JOURNAL PUNJAB ACADEMY OF SCIENCES (Peer Reviewed Open Access Journal)
JPAS, 24: 93-96 (January-December 2024)
Available online at www.jpas.in
ISSN 2229-7014 (P)

SOIL POLLUTION DETECTION USING MACHINE LEARNING: A REVIEW

Nirvair Neeru[1], Navjot Kaur[2]
[1,2]Department of Computer Science & Engineering, Punjabi University Patiala.

ABSTRACT
Soil pollution is a developing natural issue, which leads to extreme biological, horticultural, and general wellbeing issues. Conventional strategies for detection and to monitor soil pollution such as chemical analysis and manual sampling are very tedious, costly, and limited in scope. Machine learning (ML) presents a promising way to overcome these problems by automatic detection of pollutants, prediction of contamination trends, and optimization monitoring strategies. This paper reviews the present status, the difficulties and future capability of ML in the field of soil pollution detection and mitigation.

KEYWORDS: Soil contamination, agriculture, pollutant detection, sensor networks, monitoring, and prediction.

INTRODUCTION
Soil contamination is an ecological worry that influences soil quality, biodiversity, and human wellbeing. It arises by the accumulation of unsafe substances like metals, pesticides, natural pollutants, and plastics. It decrease the quality of the crop. These contaminations can diminish horticultural efficiency, hurt environments, and posture long-term dangers to humans and animals through the food chain[4]. This will lead to decrease in soil quality and poor quality of crops. Over the time, the soil become less productive due to the accumulation of toxic chemicals in large quantity. Recognizing soil contamination is a basic move toward its mitigation. Conventional soil contamination detection strategies depend on manual examining, investigation, and laboratory tests, which are frequently sluggish, costly, and spatially restricted. Moreover, the far, wide, and diffuse nature of soil pollution makes it challenging to successfully monitor enormous regions. Subsequently, there is a developing interest in applying ML procedures to mechanize and improve soil contamination recognition. ML calculations can handle tremendous measures of information from sensors, remote detecting devices, and satellite

images, enables detection and checking of soil pollution very fast. Soil pollution can be caused by various anthropogenic activities like industrial activities, mining, agricultural practices, urbanization and disposal of waste material, oil spills and chemical leaks etc. It has very negative effects on soil[1]. It results into loss of fertility of soil.it develops many health risks like cancer and neurological disorders. It also causes ecosystem disruption. Because of soil pollution, ground water is also contaminated.

RELATED WORK
This section discusses the survey of literature regarding detection of soil pollution using ML techniques.
In Jha, Sunil. (2018) et al. [11] predicted soil pollution trends in urban environments by using geospatial machine learning models. They utilised environmental factors, and ground truth soil sampling data to create detailed pollution maps. In 2019, Tan et al. [14]predicted the pollution level by implementing Random Forest models. They achieved high accuracy with the samples of soil contaminated with heavy metals. In 2021,Gao et al. [7] used soil properties as input features and implemented K-means clustering to identify clusters which are contaminated with heavy metals and analysed microbial diversity in the agriculture soil. Golobočanin, Dušan D. et al.[8] detected patterns in soil contamination within multiple regions. They applied PCA with clustering algorithms and concluded that PCA effectively reduced the complexity of the dataset while retaining important information related to pollution.
In 2024, In 2024, Yulong Wangh et al. [19] detected soil pollution by combining satellite data with machine learning models. Using spectral data, they predicted soil contamination hotspots with high accuracy. Zhaoming Chen et al. [20] classified soil contamination levels using SVM. They achieved very good accuracy in predictions. The experiments demonstrated that SVM effectively distinguished polluted and non-polluted. Anifowosei et al. [4]

applied a deep learning ANN to predict soil contamination by heavy metals. The model was trained on soil texture, pH, and metal concentrations and demonstrated strong performance in detecting polluted sites. Kim YI g et al. [10] detected pesticide contaminated area by using satellite images and ML. The use of a SVM model for anticipating soil weighty metal tainting utilizing information from soil properties and contamination sources, accomplishing high prescient exactness. ML models are also very capable to capture pollutants from soil through remote sensing, multispectral and hyperspectral imaging [13]. The sensor data and ML models can work together to predict soil pollution [12].

