A MACHINE LEARNING BASED WSN MODEL FOR WEEDS ASSESSMENT AND MONITORING

M. Nivetha
Department of Electronics and Communication Engineering
Kamaraj College of Engineering and Technology
Virudhunagar, India
nivetha13ece@gmail.com

Dr. V. Jeyalakshmi
Department of Electronics and Communication Emgineering
Kamaraj College of Engineering and Technology
Virudhunagar, India
jeyalakshmiece@kamarajengg.edu.in

***Abstract*—A vast amount of weed plants were accidentally or purposely grown in all the places, and most of them are becoming part of historicity. Many of the introduced weeds have become invasive and problematic. Weeds are species which have been introduced and/or spread from outside their natural range, posing a threat to biological diversity. Weeds have both beneficial and harmful ecological consequences on the ecosystems they invade, hence the study's goal is to evaluate the diverse ecological implications of weeds in various ecosystems. The project's aim is to map the distribution of weeds, as well as their influence on biodiversity and ecosystem services, in order to develop species-specific management strategies and a long-term monitoring program. In this paper, assessing and monitoring weed species, such as Cynodon doctylon and many, as well as developing management plans for major weed species is proposed. Temperature, relative humidity and moisture in different depths are collected over a duration of time where the weeds are located. The evaluation of weeds entails comparing various supervised machine learning algorithms to time - series environmental data and comprehensive review for the collected weed dataset provides accuracy of 99.68% in predicting appropriate weed management strategy using SVM classifier and Logical regression classifier.**

***Keywords*—Weeds, Accuracy, Machine learning, Supervised, Assessment, Monitoring**

I. INTRODUCTION

Man has chosen his food crops among the thousands of plant species that are existing for their nutrient and flavour qualities rather than their competitive ability. Crop production has influenced natural vegetation for ages as part of agriculture. Weeds thrive in the same soil as crop plants. However, in a monoculture system, weed growth is undesirable. These unwanted plants consume the nutrients, water, and space allocated to the intended crop, resulting in a significant decrease in crop yield. Weeds in the crop field lower input efficiency, disrupt agricultural operations, degrade quality, and serve as alternate hosts for a variety of insect pests and illnesses. Some weeds produce a poisonous chemical that inhibits crop growth. The most noticeable result of these features is a multi-fold increase in the expense of cultivation. Furthermore, weeds have an impact on and interfere with the operation of all aquatic and terrestrial resources. They put native biodiversity at risk by suffocating and deliberately taking over native plants. Animals that rely on the native biodiversity for its existence are also being harmed.

Weed scientists can increase effectiveness of these approaches for weed control by learning more about four key ecological factors. First, via competition for resources and niche disruption, multispecies crop rotations, intercrops and cover crops may restrict possibilities for weed growth and regeneration. Second, weed species tend to be more vulnerable than crop species to phytotoxic effects of agricultural residues as well as other organic soil additions [1]. [2] discussed that The study of viruses and their genetics has been an opportunity as well as a challenge for the scientific community. The recent ongoing SARSCov2 (Severe Acute Respiratory Syndrome) pandemic proved the unpreparedness for these situations.. Not only early weed control reduces crop output loss by up to 34%, but it can also reduce the effects of diseases and pests [3]. Precision agriculture, often known as smart farming, has emerged as a cutting-edge solution for addressing today's agricultural sustainability concerns. Machine learning is the mechanism behind this cutting-edge technology (ML). It enables the machine to learn without having to be programmed directly. The next agricultural revolution will rely heavily on machine learning and IoT-enabled farm gear. The authors of this paper provide a comprehensive overview of machine learning applications in agriculture. Predictions of soil parameters like organic matter content and moisture, crop yield predictions, disease and weed identification in crops, and species detection are among the areas where research is concentrated [4].

