International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Identification of Medicinal Leaves and Recommendation of Home Remedies using Machine learning

[1]Dr. Swati G. Kale, [2]Dr. Sachin Jain, [3]Dr. Sudhir Rangari, [4]Dr. R. C. Dharmik, [5]Dr. Manish Gardi, [6]Vipul S. Lande,

Submitted: 22/11/2023 **Revised**: 27/12/2023 **Accepted**: 08/01/2024

Abstract: In today's era there is a growing need for programmed medical assistance systems. Many medical assistance systems are available for recommendation of allopathic medicine but ayurvedic medicine identification and assistance system is still unexplored. However, with the growing urbanization and diminishing knowledge of traditional practices, the accurate identification of ayurvedic leaves and the appropriate recommendation of home remedies and medicines have become challenging. To address this issue the proposed system presents state-of-art deep learning algorithms to identify the ayurvedic leaf based on images. This system also helps in recommendation for ayurvedic medicine based on symptoms of patient along with some suggested home remedies. This study involves the development of an Android-based application aimed at classifying medicinal leaves and identifying diseases. The dataset comprises 6,541 images representing 115 distinct species of medical leaves. To achieve leaf classification, well-known pre-trained neural networks like Convolutional Neural Networks (CNN), VGG16, MobileNet, and Inception are utilized. Furthermore, disease identification is a part of this research, which involves a dataset encompassing 35 symptoms and 20 diseases.The conventional approach of splitting the dataset into training and testing subsets is employed, a practice that not only assists in training the model but also in evaluating its performance without overfitting. To evaluate the system's performance comprehensively, diverse measurement parameters are implemented. These parameters encompass widely used evaluation metrics such as accuracy, precision, recall, F1-score, as well as confusion matrices. This array of metrics enables a thorough understanding of the model's effectiveness and its ability to classify different classes accurately.

Keywords: *Machine learning, Ayurveda, Home remedies*

1. Introduction

Ayurveda is one of the old medicinal system in India is Ayurveda. It was developed more than 5,000 years ago in India. In Ayurveda Natural remedies are used to heal the human body. Natural remedies are made up from herbs, roots, leaves, stem and other natural ingredients. In this era there is a need to get the knowledge of these natural remedies which is used to heal the human body naturally and build the immune system. Ayurvedic remedies are made from. Ayurvedic leaf identification is an important part of Ayurveda as it helps identify the right leaves for the right remedy. The natural elements like alovera,Tulsi,Haldi(turmeric),black pepper, elachi, ginger, saffron, cinnamon are highly valued for their effectiveness in addressing throat ,skin-related ailments, fever, boils [1].The research work presented in this paper has two modules. First module introduces an automated system, designed to accurately identify Ayurvedic herbs from a wide range of plant species. Additionally, the system offers essential information about each herb, including its scientific name, natural habitat, weather requirements, potential side effects, recommended consumption methods, as well as its vitamin and mineral content. The recognition of these medicinal herbs is accomplished by leveraging the knowledge and expertise of Ayurveda practitioners and knowledge passed down from older generations. In second module based on symptoms of patient the system will identify the correct disease and suggest the recommended medicines along with home remedies. This study represents a cooperative effort that unites specialized technical knowledge with the insights of Ayurvedic physicians from within the Ayurvedic medical community. Together, they have joined forces to work on this project and combine their respective knowledge domains to achieve the study's objectives.

[1]*Assistant Professor Department of Information Technology, Yeshwantrao Chavan College of Engineering, Nagpur, India*
s12_kale@yahoo.com
[2]*Assistant Professor Department of Computer Science, Oklahoma State University Stillwater, United States.*
sachin.jain@okstate.edu
[3]*Assistant Professor Department of Information Technology*
JSPM'S Jayawantrao Sawant College of Engineering, Hadapsar, Pune, India
rangari.sr@gmail.com
[4]*Assistant Professor Department of Information Technology, Yeshwantrao Chavan College of Engineering ,Nagpur, India*
raj_dharmik@yahoo.co.in
[5]*Assistant Professor Department of Computer Engineering JSPM'S Jayawantrao Sawant College of Engineering, Hadapsar, Pune, India*
manishgardi@gmail.com
[6]*Department of Electronics and Communication Engineering Shri Ramdeobaba College of Engineering and Management, Nagpur, India*
landevs@rknec.edu