SOIL POLLUTION DETECTION

ML plays an important role in the detection of soil pollution. There are several key areas in which ML outperforms.

Pre-processing of Data

ML can deal with complex parameters of, multi-dim information remote detecting symbolism, soil creation, and contamination related boundaries. Remote detecting information, like satellite pictures and robot symbolism, can be investigated with ML calculations to distinguish soil contamination designs over enormous geological regions. These data require feature extraction, normalization, and dimensionality reduction. These tasks require which machine learning techniques like Principal Component Analysis (PCA) and Auto encoders [18].

Detection of Pollutants

The detection of pollutants, which are present in soil, is a tedious task. The supervised ML algorithms effectively predict the pollutants from soil Features like soil pH, electrical conductivity, organic matter content, and concentrations of pollutants can be calculated by using Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), and Neural Networks (NN) [5].

To Identify Pollution Sources

It is very important to find out the pollution sources and hotspots for the detection of pollution. There are many clustering algorithms like k-means, DBSCAN, which comes under unsupervised machine learning techniques, are useful for identifying pollution hotspots or sources of Pollution. These strategies bunch significant pieces of information with comparable qualities and can assist with finding regions with high poison levels without earlier information on the specific pollution dissemination. Clustering algorithms were applied to soil samples to locate regions with excessive concentrations of heavy metals, helping to pinpoint areas in need of remediation [3].

IoT Integration and Real-Time Monitoring

Machine learning is likewise utilized related to sensor organizations and the Web of Things (IoT) for ongoing soil contamination checking. Sensors implanted in the soil can consistently gather information on different natural boundaries, including temperature, dampness, and poison focuses. ML algorithms detect anomalies, trends, or sudden spikes in pollution levels or abrupt spikes in contamination levels [18]. Machine learning algorithms and soil moisture and pollutant sensors work together to explore the track the spread of contaminants in agricultural areas[6]. Such frameworks empower persistent observing of large agricultural fields, guaranteeing that soil quality is kept up with and pollutants are recognized before they cause critical harm.

Analysis for Risk Assessment

ML can likewise help with performing spatial and transient examination to figure framework, which gives the knowledge that how the pollution can be controlled and also evaluate the risk for further accumulation of pollutants in the soil [16].

Mitigation Strategies

With the analysis of historical data and identification of most effective intervention strategies, ML models can suggest the best strategy for soil cleaning. Methods like reinforcement learning, which optimize decision-making over time, can be very valuable in creating versatile techniques for out the elements of soil pollution across different locations. The deep learning models have also been used to predict soil pollution and provide more understanding regarding risk of soil pollution and its impact on living organisms [4]. ML has been combined with geographic data soil remediation [9].

BASIC METHODOLOGY

The Basic methodology of detection of pollution from soil has been shown in **Figure-1**. The first step is collection of sample data. Sample data can be collected through remote sensing, sensor networks and spectroscope or by field survey. It is very important to remove noisy data and to handle missing values, so data pre-processing is performed. After pre-processing, the key features are extracted on

which various machine learning algorithms van be applied like supervised

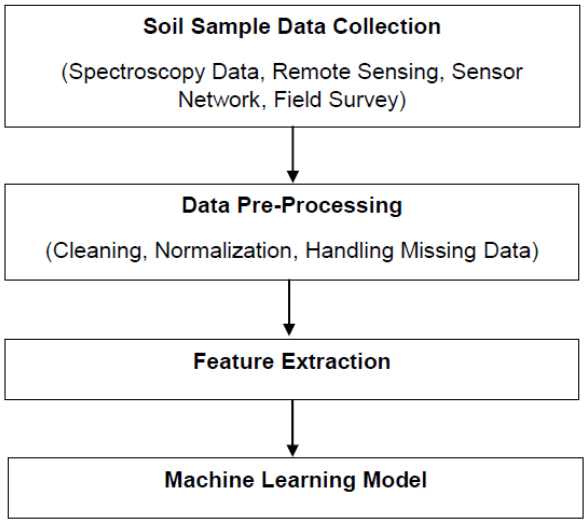

Figure:1 Basic Methodology of Pollution Detection from Soil.

