manufacturing processes and industrial production. He discovers that machine learning can indeed help in the manufacturing process especially in predictive maintenance, process optimization, quality control and scheduling of resources.

Also in the domain of land management and classification, machine learning methods were applied successfully. In their work [19], the authors were able to use

deep learning and other machine learning algorithms to effectively and efficiently classify lands. They succeeded despite the multiple challenges with respect to the vegetation, the land cover and hilly areas.

Finally, Junye Wang and colleagues provided an excellent survey of machine learning in modelling land-use and land cover-change (LULCC) in their seminal work [20].

3. Methodology

For this work, we apply machine learning tools to the task of predicting the true ownership of a land in Kumasi (Ghana) using the ZEF dataset. In the first phase of our work, we had to select a dataset on land in Ghana. This was the most difficult because such specific datasets are lacking. However, we discovered the ZEF research dataset on land plots in Kumasi (ZEF dataset). In the second phase of our study, we did some feature selection and feature extraction in order to put the dataset in a good condition so we could apply our machine learning algorithms. Finally, we selected the most prominent ML (Machine Learning) algorithms to conduct our study and designed a rule-based convolution algorithm, which performed better than the other algorithms.

3.1. ZEF Kumasi Dataset

The dataset we used is from ZEF (Zentrum für Entwicklungsforschung), meaning center for Development Research of the University of Bonn in Germany. ZEF conducted various research in Africa especially in Ghana. Therefore, as part of their work they published the given dataset [9]. The dataset is about land plots in Kumasi. It has 465 rows and seven columns or attributes. The dataset is highly sparse. As a result, we needed to do some preprocessing work to identify the features of interest. The preprocessing consisted of identifying and selecting the attributes that were not too sparse and were of interest to our target attribute, *i.e.*, determining whether a land belongs to the government or not. The attributes of the dataset are Submetro, Ownership, Urban zone, #Trees, LandUse, Plot_area, ID, and plot. The submetro shows the submetro of Kumasi where the land exists. Its values consist of “Asawasi”, “Asukwa”, “Bantama”, “Kwadaso”, “Menhyia”, “Nhyiaeso”, “Oforikro”, “Oforikrom”, “Suame”, “Subin” and “Tafo”. Though the submetro is important when considering the price of the land, we judged it as not important in determining the ownership of the land. The ownership attribute tells whether the land belongs to the government, private or mixed, *i.e.*, co-shared between the government and a private entity. These qualitative attributes were turned into quantitative values by assigning them discrete values. Thus, government land was assigned 1 and private 0. We considered that co-shared land belonged to the government; so all mixed attributes were given a value of 1 for public or state land. Urban zone means where the land is located, whether inside the city or in a suburb. A value of “Core” means inside the city and a value of “Periurban” means suburban land. Similar to the “Ownership” attribute, we also assigned discrete

values to this attribute. Henceforth, “Core” received 1 and “Periurban” 2. The “#tree” attribute tells how many trees are present, so it addresses the vegetation of the land. We did not consider “#tree” attribute relevant in determining land ownership, because it is highly sparse and the variance between values is too high. “Landuse” attribute deals with whether the land is already in use and what it is used for. The values go from 1 to 5. “Plot area” gives the size of the land. Finally, the ID and plot number are used for land addressing or localization. Out of these attributes, only the urban zone, #trees, land use and area were used as features to determine the land ownership. The dataset is divided into two parts: one fourth (117 entries) is used as the test set and three fourths (348 entries) is used as the training set.

3.2. Our Convolved Rule-Based Algorithm

Our convoluted rule-based algorithm is like the traditional majority vote [11], but with some subtleties. In the traditional majority vote method, we apply various machine learning algorithms to our dataset and take the majority or the popular vote of the algorithms as the prediction. In a similar fashion, we first apply machine learning algorithms and then from their predictions we compute our own prediction. The algorithms considered are KNN, Logistic Regression, Naïve Baye, Neural Network, Linear Regression and SVM. These algorithms in particular were selected for our project because they performed very well in the work of our colleagues from the University of Ghana Yao Kumah and others [8] on the BOT-IOT dataset. Their algorithm was even able to achieve 100% accuracy. Thus, we hoped that by using the algorithms they used in their work, we would be able to reach a higher accuracy rate. From the training, we realize that Naïve-Bayes is less efficient than Logistic Regression has the same predictions, so we focused just on Logistic regression in our convoluted algorithm. Likewise, only high-performing algorithms were included in our rule-based algorithm (as can be attested in the results section).