2. Litreture Review

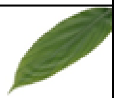
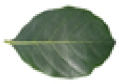
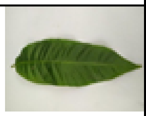
In their publication [2], Pushpa BR, Anand C, and Mithun Nambiar P detail their approach, which encompassed preprocessing, feature extraction, and classification techniques applied to a diverse set of 208 samples from 28 distinct species. Their methodology involved utilizing k-means clustering for classification, yielding an impressive accuracy rate of 93.75%.In their research article [3], Sultana Umme Habiba, Md. Khairul Islam, and Sk. Md. Masudul Ahsan explore the application of deep convolutional neural networks. Their investigation encompasses an array of prominent models including VGG16, VGG19, ResNet50, InceptionV3, Inception-ResNetv2, and Xception. Notably, they achieved the highest accuracy of 96%, a feat accomplished by the VGG16 model. The dataset utilized in their study encompasses both Ayurvedic plants and endangered species. A. Anushya mentioned in their paper [4] classification of herbal leaves using image clustering k-means. They extracted features and clustered using Gray level concurrence matrix and k-means. Their image dataset has 7 common herbal leaves such as coriander, betel, neem mint, etc, with 25 to 50 images for each leaf, making a total of 307 images. They got an accuracy of 86.96%.Sethulekshmi. A.V and Sreekumar K present in their paper [5] the feature that are ignored most of the time. They mentioned different features of leaves that are not well used. They compared performance of different features, extraction methods.M. Arun , R. Ahila Priyadarshani and S. Arivazhagan mentioned in their Springer paper [6] the usage of six level convolutional neural network (CNN) and has accuracy of 87.25%. Their dataset contains 50 species of selected leaves. This researches presents the results on recognizing the plants based on spatial, spectral and machine learnt features.Madhuri Bnadara and Lochandaka Ranathunga in their IEEE paper [7] compared different algorithm such as K-NN, decision trees, SVM to classify ayurvedic leaves. They proposed approach extracts color, shape and different texture features form leaves which extracts three features vectors such as Haralick, Tamura and Gabor textures. The accuracy obtained was 86.52%.Amrutha Raghukumar and Gayatrhi Narayan in their IEEE paper [8] mentioned the use of feature of plants such as shape, texture and color is done. Then they applied classification techniques such as KNN and SVM and compared their performances. The recorded accuracy stood at 93.23%. Their dataset comprises images sourced from the Amritavanam Herbal Garden, affiliated with the School of Ayurveda. In her scholarly work [9], C K Gomathy introduced a system harnessing the capabilities of Machine Learning Technology. Within this framework, the Naïve Bayes algorithm takes center stage, employed for disease prediction through symptom analysis. The process involves employing the K-Nearest Neighbors (KNN) algorithm for classification purposes, while Logistic Regression serves as the mechanism for extracting impactful features. Additionally, the system benefits from the utilization of Decision Trees to segment the extensive dataset into more manageable components. Ultimately, the system culminates in generating disease predictions as its final output.Md. Maniruzzaman, In their scholarly publication [10], Md. Jahanur Rahman, Benojir Ahammed, and Md. Menhazul Abedin outline their research, which focuses on diabetes risk assessment. They employ a Logistic Regression (LR) model to discern diabetes risk factors through the evaluation of p-values and odds ratios (OR). Additionally, they incorporate four essential Machine Learning (ML) classifiers—Naïve Bayes (NB), Decision Tree (DT), AdaBoost (AB), and Random Forest (RF). The primary aim of their study is to pinpoint the most influential factors associated with diabetes using the LR model, while concurrently developing an ML-based system for precise risk stratification of the disease. In his scholarly contribution [11], Min Chen introduces a novel algorithm termed the Convolutional Neural Network-based Multimodal Disease Risk Prediction (CNN-MDRP). This algorithm uniquely incorporates both structured and unstructured data from hospital sources. Remarkably, this research distinguishes itself by addressing both data types in the realm of medical big data analytics—an aspect previously unexplored by existing studies. Notably, when compared to various conventional prediction algorithms, the proposed algorithm achieves an impressive prediction accuracy of 94.8%. Additionally, it demonstrates a faster convergence speed in comparison to the CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm. In their paper [12], Amin Ul Haq, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir, and Ruinan Sun present a comprehensive framework called the Hybrid Intelligent System for Heart Disease Prediction using Machine Learning Algorithms. This framework centers around a diagnostic system tailored for heart disease. The approach involves the integration of three Feature Selection (FS) algorithms, seven distinct classifiers, a cross-validation technique, and performance evaluation metrics dedicated to heart disease diagnosis.The system's efficacy is demonstrated through testing on the Cleveland heart disease dataset, where its capability to distinguish between heart disease cases and healthy subjects is assessed. The authors emphasize the benefits of constructing a decision support system through machine learning techniques, providing an advantageous approach for heart disease diagnosis.Of particular note, the study introduces a creative facet by incorporating feature selection algorithms. These algorithms play a pivotal role in enhancing classification accuracy and diminishing computation time by identifying and employing only the most pertinent features, thus mitigating the impact of irrelevant attributes on system performance.

