


MACHINE LEARNING IN ENVIRONMENTAL MONITORING

UDC 004.8:504.06

Mirjana Milutinović

University of Niš, Faculty of Occupational Safety, Niš, Republic of Serbia

ORCID iD [Mirjana Milutinović](#)  N/A

Abstract. *Rapid advances in artificial intelligence and machine learning techniques, the availability of large-scale data and the increased computational capabilities of machines open the door to the development of sophisticated methods in environmental monitoring and management systems. Machine learning is becoming a powerful tool that is changing the way environmental data is collected, analyzed and interpreted. By going beyond conventional methods, it enables systems to learn from data, identify patterns and draw conclusions with minimal human intervention. Machine learning algorithms can process vast amounts of complex data from diverse sources to provide insight into environmental changes, predict future trends and support decision-making. Using machine learning in the fight for a better and healthier ecosystem has a number of potential long-term benefits. This paper will outline how machine learning can be put to use in environmental monitoring.*

Key words: *machine learning, environmental monitoring*

1. INTRODUCTION

The role of artificial intelligence in environmental monitoring systems has grown exponentially in recent years, fueled by the increasing availability of large amounts of data, advances in computing, and the urgent need to address global environmental challenges. Environmental monitoring involves systematic data collection, analysis and interpretation to assess the state of natural resources and ecosystems. Also, environmental management focuses on using this information to make decisions and take actions to mitigate environmental risks. Because they can analyze vast amounts of data and identify complicated patterns that may not be clearly discernible using existing methods, machine learning techniques can be particularly successful for environmental monitoring and management.

Received October 29, 2024 / Accepted November 8, 2024

Corresponding author: Mirjana Milutinović

University of Niš, Faculty of Occupational Safety, Čarojevića 10a, 18000 Niš, Serbia

E-mail: mirjana.milutinovic@znrfak.ni.ac.rs

Traditional environmental monitoring systems are characterized by disadvantages such as limited coverage, expensive infrastructure, high energy consumption, limited scalability, as well as limited real-time data. In recent years, machine learning has emerged as a game-changing technology in the field of environmental monitoring. It represents enormous potential in environmental monitoring and management, allowing us to gain accurate insight, increase efficiency and detect environmental problems early. The application of machine learning in environmental monitoring can significantly improve our understanding and conservation of natural resources, as well as more effective response to environmental challenges.

Machine learning is suitable for implementation in situations where a large amount of data is available that in some way illustrates solutions to individual parts of the problem. With incredible speed, machine learning algorithms can detect anomalies or deviations in data that signal unusual events or potential hazards in the environment.

This sophisticated branch of artificial intelligence has contributed to revolutionary changes in environmental monitoring and management, empowering scientists, politicians and activists to gain deeper insight, make more accurate predictions and implement more effective strategies.

2. BASIC OF ARTIFICIAL NEURAL NETWORKS AND MACHINE LEARNING

Artificial intelligence represents the development of software that includes the intelligent behavior of computers: acquisition, memory and processing of certain knowledge. It is based on computer science, information technology, mathematics, statistics, psychology, linguistics and many other fields. Artificial intelligence implies a way of reasoning and acting on derived conclusions, with complete reliance on logic, whereby analysis and reasoning are carried out by software applications that do not have to be physically visible. They are often not located in one place, but are active at a large number of virtual addresses and can perform tasks at the same time at billions of different spatial points.

Machine learning is a specific and very narrow branch of artificial intelligence that is the process of discovering patterns in massive data sets, and then making predictions based on decisions made from those patterns. It is based on the idea that systems can learn from the data, identify patterns and draw conclusions independently, without explicit participation i.e. human programming. This approach gives computers the power to learn from experience, which has so far resulted in the surprising performance of computer systems, surpassing the success of human experts in some fields.