[5] emphasized that Security is an important issue in current and next-generation networks. Blockchain will be an appropriate technology for securely sharing information in next-generation networks. Digital images are the prime medium attacked by cyber attackers. ML is characterised as a scientific subject

that allows machines to learn without even being strictly programmed, among other things. The threshold measurements were obtained using soil moisture sensors and relative humidity sensors linked in the WSN to assess the possibility of weed growth [7].

Temperature, soil moisture content, and humidity are the three parameters considered. These criteria would aid the recommendation system in providing an accurate prediction of weed management [8]. By collecting the real time sensor data form wireless sensor node, the weeds management strategies are classified to different classes. Section II, presents the proposed work whereas section III lists the methodology and finally section IV tells the results of ML implementation.

II. PROPOSED WORK

The research attempts to provide a simple and user-friendly algorithm that would aid in weed detection and eradication tactics in agriculture. Wireless sensor networks have proven a game-changer, particularly in agriculture. Our model combines WSN and machine learning to produce a model that is 99 percent accurate. The environmental factors observed by the WSN model's sensors are sent into our system, which extracts important information from the factors and compares four different classification algorithms to forecast the best weed removal method for the farms. Figure. 1, shows the overall workflow of the proposed work.

The temperature sensor, humidity sensor, and moisture sensor of the wireless sensor node are used to collect real-time data in the proposed work. Sensor data is collected at pre-determined intervals and kept on-site, where it can be retrieved as needed. The data is in.csv format, which is utilised for machine learning analysis. As there may be some sensor faults in real time, the dataset obtained will have some missing values and outliers. Data pre-processing techniques can eliminate these mistakes in the dataset. [6] discussed about diabetic retinopathy from retinal pictures utilizing cooperation and information on state of the art sign dealing with and picture preparing. The Pre-Processing stage remedies the lopsided lighting in fundus pictures and furthermore kills the fight in the picture. [10] discussed that Helpful correspondence is developing as a standout amongst the most encouraging procedures in remote systems by reason of giving spatial differing qualities pick up. The transfer hub (RN) assumes a key part in agreeable correspondences, and RN choice may generously influence the execution pick up in a system with helpful media get to control (MAC).

The pre-processed dataset is then normalized to a common scale. The normalized dataset is now feature extracted and trained using different ML algorithms and then tested to determine which algorithm provides weed management strategy with highest accuracy for the real time dataset.

III. METHODOLOGY

A. *Hardware used*

The PR2 profile probe measures soil moisture content profiles down to a depth of 1 m. It can be either installed or portable. Access tubes are used to make probe insertion and removal simple. At 10/20/30/40/60 and 100 cm, the PR2/6 features sensing elements. The moisture value of the sensor implies how the ground moisture is affected due to weeds.

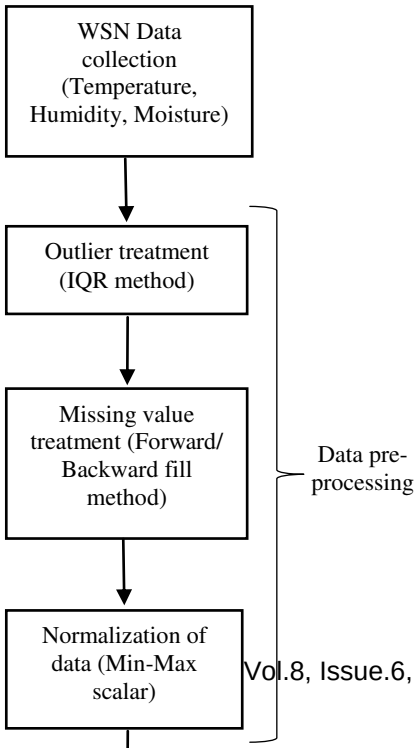
The RHT2nl-02 is a solar radiation shield-mounted relative humidity (RH) and air temperature transducer. It has a high level of RH precision and consistency. Simple 0-1V linear outputs make connecting simple. Transducer power consumption is low. It is made up of sensor modules that may be plugged in and replaced.