Regenerate. Rinkal Keniya, Aman Khakharia ,Vruddhi Shah , Vrushabh Gada , Ruchi Manjalkar , Tirth Thaker · Mahesh Warang , Ninad Mehendale presented in their paper [13] the technique of predicting the disease based on the symptoms, age, and gender of an individual patient. The Weighted KNN model gave the highest accuracy of 93.5 % for the prediction of diseases using the above-mentioned factors. Almost all the ML models gave good accuracy values. Due to the parameter-dependent nature of certain models, their disease prediction capabilities were hindered, resulting in relatively low accuracy percentages. However, once an accurate disease prediction is achieved, it would streamline the allocation of necessary medicinal resources for treatment. Consequently, this model holds the potential to significantly reduce the costs associated with managing the disease while concurrently enhancing the overall recovery process.

3. Dataset Creation

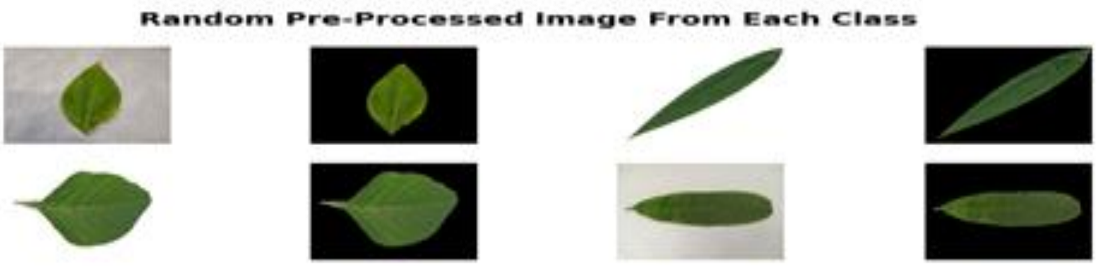
Most crucial part of any research is acquiring a data set. For our research we have used a fusion of two data set,a custom data set and online data set. Custom Dataset is acquired from the nearest ayurvedic college Datta Meghe Ayurved College, Hospital and Research Centre, Nagpur, which has a wide range of ayurvedic plant species in its herbal garden. The custom dataset contains 35 different species of different medicinal leaves, with around 25 – 30 photos of each species which adds up to 929 images. The online dataset contained 80 different species of different medicinal leaves, which adds up to 5612 images. So, total images collected are 5,612+929 = 6,541 images. During the training of the dataset, the model was becoming "overfitted," and it was unable to generalize well to new data. So, we augmented the data. Data augmentation acted as regularize and assisted in managing the overfitting of data. It also increases the capacity of the model by creating additional information and extending the model for different data types. Augmentation techniques such as horizontal, vertically flipping, positive negative 45-degree rotation, wrap shift, adding noise in the images, etc. were used. After augmentation the online dataset and custom dataset are 12,500 and 4,390 images respectively. So, total images by combing both the dataset are 12,500+4390 = 16,890 images. Table 1 shows the sample images from dataset.

Table I: Sample images from data set

Leaf Image	Leaf Name	Leaf Image	Leaf Name	Leaf Image	Leaf Name
	Adhatoda Vasica_		Alpinia Galanga		Artocarp us Heteroph yllus
	Amaranth us Viridis		Annona retisulate		Asthma Plant
	Azadiracht a Indica		Basella Alba		Gommip hora mukul

For improving the data quality Image processing is done. Color segmentation method is used for pre-processing, which blacks out background . To mitigate camera noise, image filtering techniques are employed, including the utilization of a low-pass filter for noise reduction and filtration. An added benefit of this pre-processing approach is the enhanced treatment of pixel surroundings. To enhance the efficacy of automated image-processing procedures, it is imperative to enhance the interpretation of information contained within images. Once the pre-processing phase is completed, the images are primed for utilization as input data for the algorithm.To enhance the dataset's consistency, reliability, authenticity, and usability, we subjected it to cross-validation and expert testing. The dataset were divided into 80% training and 20% images for testing. Table 2 shows some random pre-processed images.

Table II: Random pre-processed images from each class.



We have also created a database for identification of disease which consist 35 symptoms which are used to identify 20 disease. The labelled matrix for identification of disease is shown in Table 3.

