THE APPLICATION OF MACHINE LEARNING ALGORITHMS FOR PREDICTION OF PERFORMANCE ACCURACY FOR METAL(LOID) ADSORPTION IN SOIL AND UPTAKE IN WEEDS

Sudipta Biswas[1*], Phani Bhusan Ghosh[1], Soumendra Nath Talapatra[2]

[1]Department of Chemistry
Seacom Skills University
Kendradangal, Shantiniketan,
Birbhum – 731236, West Bengal, India

[2]Department of Bio-Science
Seacom Skills University
Kendradangal, Shantiniketan,
Birbhum – 731236, West Bengal, India

*Corresponding author: Ph.D Scholar

Abstract: The present study was attempted to predict the performance accuracy of datasets of metal(loid) contaminated and uncontaminated soil and uptake capability in weeds through machine learning (ML) classification models by using WEKA tool, version 3.8.5. Different ML algorithm models along with 4 attributes viz. Pb in soil (SPb), Cd in soil (SCd), As in soil (SAs) and effect (normal as N and abnormal as A content) as well as Pb accumulation in plant (PlPb), Cd accumulation in plant (PlCd), As accumulation in plant (PlAs) and effect (N and A content) separately predicted to know overall performance accuracy as per 10-fold cross validation. In the present study, PRC values were recorded the ranged between 87% to 100% for the prediction of metal(loid) content performance accuracy in the soil while PRC values were recorded the ranged between 73% to 96% for the prediction of metal(loid) accumulation performance accuracy in the weeds. It is concluded that ML algorithms performed accurately from the dataset and obtained rich information with statistical validation. The future study in WEKA tool can easily be analysed with more dataset to predict classifier accuracy related to metal(loid) phytoremediation efficiency through weeds.

Keywords: Machine learning algorithm, Model classifier accuracy, WEKA tool, Predictive soil adsorption, Predictive plant accumulation, Metals and metalloids

1. INTRODUCTION

The municipal solid wastes (MSW) comprise different wastes viz. raw vegetable and cooked food wastes, garden wastes, papers, woods, plastics, construction and demolition wastes, glass, ceramics, electrical and electronic wastes, etc. [1] in which few are biodegradables, but majority wastes are non-biodegradable. In the municipality area, the specified barren land is used to dump MSW, which could produce environmental pollution [2,3]. The leachate runoff from open dumping sites showed dominant source of metal(loids) in the surface water and underground water, soil, and finally uptake by plants [4-14].

Several investigations have been carried out especially in and near the solid wastes dumping sites in India and abroad [15-19]. On the other hand, Mandal et al. [20] observed that soil is contaminated due to unsafe disposal of large quantity of arsenic contaminated sludge, which is generated from arsenic removal water treatment plant and arsenic could be adsorbed through waste candles containing arsenic.

In the technique of phytoremediation, several studies have been reported nationally and internationally that some weeds are also able to extract the heavy metals from the soil and could remediate easily from the medium both *in situ* as well as *ex situ* conditions [21-29]. Shahid et al. [30]) revealed that metals or metalloid have tendency to accumulate and translocated to roots and aerial parts viz. stem, leaves, etc. of plant species. It was reported that root is the main target to accumulate but it translocated to the different parts of the shoot of plant species [31]. Biswas et al. [19] observed two weed species (*Lantana camara* and *Sida* sp.) accumulated Pb and Cd into the leaves as hyper-accumulator and could be efficiently used for phyto-remediation for these toxic elements from the soil around solid waste dumping ground of Berhampur Municipality, West Bengal, India.

Interestingly, recent research revealed that big data mining is demanding research in which the endeavour from dataset to valuable information through statistical interpretation. It can easily be accomplished through ML models or artificial intelligence (AI) algorithms, which is predicted the performance accuracy of the dataset [32,33]. On the other hand, several big data analysis on finance, agriculture, biomedical science, bio-science, etc. well established by many researchers [33-38], the data analysis by using ML models to establish limits in the classification of hyperaccumulator plants growing on different metals contaminated soils has already been achieved to know plant mineral composition [38] and recently many investigations are showing interest on the big data analysis by using ML and AI classification algorithms to obtain the accuracy in the big dataset [34-38].

The objective of the present study was to predict the performance accuracy of datasets of uncontaminated and contaminated soil and uptake potential in weeds through machine learning (ML) classification models in the WEKA (Waikato Environment for Knowledge Analysis) tool (version 3.8.5).

2. MATERIALS AND METHODS

In the present study, we used data mining tool namely WEKA (Waikato Environment for Knowledge Analysis) tool (version, 3.8.5) developed by Frank et al. [39] in which performance accuracy could be achieved through ML modelling algorithms. The WEKA explorer was developed with data pre-processing, classification, regression, and association rules [40]. In pre-processing, all the data were made through unsupervised instance and 10-fold cross validation data was used.