The ST4 is a 10K thermistor-based soil temperature probe. The thermistor is enclosed in a stainless-steel sheath with a nominal outer diameter of 6 mm and a length of 120 mm, which is connected to a 5 or 10 metre wire.

The GP2 data recorder is a robust and powerful field data logger. It's suitable for high-stakes research projects. It can log most sensor kinds. 2.5 million readings can be stored in 4 MB of FLASH memory (typical). Data can be collected locally through USB/RS232 or remotely via cellular modem choices by a laptop.

B. *Data Pre-processing*

The raw data obtained from sensors is not suitable for ML models, so the raw data is pre-processed in order to train and test the dataset using ML algorithms. The data obtained may contain several outliers and missing values because of sensing and storing errors.



Outliers in the dataset are identified using the Interquartile Range method (IQR method), in which the dataset is bound between the 75th and 25th percentiles of the values [9]. Interquartile range of the dataset is calculated using equation (1) and the lower and higher outlier in the dataset is found by using equation (2) and (3) respectively. Let Q_1 be the 25 percentile of each column values of the dataset and Q_3 be the 75 percentile of each column values of the dataset.

$$IQR = Q_3 - Q_1 \quad (1)$$

$$\text{Lower outlier} = Q_1 - (1.5 * IQR) \quad (2)$$

$$\text{Higher outlier} = Q_3 + (1.5 * IQR) \quad (3)$$

Thus, the detected higher and lower outlier can be treated using either flooring and capping techniques. If the value of the dataset exceeds the higher outlier threshold, it should be set to the higher outlier value which is called capping and similarly if the dataset value lies below the lower outlier threshold, it is set to be lower outlier value which is said to be flooring.

The missing values in the data set is treated using either forward fill or backward fill method by assigning previous value or next value for NaN because there will not be drastic change in data which is taken in small interval of time.

Normalization is the process of converting real-valued numeric values to a 0–1 range using Min–Max scaling method given in equation (4). Because most machine learning algorithms will reject datasets with varying scales, this is a vital step before splitting the dataset into training and testing data.

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (4)$$

C. *Feature Extraction*

The process of translating raw data into numerical features that may be handled while keeping the information in the original data set is known as feature extraction. It produces better outcomes than applying machine learning to raw data directly. It also decreases the dataset's dimensionality without sacrificing essential information. In exploratory data analysis and ML for prediction algorithms, Principal Component Analysis (PCA) is the most extensively used tool [11]. It is also an unsupervised statistical tool for examining the interrelationships between a set of variables. Regression determines a best fit line, which is also known as a generic factor analysis. PCA is a technique for reducing the complexity of high-dimensional data while preserving trends and patterns. It accomplishes this by condensing the data to fewer dimensions that serve as feature summaries. PCA technique is used here to reduce three variables temperature, moisture and humidity into two major principal components without loss of information.

D. *Machine Learning algorithms*

The KNN (k-nearest neighbor) algorithm is a supervised machine learning technique that can be used to handle regression and classification problems. It classifies the data based on the classification of its neighbors. New data points are classified by KNN using the similarity metric of the previously stored data points.

The supervised learning algorithm for classification, logistic regression is used to predict the likelihood of a target variable. Because the nature of the target is dichotomous, there are only two classifications. The sigmoid function is used in logistic regression to transfer predicted values to probabilities. This method converts any real value to a number between 0 and 1.

The Bayes' Theorem is used to create a collection of different classifiers known as Naive Bayes classifiers. It is a collection of algorithms that share a similar idea, namely that each pair of variables being classified is autonomous of the others. The feature matrix and the response vector are the two sections of the dataset.

SVM (Support Vector Machine) is a common Supervised Learning technique for Classification and Regression. However, it is most commonly employed in Machine Learning for Classification issues. The data can be linear or non-linear when using SVM Classification. The SVM algorithm's purpose is to find the best decision boundary that can divide n-dimensional space into classes so that fresh data points can be readily placed in the correct category in the future. A hyperplane is a term for the optimal decision boundary.