Machine learning methods are tasked with identifying relevant regularities in the provided data and defining a mathematical model based on them. Such a mathematical model is applied in situations that were not described within the basic, available database. A good mathematical model that is useful for making inferences in solving future problem instances is the main goal of machine learning. Although errors are always to be expected in solutions based on machine learning, this approach in many fields produces great results that far surpass other approaches.

Although the machine learning method itself can be created by combining several techniques, there is a basic division of machine learning types:

- **Supervised learning:** a type of machine learning where a model is trained on labeled data, meaning that the correct output (also known as a label) is provided for each example in the training data. The goal of supervised learning is to build a model that can make predictions or take actions based on new, unseen input data, generalizing from the patterns it has learned in the training data. If the model performs well on test data, then it can be used to make predictions or take action on new data in the future. For example, a supervised learning algorithm could be trained on a dataset of species identification, land use classification and pollution level prediction.
- **Unsupervised learning:** involves training a model on unlabeled data and learning to identify patterns and relationships in the data on its own, without any guidance from a human supervisor. This type of learning is useful for tasks such as clustering and dimensionality reduction, where the goal is to discover the underlying structure in the data without providing explicit labels for different classes or groups. For example, an unsupervised learning algorithm could be used to identify ecosystem types from remote sensing data, as well as to cluster climate zones.
- **Semi-supervised learning:** is a combination of supervised and unsupervised learning, where the model is trained on a dataset that includes both labeled and unlabeled data. It can be useful when a large amount of unlabeled data is available, but only a small amount of labeled data.

2.1. Types of machine learning algorithms

Machine learning algorithms are great for analyzing large amounts of environmental data, such as air quality measurements, water quality readings, and satellite imagery. By processing this data, machine learning models can uncover hidden patterns, correlations and trends, providing a comprehensive understanding of environmental dynamics.

There are many different machine learning algorithms, which, among other things, continue to be developed and improved to minimize the value of the error. The choice of the specific algorithm to be used for a particular set of data primarily depends on the type of machine learning being performed, as well as the nature of the data and the goals of the model.

Using supervised machine learning regression algorithms on a dataset that includes the height and weight of a group of wild animals, a model could be trained to predict an individual's weight based on their height. During training, the model would use input features (heights) and target values (weights) to learn the relationship between them. Once the model is trained, it can be used to predict new data by providing the model with input features (height) and the model predicts the corresponding target value (weight). Some of the more important types of regression methods in machine learning are linear regression, K nearest neighbors (kNN), decision trees, random forests, and neural network.

Supervised machine learning classification algorithms are used to categorize data into one of several predefined classes. Suppose that in a certain set, there is data that includes the height and weight of a group of individuals, along with their gender (male or female). Using supervised learning classification, a model could be trained to predict an individual's gender based on their height and weight. Such a model can show great efficiency in capturing and classifying wild animals caught on camera. It has also been successfully used to classify land cover types from satellite images, helping to track deforestation and urban expansion. There are several types of classification methods in machine learning, such as logistic

regression, naive Bayes classifier, K nearest neighbors (kNN), support vector machine (SVM), decision tree, random forest, and neural networks.

The goal of unsupervised learning algorithms is to find the underlying structure of a data set, group that data according to similarities, and present that data set in an appropriate format. Suppose an unsupervised learning algorithm is given an input dataset containing images of different types of cats and dogs. The algorithm is never trained on a given set of data, which means it does not know its characteristics, but instead independently identifies common features in the data and reacts based on the presence or absence of such common features in each new piece of data. This approach helps detect anomalous data points that do not fit into any group. Thus, it shows great efficiency in the detection and classification of species in the animal and plant world, as well as in the detection of anomalies in environmental time series data. Some common unsupervised machine learning models are:

- Clustering algorithms that group similar data points together based on their proximity in feature space: K-means clustering, hierarchical clustering, Density-based spatial clustering of applications with noise (DBSCAN).
- Dimensionality reduction algorithms that reduce the dimensionality of data by projecting it onto a lower dimensional space while preserving the most important information: Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Generative models used to generate new data similar to training data, such as Autoencoder and Generative adversarial networks (GAN).