The predictive accuracy of dataset on normal and abnormal metal(loid) content in soil and maximum accumulation in the weeds of MSW dumping ground through ML modelling algorithms especially different classifiers viz. BayesNet (BN), NaiveBayes (NB), logistic regression (LR), Lazy.KStar (K*), decision tree (DT) J48, Random forest (RF), Random tree (RT) and Class implementing minimal cost-complexity pruning (CART) along with 4 attributes viz. Pb in soil (SPb), Cd in soil (SCd), As in soil (SAs) and effect (normal as N and abnormal as A content) as well as Pb accumulation in plant (PlPb), Cd accumulation in plant (PlCd), As accumulation in plant (PlAs) and effect (N and A content) separately studied from dataset to predict the overall performance accuracy from the dataset of our earlier study of Biswas et al. [19].

The performance accuracy of above-mentioned ML model classifications related to correctly and incorrectly classified instances, Kappa statistics (KS), mean absolute error (MAE) and root mean squared error (RMSE) were studied for 10-fold cross validation test as per earlier study by Talapatra et al. [37] and Bhattacharya et al. [41]. As per Bouckaert et al. [42], the results for each algorithm model summary were retrieved from WEKA tool. The prediction accuracy of studied ML models as per 10-fold cross validation test was retrieved from summary results and the statistical parameters such as true positive (TP), false positive (FP), Matthew's correlation coefficient (MCC), receiver operating characteristic (ROC) and Precision-recall curve (PRC), respectively were recorded.

3. RESULTS AND DISCUSSION

In the pre-processing step, graphical representation of statistical data of different attributes (SPb, SCd, SAs and effect (N and A) (Fig 1) as well as (PlPb, PlCd, PlAs and effect (N and A) (Fig 2) were obtained. It is not always possible to identify that which part of plants accumulate metal(loids) and these problems can easily be explained by resorting to big data mining, which is the abstraction of implicit, previously unknown, and potentially useful information in data [40]. Generally, ML is used to extract information from raw data of metal(loids) adsorbed soil and accumulated plants [38]. The process is based on abstraction in which data were collected, with all their defects, and the underlying structure is represented [40].

In Fig 1, visual qualitative and quantitative understanding of the distribution class (class effect N as blue coloured and A as red coloured under nominal) in which SPb attribute was obtained ranged between 3.79-26.05 for N category and 26.05-48.32 for A category (11 nos. in each), SCd attribute was obtained ranged between 1.59-3.40 (18 nos.), 3.40-5.22 (2 nos.) and 5.22-7.03 (2 nos.), SAs was found ranged between 2.39-4.54 (5 nos.), 4.54-6.68 (9 nos.) and 6.68-8.83 (8 nos.) and effect attribute viz. N and A category were obtained 11 nos. in each case.

In Fig 2, visual qualitative and quantitative understanding of the distribution class (class effect N as blue coloured and A as red coloured under nominal) in which PlPb attribute was obtained ranged between 0.19-9.68 (12 nos.), 9.68-19.18 (6 nos.) and 19.18-28.67(4 nos.), PlCd attribute was obtained ranged between 0.38-1.01 (16 nos.), 1.01-1.64 (3 nos.) and 1.64-2.27 (3 nos.), SAs was found ranged between 0.35-1.38 (9 nos.), 1.38-2.40 (6 nos.) and 2.40-3.43 (7 nos.) and effect attribute viz. N and A category were obtained 11 nos. in each case.