Performance parameters like accuracy, Precision, Recall, Root Mean Square Error, F1Score etc. are used to check the performance of models.

learning professionals face is the absence of good quality data. Inaccurate detection can be made with noisy data and unclean data. Hence, the good quality of input data is necessary to increase the output. It need to ensure that the process of data pre-processing which includes removing outliers, filtering missing values, and removing unwanted features, is done with the utmost level of perfection [2].

The field of machine learning industry is growing very fast. Frequently, new experiments are introduced which makes it very complex and difficult to understand. It includes analysis of input data, training and testing of data, complex mathematical calculations, and many more things. Hence detection of soil pollution it is a complex process which is a big challenge. In addition, detection of soil pollution depends on number of factors like soil type, climate, and land use. ML models are required to deal with these complexities, which can be challenging due to the heterogeneous nature of soil and pollutants [17].

CONCLUSION

Machine learning provides a powerful tool for the detection of soil pollution and to monitor the quality of environment. By leveraging data from remote sensing, sensor networks, and geospatial analysis,

CHALLENGES

Challenges in Machine Learning for Soil Pollution Detection

Data is main factor of machine learning process. The main issue which machine learning methods like regression models and classification models, unsupervised methods like clustering models and deep learning methods like neural networks. **Figure-2** shows the various types of techniques based on machine learning. ML algorithms can automate the detection of pollutants, predict contamination trends, and improve environmental management practices. Despite challenges related to data quality and model interpretability, the integration of machine learning with traditional soil monitoring methods holds great promise for advancing the detection and remediation of soil pollution.

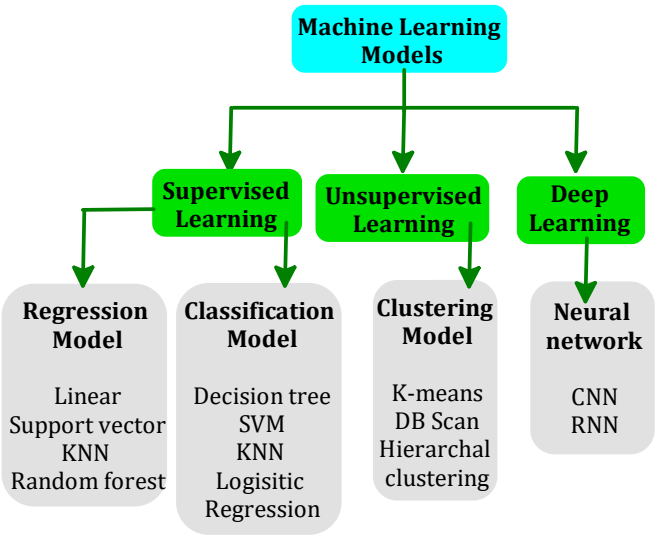

Figure:2 Various Types of Techniques Based on Machine Learning

FUTURE DIRECTIONS

By combining different ML techniques, such as deep learning and traditional statistical models, a hybrid model could improve the accuracy and interpretability of soil pollution detection systems. A continuous, real-time monitoring of soil contamination can be done with the integration of sensor networks with machine learning models. A large-scale collaboration among researchers, policymakers, and environmental organization can be facilitated [15].

REFERENCES

Alloway, B. J. 1996. Soil pollution and land contamination. Pollution, causes, effects and control 3 (1996).

Ananthi, Narayanasamy & Divya, J. & Divya, M. & Janani, V. 2017.IoT based smart soil monitoring system for agricultural production", 209-214. 10.1109/TIAR, 2017.8273717.

Aniagor, Chukwunonso O., Marcel I. Ejimofor, Stephen N. Oba, and Matthew C. Menkiti 2022. Application of artificial intelligence in the mapping and measurement of soil pollution. In Current trends and advances in computer-aided intelligent environmental data engineering, pp. 297-318. Academic Press.

Anifowose, Babatunde, and Fatai Anifowose 2024.Artificial Intelligence and Machine Learning in environmental impact prediction for soil pollution management–Case for EIA Process." Environmental Advances: 100554.