From the training set, we realize that whenever SVM predicts 0, KNN predicts 1, Logistic Regression 1, Neural Network/ Random Forest 0 and Linear regression 0, the actual prediction is 0 (private). Also, whenever SVM predicts 0, KNN predicts 1, Logistic Regression 1, Neural Network 0 and Linear regression 1, the actual prediction is 1 (public). Again, whenever SVM predicts 1, KNN predicts 1, Logistic Regression 1, Neural Network 1 and Linear regression 1, the actual prediction is 1 (public). Furthermore, whenever SVM predicts 0, KNN predicts 0, Logistic Regression 1, Neural Network 0 and Linear regression 1, the actual prediction is 0 (private). Finally, whenever SVM predicts 0, KNN predicts 0, Logistic Regression 1, Neural Network 1 and Linear regression 1, the actual prediction is 1 (public).

The algorithm can be stated as:

if ((SVM==0) and (KNN==1) and (Logistic==1) and (Neural Network==0) and (LinearRegression==0)), then

```

MyOwn=0
else if ((SVM==0) and (KNN==1) and (Logistic==1) and (Neural Network==0)
and (LinearRegression==1)), then
    MyOwn=1
else if ((SVM==1) and (KNN==1) and (Logistic==1) and (Neural Network==1)
and (LinearRegression==1)), then
    MyOwn=1
else if ((SVM==0) and (KNN==0) and (Logistic==1) and (Neural Network==0)
and (LinearRegression==1)), then
    MyOwn=0
else if ((SVM==0) and (KNN==0) and (Logistic==1) and (Neural Network==1)
and (LinearRegression==1)), then
    MyOwn=1

```

In the preceding pseudocode MyOwn denotes my convolution algorithm, SVM is the SVM prediction, KNN stands for the KNN prediction, Logistic stands for the Logistic Regression prediction and Neural Network for the Neural Network prediction and finally Linear Regression prediction. Random forest yields the same prediction as Neural Network, so we did not add it.

3.3. Performance Metrics

Before describing our results, it is expedient to first describe the performance metrics used. Indeed, we used metrics like TP rate (recall), precision, accuracy, F-measure and MCC.

TP rate is also called recall; it is described as the proportion of positive data instances that are correctly classified. It is given by $TPR = TP / (TP + FN)$, where TPR stands for True Positive Rate, TP for number of True Positive instances and FN for number of False Negative instances.

Precision measures the strength of a positive prediction; it is given by the ratio of the number of true positives to the total number of positive predictions.

Precision = True Positives / (True Positives + False Positives).

Accuracy reveals how often the algorithm makes a correct prediction. It is given by the ratio of the number of correct predictions to the total number of predictions or

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN).$$

F-measure combines both precision and recall in a single metric; it is given as

$$F1 = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall}).$$

The Matthews Correlation Coefficient (MCC) evaluates the performance of a binary classifier as in our case. It is given by

$$MCC = (TP * TN - FP * FN) / \sqrt{((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN))}.$$

3.4. Experimental Setup

The computer and software we used for this study are described in this section.

The system (laptop) manufacturer is Lenovo, the system model is 20TD00HASP, the system type is x64-based PC and the system name is DESKTOP-IR1UGN4. As for the operating system, it is from Microsoft Corporation, the name is Microsoft Windows 10 Pro and the version is 10.0.19045 Build 19045. Also, the software we used to implement the machine learning algorithms is Anaconda Navigator 2.6.5 with Jupyter Notebook.