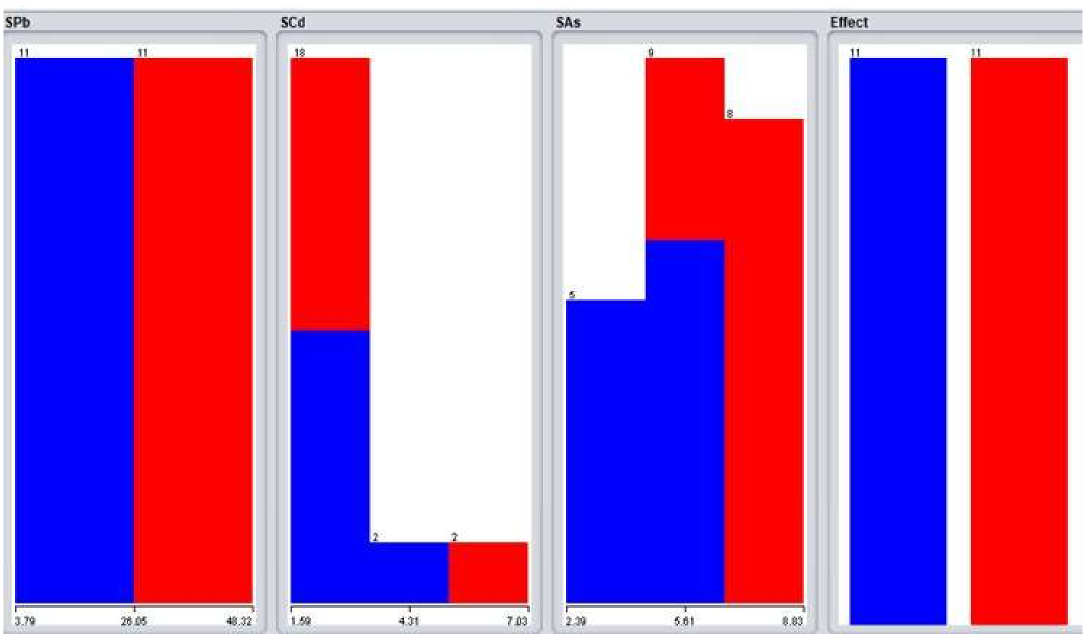

Figure 1: Representation of different attributes of soil after pre-processing in WEKA tool

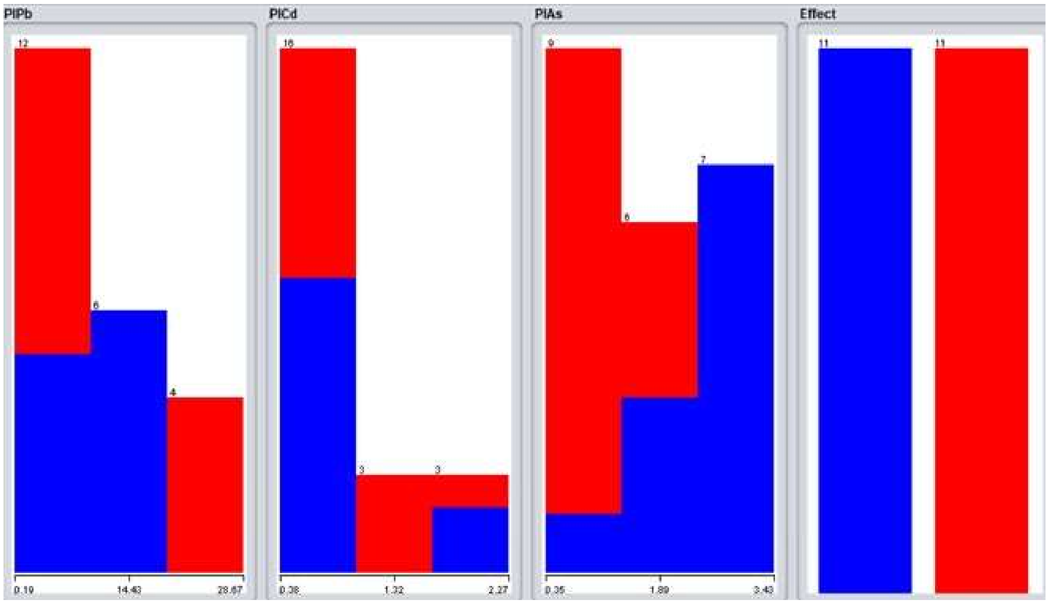

Figure 2: Representation of different attributes of weeds after pre-processing in WEKA tool

Table 1 describes the summary results of studied ML algorithm models such as BayesNet (BN), NaiveBayes (NB), logistic regression (LR), Lazy.KStar (K*), decision tree (DT) J48, Random forest (RF), Random tree (RT) and Class implementing minimal cost-complexity pruning (CART) related to 4 attributes of soil. The performance of model accuracy of above-mentioned ML algorithm classifications as per correctly and incorrectly classified instances, Kappa (K) statistics, mean absolute error (MAE) and root mean squared error (RMSE) were studied as per 10-fold cross validation test. In the case of algorithm model classification, the higher values were observed in BN, NB, LgR, RF and CART (100.00%) followed by J48 (95.45%), and lower value in RT (90.91%) as per 10-fold cross validation test.

Table 1: Results on different classified instances and statistical values for different algorithm models for soil

Classifier model	Correctly classified instances	Incorrectly classified instances	KS	MAE	RMSE
BN	100.0	0.0	1	0.03	0.07
NB	100.0	0.0	1	0.00	0.00
LgR	100.0	0.0	1	0.00	0.00
K*	100.0	0.0	1	0.0005	0.002
J48	95.45	4.54	0.91	0.04	0.21
RF	100.0	0.0	1	0.03	0.06
RT	90.91	9.09	0.82	0.09	0.30
CART	100.0	0.0	1	0.03	0.06

BN = Bayes Network; NB = NaiveBayes; LgR = Logistic Regression; K* = Lazy.KStar; J48 = Pruned and unpruned decision tree C4; RF = Random Forest; RT = Random tree; CART = Class implementing minimal cost-complexity pruning; KS = Kappa Statistics; MAE = Mean Absolute Error; RMSE = Root Mean Squared Error

Table 2 describes the summary results of studied ML algorithm models such as BayesNet (BN), NaiveBayes (NB), logistic regression (LR), Lazy.KStar (K*), decision tree (DT) J48, Random forest (RF), Random tree (RT) and Class implementing minimal cost-complexity pruning (CART) related to 4 attributes of weeds. The performance of model accuracy of above-mentioned ML algorithm classifications as per correctly and incorrectly classified instances, Kappa (K) statistics, mean absolute error (MAE) and root mean squared error (RMSE) were studied as per 10-fold cross validation test. In the case of algorithm model classification, the higher values were observed in K* (95.45%) followed by LogR (86.37%), RF (86.36%), NB and RT (81.82%), J48 and CART (72.72%) and lower value in BN (68.18%) as per 10-fold cross validation test.