IV. RESULTS AND DISCUSSION

For analysis, a dataset of 1000 samples were obtained from the wireless sensor node. There are three independent variables in the dataset: temperature, humidity, and moisture, as well as one dependent variable, target. The samples are intended to fall into one of two categories: clear cutting or land restoration. Figure 2(a) shows the raw data set, which includes some missing values and outliers. After that, the raw data set is treated for missing values using the forward fill method, as well as outliers using the IQR method along with flooring and capping. The dataset is shown in figure 2(b) after these procedures.

	Temperature	Humidity	Moisture
count	998.000000	995.000000	1000.000000
mean	29.076772	57.666468	871.877361
std	5.743891	16.715253	153.691408
min	10.000000	13.428571	652.330000
25%	27.000000	46.875000	698.770000
50%	30.000000	58.142857	997.687500
75%	32.125000	68.593750	1004.250000
max	100.000000	200.000000	1019.750000

(a)

	Temperature	Humidity	Moisture
count	998.000000	998.000000	998.000000
mean	29.233288	57.568016	871.877361
std	4.774248	16.150432	153.691408

The pre-processed data is then normalized to the range of 0 to 1. The correlation of these three independent variables is shown in figure 3.

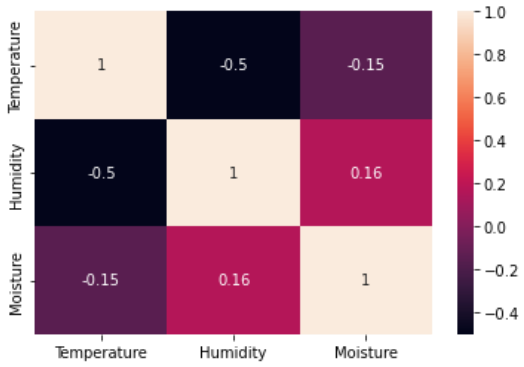


Figure 3. Correlation heatmap of dataset features

The dataset is visualized after PCA, with the x-axis displaying the principal component-1 and the y-axis displaying the principal component-2, and the targets plotted to differentiate the two unique classes in figure 4.

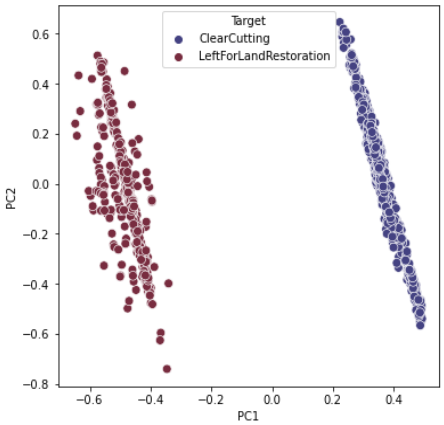


Figure 4. Dataset visualization after PCA

The data is split into training and testing data in the ratio of 4:1 so that there will be 800 training data and 200 testing data. Different ML algorithms are trained using training data and then predictions are made for testing data. The prediction accuracy shows the performance of ML algorithms for the real time dataset.

Table 1. Performance comparison of ML algorithms for weed management strategies

ML Algorithms	Prediction accuracy
KNN classifier	99.36%
Naïve Bayes classifier	99.36%
SVM classifier	99.68%
Logical regression classifier	99.68%

Table 1 shows a comparison of several machine learning algorithms based on prediction accuracy. Although all of the strategies have a respectable accuracy score, we selected to sstart with SVM classifier and Logical regression classifier because it has been shown to have higher accuracy and better prediction outcomes than the others. Since the number of training data is very large when compared with number of features, SVM classifier outperforms over other classifier. The dataset is almost linear and the samples are dependent on each other, so Logical regression algorithm which is discriminative provides high accuracy.