3. APPLICATION OF MACHINE LEARNING IN ENVIRONMENTAL MONITORING

Environmental monitoring refers to the strategy of monitoring and assessing the state of both natural and built environments. Its purpose is to collect, identify and quantify information that can be used to spot patterns and hazards, but also to assess the effectiveness of environmental protection policies and check the impact of human activity on the environment. Machine learning has ushered in a new era in environmental monitoring and management, improving our ability to understand complex ecosystems and develop sustainable practices. The applications of machine learning in environmental monitoring are diverse and constantly expanding.

- Air Quality Monitoring: Machine learning algorithms by analyzing measurement data, above all, can help expand the air pollution monitoring network around the world, predict pollution levels, identify pollution sources and monitor changes over time. They offer the possibility of pairing data from various modern sources that are becoming more and more available and cheaper. Air quality sensors, satellite imagery and meteorological data provide information that can be comprehensively analyzed by machine learning algorithms to predict air quality conditions, as well as identify the main contributors to air pollution in different locations. Pattern prediction, offered by machine learning, enables authorities to take proactive measures to mitigate pollution and protect public health.
- Water Quality Monitoring: Using algorithms to analyze data from water systems, water quality can be monitored, including levels of toxic substances or changes in chemical composition. These insights enable timely action to be taken to prevent contamination, ensure safe drinking water and protect aquatic ecosystems. In water

resource management, machine learning offers help in limiting costs and minimizing environmental impact while reducing or eliminating waste. Excessive water use can be reduced with the help of localized weather forecasts driven by artificial intelligence.

- **Soil Quality Monitoring:** The application of machine learning in this area is promising and can contribute to a more sustainable management of natural resources. Data collection, data analysis, soil quality prediction, problem identification as well as optimization of agricultural practices are just some of the items that machine learning algorithms can improve. Using sensors, drones and satellites to collect data on the physical, chemical and biological properties of the soil, machine learning algorithms integrate the data and process it to identify patterns and trends in soil quality. Automated analysis can detect problem areas, such as pollution or land degradation, enabling rapid response. By generating maps of soil quality, machine learning makes it possible to predict soil erosion and identify optimal conditions for crop growth.
- **Biodiversity Conservation:** In biodiversity monitoring, machine learning is transforming the way we monitor and understand ecosystems. Automatic identification and classification of species based on photographs and acoustic recordings is one of the most interesting applications in this field. By predicting threatened species based on trends in their populations and changes in the environment, as well as early identification of invasive species, machine learning contributes to biodiversity conservation. Species monitoring, optimization of conservation measures, training and education are just some of the techniques that can be improved by analyzing complex ecosystem data, including climate data, soil properties and human activities, provided by machine learning algorithms.
- **Prediction of Natural Disasters:** Using data from the past, machine learning can help predict extreme natural events such as hurricanes, floods, wildfires or droughts. By monitoring and analyzing parameters such as temperature, precipitation, sea level, and concentration of greenhouse gases, this branch of artificial intelligence offers great help for effective planning, response, prevention of natural disasters, as well as assistance for recovery in emergency situations.

In addition to environmental monitoring, machine learning can be applied to climate modeling and prediction, energy efficiency and resource management, natural disaster management, waste and recycling management, ecosystem restoration and biodiversity conservation. Using the power of machine learning algorithms and data analysis, we can gain valuable insights, make informed decisions and take proactive measures to protect our environment and ensure a sustainable future.

4. CONCLUSION

The application of applications based on machine learning algorithms is represented in a wide range of numerous fields. Its influence only continues to grow within all spheres of life. Also, it achieves results that greatly exceed previous achievements and predictions. Together with the Internet, machine learning is changing the way we experience the world and has the potential to be a new driver in solving global environmental challenges.