Table 2: Results on different classified instances and statistical values for different algorithm models for weeds

Classifier model	Correctly classified instances	Incorrectly classified instances	KS	MAE	RMSE
BN	68.18	31.82	0.36	0.38	0.48
NB	81.82	18.18	0.63	0.17	0.36
LgR	86.37	13.64	0.73	0.17	0.34
K*	95.45	4.54	0.91	0.07	0.22
J48	72.72	27.27	0.45	0.32	0.48
RF	86.36	16.64	0.73	0.24	0.24
RT	81.82	18.18	0.64	0.18	0.43
CART	72.72	27.27	0.45	0.32	0.47

BN = Bayes Network; NB = NaiveBayes; LgR = Logistic Regression; K* = Lazy.KStar; J48 = Pruned and unpruned decision tree C4; LMT = Logistic Model Tree; RF = Random Forest; RT = Random tree; CART = Class implementing minimal cost-complexity pruning; KS = Kappa Statistics; MAE = Mean Absolute Error; RMSE = Root Mean Squared Error

Table 3 describes the representation of the detailed accuracy of studied models for the studied dataset. In case of the accuracy of a class of values of TP, FP, precision, MCC, ROC and PRC, the better performances were observed in BN, NB, LgR, RF and CART followed by J48 and RT for soil. In the present study, PRC values were recorded the ranged between 87% to 100% for the prediction of metal(loid) content performance accuracy in the soil. The ROC curve is depicted (Figs 3, 4 and 5).

Table 3: Statistical data for prediction accuracy of studied algorithms for soil

Classifier model		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC area	PRC area
BN	N	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
	A	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
NB	N	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
	A	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
LgR	N	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
	A	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
K*	N	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
	A	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
J48	N	0.91	0.0	1.0	0.91	0.95	0.91	0.95	0.95
	A	1.0	0.09	0.92	1.0	0.96	0.91	0.95	0.92
RF	N	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
	A	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
RT	N	0.91	0.09	0.91	0.91	0.91	0.82	0.91	0.87
	A	0.91	0.09	0.91	0.91	0.91	0.82	0.91	0.87
CART	N	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
	A	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0

BN = Bayes Network; NB = NaiveBayes; LgR = Logistic Regression; K* = Lazy.KStar; J48 = Pruned and unpruned decision tree C4; LMT = Logistic Model Tree; RF = Random Forest; RT = Random tree; CART = Class implementing minimal cost-complexity pruning; TP = True positive; FP = False positive; MCC = Matthew's correlation coefficient; ROC = Receiver operating characteristic; PRC = Precision-recall curve

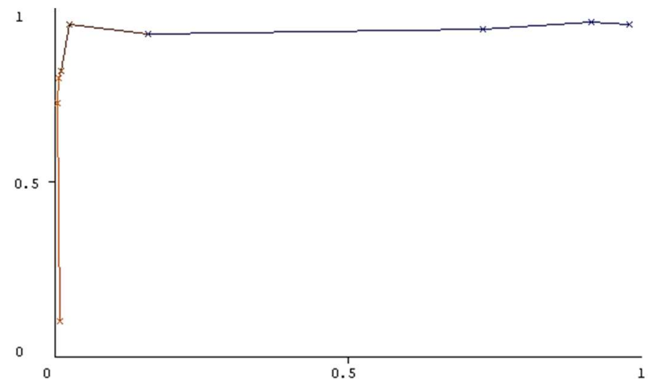

Figure 3: Area under ROC (=1) plot for BN, NB, LgR, K*, RF and CART algorithms of N and A effect for soil

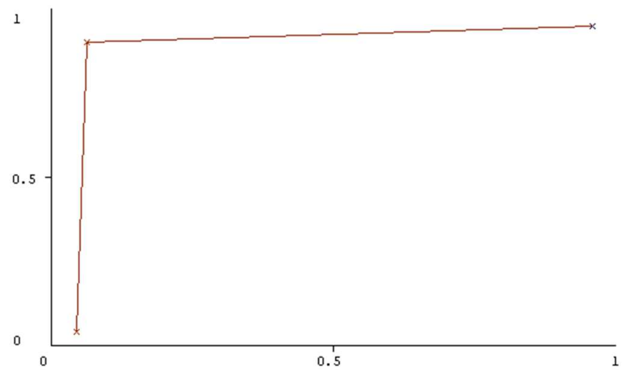

Figure 4: Area under ROC (=0.95) plot for J48 algorithm of N and A effect for soil