V. CONCLUSION

Agriculture is equipping farmers with technologies in order to provide the best results with the most exact inputs. Smart sensors and actuators, the Internet of Things, and wireless sensor networks are just a few of the significant technical advancements that have benefited agriculture. These components are critical for obtaining real-time data and making judgments without the need for human intervention. Artificial intelligence, or automation of intelligent behavior, continues to benefit our planet and assist humans in a variety of ways. We have proposed a model for giving appropriate strategy in weed management based on temperature, humidity and moisture characteristics is built using the currently popular machine learning algorithm. For data samples taken over a month, the Logical regression classifier and SVM classifier have a greater accuracy of 99.68 percent.

As scope of future work, the work can be extended by exploring deep learning algorithms and other hybrid algorithms with more precise management strategies.

REFERENCES

[1] M. Liebman and A. S. Davis, “Integration of soil, crop and weed management in low-external-input farming systems,” Weed Research, 2002, vol. 40, USA, pp. 27-47.

[2] Christo Ananth, Pranav Pushkar, Preeti Nagrath, Jehad F. Al-Amri, Vividha, Anand Nayyar, “Mutation Prediction for Coronaviruses using Genome Sequence and Recurrent Neural Networks”, Computers, Materials & Continua, Tech Science Press, Volume 73, Issue 1, April 2022,pp. 1601-1619.

[3] Mojtaba Dadashzadeh, Yousef Abbaspour-Gilandeh, Tarahom Mesri-Gundoshmian, Sajad Sabzi, Jose Luis Hernández-Hernández, Mario Hernández-Hernández and Juan Ignacio Arribas, “Weed Classification for Site-Specific Weed Management Using an Automated Stereo Computer-Vision Machine-Learning System in Rice Fields,” Plants, 2020, vol. 9, pp. 1-19.

[4] Abhinav Sharma, Arpit Jain, Prateek Gupta and Vinay Chowdary, “Machine Learning Applications for Precision Agriculture: A Comprehensive Review,” IEEE Access, 2021, vol. 9, pp. 4843-4873.

[5] Christo Ananth, Denslin Brabin, Sriramulu Bojjagani, “Blockchain based security framework for sharing digital

images using reversible data hiding and encryption", Multimedia Tools and Applications, Springer US, Volume 81,Issue 6, March 2022,pp. 1-18.

[6] Christo Ananth, D.R. Denslin Brabin, Jenifer Darling Rosita, "A Deep Learning Approach To Evaluation Of Augmented Evidence Of Diabetic Retinopathy", Turkish Journal of Physiotherapy and Rehabilitation, Volume 32,Issue 3, December 2021,pp. 11813-11817.

[7] Ishita Dasgupta, Jayit Saha, Pattabiraman Venkatasubbu and Parvathi Ramasubramanian, "AI Crop Predictor and Weed Detector Using Wireless Technologies: A Smart Application for Farmers," Arabian Journal for Science and Engineering, 2020, vol. 45, pp. 11115-11127.

[8] T. Banavlikar, A. Mahir, M. Budukh and S. Dhodapkar, "Crop recommendation system using Neural Networks," International Research Journal of Engineering and Technology, 2018, vol. 5, pp. 1475-1480.

[9] B. J. Sowmya, S. Seema, Daniel Monteiro, Zaid Hossain, Zubin Paymaster and Aditya Tomar, "Machine learning techniques on agro meteorological parameters to predict the crop yield in support of building decision support system to farmer's community," Journal of Physics: Conference series. 1706, 2020.

[10] Christo Ananth, Joy Winston.J., "SPLITTING ALGORITHM BASED RELAY NODE SELEC-TION IN WIRELESS NETWORKS", Revista de la Facultad de Agronomía, Volume 34, No. 1, 2018,(162-169).

[11] Shanshan Duan, Weidong Yang, Xuyu Wang, Shiwen Mao and Yuan Zhang, "Forecasting of Grain Pile Temperature From Meteorological Factors Using Machine Learning," IEEE Access, 2019, vol. 7, pp. 130721- 130733.