4. Results and Discussion

The table below (**Table 1**) shows the results of applying the ML algorithms onto the ZEF dataset using metrics described in the performance metrics section. From the table, we can clearly see that our convoluted rule-based algorithm “MyOwn” outperforms all the other algorithms with almost 90% accuracy. Our algorithm is followed by logistic regression and KNN, which have an accuracy of about 84%. Then, we have SVM and neural network with an accuracy of almost 83%. Furthermore, decision trees and Naïve Bayes are the weakest with 67% of accuracy. Finally, Random Forest and linear regression have an acceptable performance with almost 69% accuracy.

Table 1. Results of different machine learning algorithms on the ZEF dataset.

Algorithms	Recall	Precision	Accuracy	F-measure	MCC
Decision-Tree	0.658	0.774	0.667	0.711	0.329
KNN	0.849	0.886	0.838	0.867	0.66
Logistic Regression	0.849	0.886	0.838	0.867	0.66
Naive Baye	0.658	0.774	0.667	0.711	0.329
Neural Network	0.836	0.884	0.829	0.859	0.644
Random Forest	0.739	0.9	0.786	0.812	0.585
Linear Regression	0.739	0.9	0.786	0.812	0.585
SVM	0.836	0.884	0.829	0.859	0.644
MyOwn	0.89	0.942	0.897	0.915	0.787

We can clearly see that our convoluted rule-based algorithm (MyOwn) outperforms the other algorithms on the ZEF dataset. We think that the higher performance of our algorithm is due to the fact that it learns from the shortcomings of the other algorithms. Therefore, wherever any of them gets it wrong, our algorithms do not consider its prediction, but rather focus on those that get it right. Naïve Bayes performs poorly because it makes a strong assumption about the independence of the attributes, while here the attributes chosen are correlated. Decision tree also does not perform very well because we considered a single test order for the attributes, which is not optimal. On the other hand, when multiple trees are combined, we get a better result. Therefore, random forest combining multiple

trees is more successful than the single decision tree. Logistic regression performs better than linear regression because logistic regression provides a probabilistic prediction, while linear regression is more rigid in its evaluation. We know that soft thresholds perform better than hard ones. As for KNN, its good performance is due to the fact that it considers immediate neighbors ($K=5$) of our target data in its prediction in a majority voting scheme. An odd number of neighbors, 5, was chosen to ensure a majority in the voting. As for neural network, its good performance is due to the choice of the sigmoid activation function for the hidden layer units. Finally, SVM, in the same manner, performs well because of the judicious choice of the SVC (support vector classification) and C parameters of the SVM function. Indeed, we selected the linear for SVC and 1.0 for C after multiple trials with other parameters.

5. Conclusion

In this work, we applied machine learning algorithms to the problem of predicting the ownership of a land in Kumasi (Ghana), whether it belongs to the government or private sector. In addition, we developed a rule-based convolution algorithm. This problem is a serious problem in West Africa. Results show that our algorithm performs better than the algorithms considered. However, accuracy could be substantially improved if we had a bigger dataset than the limited ZEF dataset used. In fact, a larger dataset could improve model robustness and generalizability, which will greatly improve accuracy. And this will be the object of our next work. To conclude, we would like to encourage researchers to apply the machine learning tools to real problems faced by the community as we did for the good of our beloved continent.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Boafo-Anang, M., Akrofi, E.O. and Shitsi, F.J. (2021) Land Disputes and Conflicting Land-Court Judgement Resolution in Ghana: Perspectives of Regulatory Stakeholders. *African Journal on Land Policy and Geospatial Sciences*, **4**, 331-342.
- [2] Sabogu, A., Nasse, T. and Kanto, I. (2020) Understanding Land Conflicts and Food Security in West Africa: Triggers and Perspectives from Dorimon in Ghana. *International Journal of Economics, Business and Management Research*, **4**, 84-108.
- [3] <https://thevaultznews.com/2024/08/14/minister-of-lands-clarifies-on-state-of-public-lands-amidst-allegations/>
- [4] Habebh, H. and Gohel, S. (2021) Machine Learning in Healthcare. *Current Genomics*, **22**, 291-300. <https://doi.org/10.2174/1389202922666210705124359>
- [5] Ersozlu, Z., Taheri, S. and Koch, I. (2024) A Review of Machine Learning Methods Used for Educational Data. *Education and Information Technologies*, **29**, 22125-22145. <https://doi.org/10.1007/s10639-024-12704-0>
- [6] Kelly, B.T. and Xiu, D.C. (2023) Financial Machine Learning. Becker Friedman Insti-