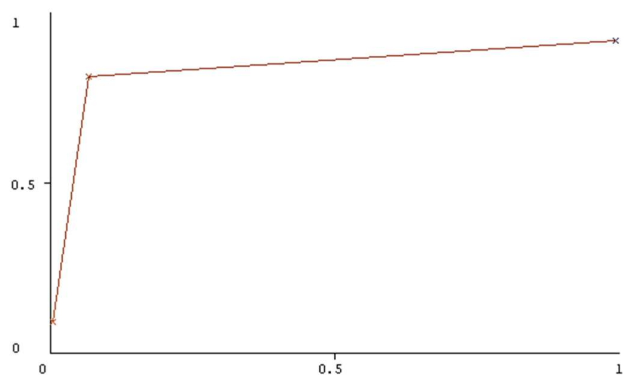

Figure 5: Area under ROC (=0.91) plot for RT algorithm of N and A effect for soil

Table 4 describes the representation of the detailed accuracy of studied models for the studied dataset. In case of the accuracy of a class of values of TP, FP, precision, MCC, ROC and PRC, the better performances were observed in BN, NB, LgR, RF and CART followed by J48 and RT for soil. In the present study, PRC values were recorded the ranged between 73% to 96% for the prediction of metal(loid) accumulation performance accuracy in the weeds. The ROC curve is depicted (Figs 6, 7, 8 and 9).

Table 4: Statistical data for prediction accuracy of studied algorithms for soil

Classifier model		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC area	PRC area
BN	N	0.73	0.36	0.67	0.73	0.70	0.36	0.76	0.80
	A	0.64	0.27	0.70	0.64	0.67	0.36	0.76	0.69
NB	N	0.91	0.27	0.77	0.91	0.83	0.65	0.94	0.96
	A	0.73	0.09	0.89	0.73	0.80	0.65	0.94	0.92
LgR	N	0.82	0.09	0.90	0.82	0.86	0.73	0.89	0.94
	A	0.91	0.18	0.83	0.91	0.87	0.73	0.90	0.80
K*	N	0.91	0.00	1.00	0.91	0.95	0.91	0.92	0.96
	A	1.00	0.09	0.92	1.00	0.96	0.91	0.92	0.85
J48	N	0.64	0.18	0.78	0.64	0.70	0.46	0.72	0.68
	A	0.82	0.36	0.69	0.82	0.75	0.46	0.72	0.73
RF	N	0.91	0.18	0.83	0.91	0.87	0.73	0.89	0.93
	A	0.82	0.09	0.90	0.82	0.86	0.73	0.89	0.79
RT	N	0.91	0.27	0.77	0.91	0.83	0.65	0.82	0.74
	A	0.73	0.09	0.89	0.73	0.80	0.65	0.82	0.78
CART	N	0.64	0.18	0.78	0.64	0.70	0.46	0.76	0.74
	A	0.82	0.36	0.69	0.82	0.75	0.46	0.76	0.75

BN = Bayes Network; NB = NaiveBayes; LgR = Logistic Regression; K* = Lazy.KStar; J48 = Pruned and unpruned decision tree C4; LMT = Logistic Model Tree; RF = Random Forest; RT = Random tree; CART = Class implementing minimal cost-complexity pruning; TP = True positive; FP = False positive; MCC = Matthew's correlation coefficient; ROC = Receiver operating characteristic; PRC = Precision-recall curve

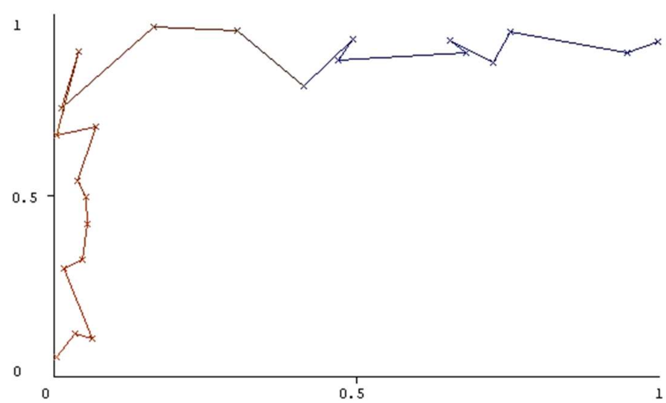

Figure 6: Area under ROC (=0.94) plot for NB algorithm of N and A effect for weeds