Environmental monitoring is becoming an increasingly important issue today when considering climate and land cover changes and their environmental consequences. Conventional monitoring methods often struggle to keep up with the scale and complexity of

environmental change occurring on a global scale. All this has led to a significant evolution of the technology behind environmental monitoring in the last few decades. Modern technical innovations provide data with advanced spatial and temporal detail that increase the potential for automatic detection of patterns and trends. In such situations, machine learning algorithms have proven to be a powerful method for relating remote sensing information to relevant environmental variables taking into account the complexity and nonlinearity found in nature.

Machine learning offers the potential to improve accuracy, efficiency and speed, as well as cost savings. It can detect subtle changes in an ecosystem, predict environmental risks, and even suggest mitigation strategies. All this at a scale and speed that would be impossible for human analysts. Overall, machine learning has the power to completely transform environmental management by providing data-driven insights and predictive capabilities for well-informed decision-making and proactive environmental conservation. The combination of mobile survey data and machine learning methods offers great, but still not fully exploited opportunities for monitoring environmental components in different disciplines.

REFERENCES

1. Branković S. (2017.), *Artificial Intelligence and Society*, Institute for Political Studies, Belgrade
2. Pečut L., (2019.), *Recurrent neural networks*, University of Zagreb, Faculty of Science
3. Mančić M., (2019.), *Neural network architecture encoder-decoder for machine translation*, Faculty of Sciences and Mathematics, University of Niš
4. Zdravković M., lectures „Data analysis in companies“, Faculty of Mechanical Engineering, University of Niš
5. Szramowiat-Sala K. (2023.), *Artificial Intelligence in Environmental Monitoring: Application of Artificial Neural Networks and Machine Learning for Pollution Prevention and Toxicity Measurements*, doi: 10.20944/preprints202307.1298.v1
6. Mahalle P.N. (2024.), *Machine Learning for Environmental Monitoring in Wireless Sensor Networks*, doi: 10.4018/979-8-3693-3940-4
7. Meyer H., Machine Learning Methods for Environmental Monitoring, *MDPI Journal Remote Sensing* ISSN 2072-4292
8. The Role of Machine Learning in Environmental Monitoring and Management: A Comprehensive Overview (2023.) available at: <https://medium.com/@binmile/the-role-of-machine-learning-in-environmental-monitoring-and-management-a-comprehensive-overview-7ca6e79e4d7f>
9. Diaz M. (), Machine Learning in Environmental Monitoring, available at: <https://gemmo.ai/machine-learning-in-environmental-monitoring/>
10. Machine Learning for Environmental Monitoring| A Comprehensive Overview (2024.), available at: <https://medium.com/@saiwadotai/machine-learning-for-environmental-monitoring-a-comprehensive-overview-98734f75248a>

MAŠINSKO UČENJE U MONITORINGU ŽIVOTNE SREDINE

Brzi napredak u tehnikama veštačke inteligencije i mašinskog učenja, dostupnost podataka velikih razmera i povećane računarske mogućnosti mašina otvaraju vrata za razvoj sofisticiranih metoda u sistemu za praćenje i upravljanje životnom sredinom. Mašinsko učenje postaje moćno sredstvo koje menja način prikupljanja, analiziranja i tumačenja podataka o životnoj sredini. Prevazilazeći konvencionalne metode, omogućava sistemima da uče iz podataka, identifikuju obrasce i donose zaključke uz minimalnu ljudsku intervenciju. Algoritmi mašinskog učenja mogu da obrađuju ogromne količine složenih podataka iz različitih izvora da bi pružili uvid u promene životne sredine, predvideli buduće trendove i podržali donošenje odluka. Upotreba mašinskog učenja u borbi za boljim i zdravijim ekosistemom ima niz mogućih dugoročnih prednosti. U ovom radu biće izloženo kako se mašinsko učenje može staviti u korist nadzoru životne sredine.

Ključne reči: mašinsko učenje, monitoring životne sredine