Azizi K, Ayoubi S, Nabiollahi K, Garosi Y, Gislum R. 2022. Predicting heavy metal contents by applying machine learning approaches and environmental covariates in west of Iran. Journal of Geochemical Exploration, 233.

Chakraborty, Tapos Kumar, Most Zaima Mobaswara, MdSimoon Nice, Khandakar Rashedul Islam, Baytune Nahar Netema, Md Sozibur Rahman, Ahsan Habib 2023. Application of machine learning and multivariate approaches for source apportionment and risks of hazardous elements in the cropland soils near industrial areas in Bangladesh. Ecological Indicators 154 (2023): 110856.

Gao, Hongju 2021. Agricultural soil data analysis using spatial clustering data mining techniques. In 2021 IEEE 13th International Conference on Computer Research and Development (ICCRD), pp. 83-90. IEEE.

Golobočanin, Dušan D., Biljana D. Škrbić, and Nada R. Miljević 2004. Principal component analysis for soil contamination with PAHs. Chemometrics and Intelligent Laboratory Systems 72, no. 2 (2004): 219-223.

Jha, Sunil 2018. Soil microbial dynamics prediction using machine learning regression methods. Computers and Electronics in Agriculture, 2018.02.024.

Kim YI, Park WH, Shin Y, Park J-W, Engel B, Yun Y-J, Jang WS 2024. Applications of Machine Learning and Remote Sensing in Soil and Water Conservation. *Hydrology*, 11(11):183.

Mohamed ElsayedSaid, Jalhoum Mohamed E. M. , Hendawy Ehab , El-Adly Ahmed M. , Nawar Said , Rebouh Nazih Y. , Saleh Ahmed , Shokr Mohamed. S. 2024. Geospatial evaluation and bio-remediation of heavy metal-contaminated soils in arid zones. Frontiers in Environmental Science,12,2024.

Molla, Abiot, Weiwei Zhang, Shudi Zuo, Yin Ren, and Jigang Han 2023. A machine learning and geostatistical hybrid method to improve spatial prediction accuracy of soil potentially toxic elements. Stochastic Environmental Research and Risk Assessment 37, no. 2 (2023): 681-696.

Shi, Shuaiwei, Meiyi Hou, ZifanGu, Ce Jiang, Weiqiang Zhang, MengyangHou, Chenxi Li, and Zenglei Xi. 2022. Estimation of heavy metal content in soil based on machine learning models. Land 11, no. 7 (2022): 1037.

Tan, Kun, Weibo Ma, Fuyu Wu, and Qian Du. 2019. Random forest–based estimation of heavy metal concentration in agricultural soils with hyperspectral sensor data. Environmental monitoring and assessment, 191 (2019), 1-14.

Tan, Min, Xiaotong Zhang, WeiqiangLuo, and Ming Hao. 2023. Deep Learning Based Spatial Distribution Estimation of Soil Pb Using Multi-Phase Multispectral Remote Sensing Images in a Mining Area. Land 12, no. 9 (2023): 1789.

Verma, P., and Meena, R. 2020. Machine learning techniques for real-time soil contamination detection using hyperspectral imaging. Environmental Pollution, 267, 115479.

Wang, Xu, Ping Gao, Daping Li, Ju Liu, Nuan Yang, WenzhiGu, Xiaohong He, and Wenzhong Tang 2019. Risk assessment for and microbial community changes in Farmland soil contaminated with heavy metals and metalloids. Ecotoxicology and Environmental Safety 185 (2019): 109685.

Yilmaz, S., &Uysal, S. 2018. Classification of soil pollution based on machine learning techniques. Environmental Monitoring and Assessment, 2018, 190(10), 598.

Yulong Wang, Bin Zou, Xuegang Zuo, Haijing Zou, Bo Zhang, Rongcai Tian, Huihui Feng 2024.A remote sensing analysis method for soil heavy metal pollution sources at site scale considering source-sink relationships. Science of The Total Environment, Volume 946,174021,ISSN 0048-9697,

Zhaoming Chen, Shengli Wang, Jun Xu, Liang He, Qi Liu, Yufan Wang 2024. Assessment and machine learning prediction of heavy metals fate in mining farmland assisted by Positive Matrix Factorization. Journal of Environmental Management, Volume 350, 119587.