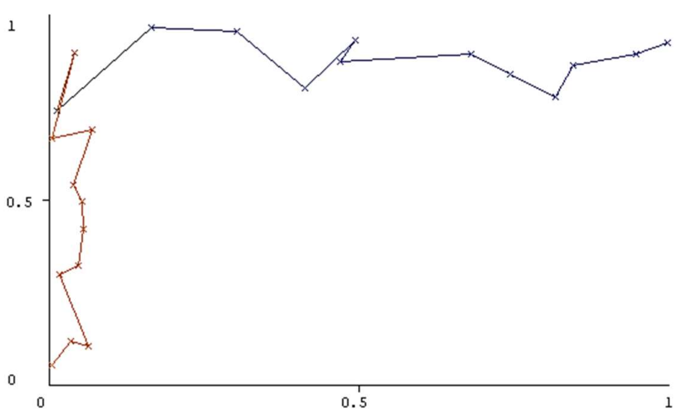

Figure 7: Area under ROC (=0.92) plot for K* algorithm of N and A effect for weeds

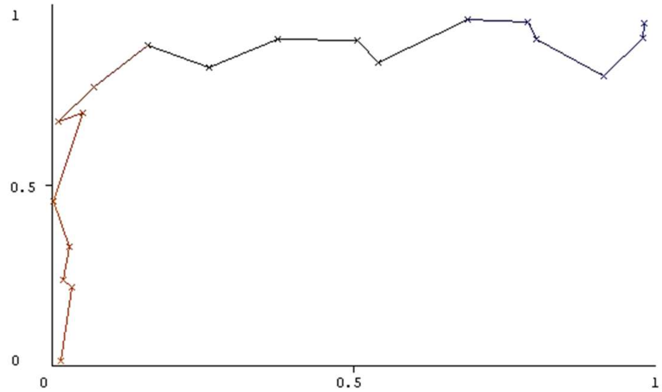

Figure 8: Area under ROC (=0.89) plot for RF algorithm of N and A effect for weeds

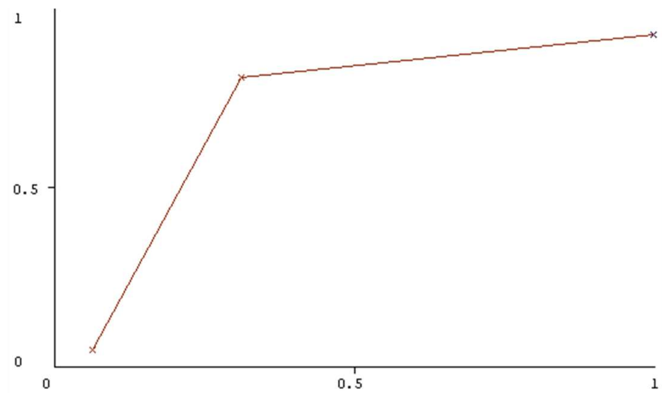

Figure 9: Area under ROC (=0.82) plot for RT algorithm of N and A effect for weeds

Several studies on ML algorithms have been carried out on biological science, [34-36,38,41] etc. but the analysis of dataset through ML modelling algorithm for soil adsorption and plant accumulation of metal(loids) to predict the classifier performance accuracy by using WEKA tool is the first-time endeavour.

4. CONCLUSION

In the present study, PRC values were recorded the ranged between 87% to 100% for the prediction of metal(loids) content performance accuracy in the soil while PRC values were recorded the ranged between 73% to 96% for the prediction of metal(loids) accumulation performance accuracy in the weeds. In conclusion, ML algorithms performed accurately from the dataset and obtained rich information with statistical validation and future study in WEKA tool can easily be analysed with more dataset to predict classifier accuracy related to metal(loid) phytoremediation through weeds.

Acknowledgement

Authors convey thanks to the developer of present tool, which used in the present study.

Funding source

This is a non-funded project.

Conflict of interest

Authors declare no conflict of interest.

REFERENCES

[1] Kolkata Municipal Corporation. Assessment Report on Dhapa Disposal Site Kolkata, India. Prepared under the support of: U. S. Environmental Protection Agency Landfill Methane Outreach Program, **(2010)**.

[2] S. Esakku, K. Palanivelu, and K. Joseph, Assessment of heavy metals in a municipal solid waste dumpsite. In Proceedings of the Workshop on Sustainable Landfill Management, Chennai, India, 3–5 December; 35, **(2003)**, pp. 139-145.

[3] L. Giusti, A review of waste management practices and their impact on human health. Waste Management, vol. 29, **(2009)**, pp. 2227-2239.

[4] R. J. Slack, J. R. Gronow, and N. Voulvoulis, Household hazardous waste in municipal landfills: Contaminants in leachate. Science of the Total Environment, vol. 337, **(2005)**, pp. 119-137.

[5] C. B. Öman, and C. Junestedt, Chemical characterization of landfill leachates-400 parameters and compounds. Waste Management, vol. 28, **(2008)**, pp. 1876-1891.