- tute for Economics-Working Paper, No. 2023-100.
- [7] Chopra, V. and Priyadarshi, D. (2019) Role of Machine Learning in Manufacturing Sector. *International Journal of Recent Technology and Engineering (IJRTE)*, **8**, 2320-2328. <https://doi.org/10.35940/ijrte.d8191.118419>
 - [8] Yaokumah, W., Appati, J.K. and Kumah, D. (2021) Machine Learning Methods for Detecting Internet-Of-Things (IoT) Malware. *International Journal of Cognitive Informatics and Natural Intelligence*, **15**, 1-18. <https://doi.org/10.4018/ijcini.286768>
 - [9] (2014) ZEF Kumasi Dataset. <https://bonndata.uni-bonn.de/dataset.xhtml?persistentId=doi:10.60507/FK2/PVXHW5&version=1.0>
 - [10] Evgeniou, T. and Pontil, M. (2001) Support Vector Machines: Theory and Applications. In: Paliouras, G., Karkaletsis, V. and Spyropoulos, C.D., Eds., *Machine Learning and Its Applications*, Springer, 249-257. https://doi.org/10.1007/3-540-44673-7_12
 - [11] Aeeneh, S., Zlatanov, N. and Yu, J. (2025) New Bounds on the Accuracy of Majority Voting for Multiclass Classification. *IEEE Transactions on Neural Networks and Learning Systems*, **36**, 6014-6028. <https://doi.org/10.1109/tnnls.2024.3387544>
 - [12] Halder, R.K., *et al.* (2024) Enhancing K-Nearest Neighbor Algorithm: A Comprehensive Review and Performance Analysis of Modifications. *Journal of Big Data*, **11**, Article No. 113.
 - [13] Kumari, K. and Yadav, S. (2018) Linear Regression Analysis Study. *Journal of the Practice of Cardiovascular Sciences*, **4**, 33-36. https://doi.org/10.4103/jpcs.jpcs_8_18
 - [14] Sperandei, S. (2014) Understanding Logistic Regression Analysis. *Biochemia Medica*, **24**, 12-18. <https://doi.org/10.11613/bm.2014.003>
 - [15] Lingaraj, H. (2016) A Study on Genetic Algorithm and Its Applications. *International Journal of Computer Sciences and Engineering*, **4**, 139-143.
 - [16] Shakya, A.K., Pillai, G. and Chakrabarty, S. (2023) Reinforcement Learning Algorithms: A Brief Survey. *Expert Systems with Applications*, **231**, Article ID: 120495. <https://doi.org/10.1016/j.eswa.2023.120495>
 - [17] Breiman, L. (1996) Bagging Predictors. *Machine Learning*, **24**, 123-140. <https://doi.org/10.1023/a:1018054314350>
 - [18] Al-Bashiri, H., Abdulgabber, M.A., Romli, A. and Hujainah, F. (2017) Collaborative Filtering Recommender System: Overview and Challenges. *Advanced Science Letters*, **23**, 9045-9049. <https://doi.org/10.1166/asl.2017.10020>
 - [19] Southworth, J., Smith, A.C., Safaei, M., Rahaman, M., Alruzuq, A., Tefera, B.B., *et al.* (2024) Machine Learning versus Deep Learning in Land System Science: A Decision-Making Framework for Effective Land Classification. *Frontiers in Remote Sensing*, **5**, Article 1374862. <https://doi.org/10.3389/frsen.2024.1374862>
 - [20] Wang, J., Bretz, M., Dewan, M.A.A. and Delavar, M.A. (2022) Machine Learning in Modelling Land-Use and Land Cover-Change (LULCC): Current Status, Challenges and Prospects. *Science of the Total Environment*, **822**, Article ID: 153559. <https://doi.org/10.1016/j.scitotenv.2022.153559>