[6] R. Bakis, and A. Tuncan, An investigation of heavy metal and migration through groundwater from the landfill area of Eskisehir in Turkey. Environmental Monitoring and Assessment, vol. 176, **(2011)**, pp. 87-98.

[7] Y. Y. Long, D. S. Shen, H. T. Wang, W. J. Lu, and Y. Zhao, Heavy metal source analysis in municipal solid waste (MSW): Case study on Cu and Zn. Journal of Hazardous Materials, vol. 186, **(2011)**, 1082-1087.

[8] S. Kanmani, and R. Gandhimathi, Assessment of heavy metal contamination in soil due to leachate migration from an open dumping site. Applied Water Science, vol. 13, **(2013)**, pp.193-205.

[9] R. G. Van Ryan Kristopher, and R. Parilla, Analysis of heavy metals in Cebu city sanitary landfill, Philippines. Journal of Environmental Science and Management, vol. 17, **(2014)**, pp. 50-59.

[10] Y. N. Vodyanitskii, Biochemical processes in soil and groundwater contaminated by leachates from municipal landfills (mini review). Annals of Agrarian Science, vol. 14, **(2016)**, pp. 249-256.

[11] B. Gworek, W. Dmuchowski, E. Koda, M. Marecka, A. H. Baczewska, P. Bragoszewska, and P. Osin´ski, Impact of the municipal solid waste Łubna landfill on environmental pollution by heavy metals. Water, vol. 8, **(2016)**, p.470.

[12] S. R. Samadder, R. Prabhakar, D. Khan, D. Kishan, and M. S. Chauhan, Analysis of the contaminants released from municipal solid waste landfill site: A case study. Science of the Total Environment, vol. 580, **(2017)**, pp. 593-601.

[13] N. Yukalang, B. Clarke, and K. Ross, Barriers to effective municipal solid waste management in a rapidly urbanizing area in Thailand. International Journal of Environmental Research and Public Health, vol. 14, **(2017)**, p. 1013.

[14] N. Vongdala, H. D. Tran, T. D. Xuan, R. Teschke, and T. D. Khanh, Heavy metal accumulation in water, soil, and plants of municipal solid waste landfill in Vientiane, Laos. International Journal of Environmental Research and Public Health, vol. 16, **(2019)**, p. 22.

[15] S. Esakku, A. Selvam, K. Joseph, and K. Palanivelu, Assessment of heavy metal species in decomposed municipal solid waste. Chemical Speciation & Bioavailability, vol. 17 no. 3, **(2005)**, pp. 95-102.

[16] H. Beyene, and S. Banerjee, Assessment of the pollution status of the solid waste disposal site of Addis Ababa city with some selected trace elements, Ethiopia. World Applied Sciences Journal, vol. 14 no. 7, **(2011)**, pp. 1048-1057.

[17] G. Demie, and H. Degefa, Heavy metal pollution of soil around solid waste dumping sites and its impact on adjacent community: the case of Shashemane open landfill, Ethiopia. Journal of Environment and Earth Science, vol. 5 no. 15, **(2015)**, pp. 169-179.

[18] S. Mouhoun-Chouaki, A. Derridj, D. Tazdaıt, and R. Salah-Tazdaıt, A study of the impact of municipal solid waste on some soil physicochemical properties: The case of the landfill of Ain-El-Hammam Municipality, Algeria. Applied and Environmental Soil Science, vol. 2019 **(2019)**.

[19] S. Biswas, S. N. Talapatra, and P. B. Ghosh, Phytoremediation potential of elements by weed species around solid waste dumping ground, Berhampur, West Bengal, India. Pollution Research, vol. 40 no. 3, **(2021)**, pp. 344-352.

[20] P. Mandal, S. R. Debbarma, A. Saha, and B. Ruj, Disposal problem of arsenic sludge generated during arsenic removal from drinking water. Procedia Environmental Sciences, vol. 35, **(2016)**, pp. 943-949.

[21] D.M. Antonsiewicz, C. Escude-Duran, E. Eierzbowska, and A. Sklodowska, Indigenous plant species with potential for the phytoremediation of arsenic and metal contaminated soil. Water, Air and Soil Pollution, vol. 19, **(2008)**, pp. 197-210.

[22] S. Wei, Q. Zhou, H. Xiao, C. Yang, Y. Hu, and L. Ren, Hyperaccumulative property comparison of 24 weed species to heavy metals using a pot culture experiment. Environmental Monitoring and Assessment, vol. 152, **(2009)**, pp. 299-307.

[23] R. D'Souza, M. Varun, J. Masih, and M. S. Paul, Identification of *Calotropis procera* L. as a potential phytoaccumulator of heavy metals from contaminated soils in urban north central India. Journal of Hazardous Material, vol. 184, **(2010)**, pp. 457-464.

[24] H. K. Gurijala, and P. K. Jasti, Phytoextraction of weed plants by survey and analysis in response to lead accumulation. International Journal of Chemical and Biochemical Science, vol. 1 no. 8, **(2014)**, pp. 16-20.

[25] M. Girdhar, N. R. Sharma, H. Rehman, A. Kumar, and A. Mohan, Comparative assessment for hyperaccumulatory and phytoremediation capability of three wild weeds. 3 Biotech, vol. 4, **(2014)**, pp. 579-589.

[26] C. O. Ogunkunle, P. O. Fatoba, A. O. Oyedeji, and O. O. Awotoye, Assessing the heavy metal transfer and translocation by *Sida acuta* and *Pennisetum purpureum* for

phytoremediation purposes. Albanian Journal of Agricultural Sciences, vol. 13 no. 1, **(2016)**, pp. 71-80.

[27] C. E. Anarado, C. J. O. Anarado, C. Agwuna, M. O. Okeke, and P. C. Okafor, Phytoremediating potentials of *Sida acuta* and *Duranta erecta* for lead, cadmium, cobalt and zinc. International Journal of Science and Research, vol. 7 no. 11, **(2018)**, pp. 969-971.

[28] S. Fu, C. Wei, Y. Xiao, L. Li, and D. Wu, Heavy metals uptake and transport by native wild plants: implications for phytoremediation and restoration. Environmental Earth Sciences, vol. 78, **(2019)**, p. 103.

[29] Natasha, M. Shahid, M. Saleem, H. Anwar, S. Khalid, T. Z. Tariq, B. Murtaza, M. Amjad, and M. A. Naeem, A multivariate analysis of comparative effects of heavy metals on cellular biomarkers of phytoremediation using Brassica oleracea. International Journal of Phytoremediation, vol. 22 no. 6, **(2020)**, pp. 617-627.

[30] M. Shahid, E. Pinelli, B. Pourrut, J. Silvestre, and C. Dumat, Lead-induced genotoxicity to *Vicia faba* L. roots in relation with metal cell uptake and initial speciation. Ecotoxicology and Environmental Safety, vol. 74 no. 1, **(2011)**, pp. 78-84.

[31] K. M. Al-Jobori, and A. K. Kadhim, Evaluation of sunflower (*Helianthus annuus* L.) for phytoremediation of lead contaminated soil. Journal of Pharmaceutical Sciences and Research, vol. 11 no. 3, **(2019)**, pp. 847-854.

[32] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning. Nature, vol. 521 no. 7553, **(2015)**, pp. 436-444.

[33] S. Mishra, A. Dash, and L. Jena, Use of deep learning for disease detection and diagnosis. In: Bio-inspired Neurocomputing. Studies in Computational Intelligence. (eds. Bhoi, A., Mallick, P., Liu, C.M. and Balas, V.), Springer, Singapore, **(2021)**, p. 903.

[34] I. Chakraborty, A. Choudhury, and T. S. Banerjee, Artificial intelligence in biological data. Journal of Information Technology & Software Engineering, vol. 7 no. 4, **(2017)**, p. 1000207.

[35] K. P. S. Attwal, and A. S. Dhiman, Exploring data mining tool - Weka and using Weka to build and evaluate predictive models. Advances and Applications in Mathematical Sciences, vol. 19 no. 6, **(2020)**, pp. 451-469.

[36] E. Ropelewska, The application of machine learning for cultivar discrimination of sweet cherry endocarp. Agriculture, vol. 11 no. 1, **(2021)**, p. 6.

[37] S. N. Talapatra, R. Chaudhury, and S. Ghosh, CellProfiler and WEKA Tools: Image analysis for fish erythrocytes shape and machine learning model algorithm accuracy prediction of dataset. World Scientific News, vol. 154, **(2021)**, pp. 101-116.

[38] M. Mota-Merlo, and V. Martos, Use of machine learning to establish limits in the classification of hyperaccumulator plants growing on serpentine, gypsum and dolomite soils. Mediterranean Botany, vol. 42, **(2021)**, p. e67609.

[39] E. Frank, M. A. Hall, and I. H. Witten, The WEKA workbench, Online appendix for data mining: Practical machine learning tools and techniques. Morgan Kaufmann, 4th edition, **(2016)**.

[40] Witten, I. H., Frank, E., Hall, M. A. Data Mining: Practical machine learning tools and techniques. 3rd edn, Morgan Kaufmann, Burlington, MA, **(2011)**.

[41] K. Bhattacharya, B. Mondal, and S. N. Talapatra, Multilayer perceptron network of machine learning for prediction accuracy after genetic biomonitoring of estuarine fish specimen. Journal of Electronics Information Technology Science and Management, vol. 12 no. 10, **(2022)**, pp. 42-50.

[42] R. R. Bouckaert, E. Frank, M. Hall, R. Kirkby, P. Reutemann, A. Seewald, and D. Scuse, WEKA manual for version 3-8-5. University of Waikato, Hamilton, New Zealand, **(2020)**.