Table III: Sample Labeled Dataset

headache	fever	cold	cough	Pail body	Yellow eyes	Dry skin	Disease
1	1	1	1	1	0	0	1
0	1	0	0	0	0	0	2
0	0	0	0	1	1	1	3

4. Proposed Methodology

We have extended our research work [14] for leaf identification by applying Convolutional Neural Network (CNN), Residual Network (ResNet), VGG neural networks, Mobile-Net, Inception pre-trained networks. The obtained results for leaf identification is given in table 3 of result and discussion section.

After doing the various experimentation using above listed network it is observed that CNN is best situated network for our dataset. After performing the task of leaf identification detection of diseases based on symptoms and provide more details about the top fetched diseases including treatment recommendation is done .The overall working of our system is depicted in fig 1.

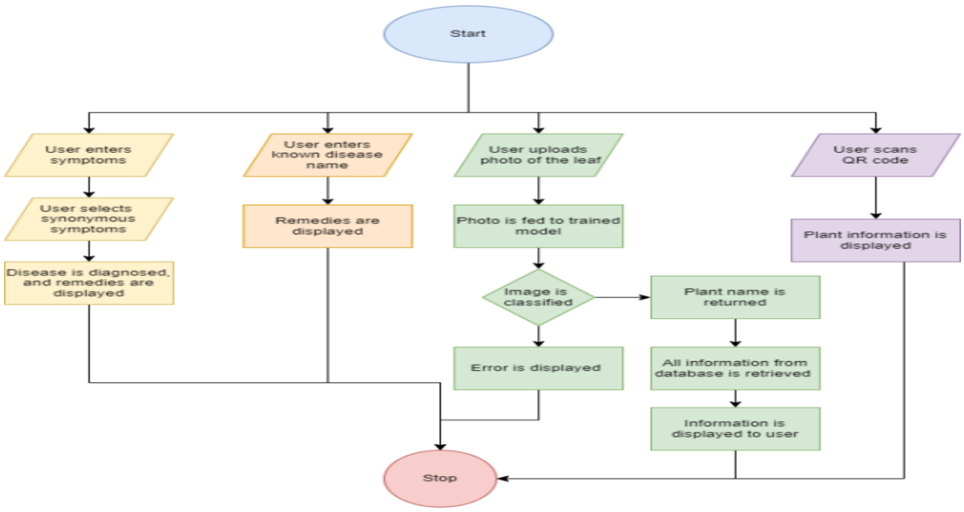


Fig 1: Working of proposed Methedology

The disease prediction Dataset is divided into a training set and a testing set. The training set goes to seven different pre-trained models. The different pertained models used for this phase are the MNB (Multinomial Naive Bayes),RFMLS (Radio Frequency Machine Learning Systems),KNN (K-Nearest Neighbours Algorithm),LR (Logistic Regression),SVM (Support Vector Machine),DT (Decision Tree),MLP (Multi-layer perceptron) models. The details phase of training and evaluation phase is given in figure 2.

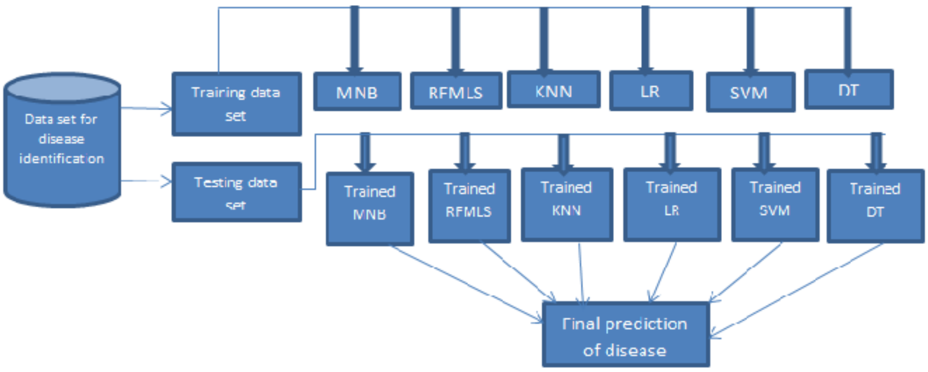


Fig 2: Proposed methodology Training and testing Phase for disease identification.

5. Results & Discussions

The devised methodology, which leverages Machine Learning, is executed through Python. The effectiveness of the resultant predictive model is assessed by employing established evaluation metrics, including accuracy, F-measure, precision, and recall.In this approach, 16890 leaf images are utilized by considering 80% for training and 20% for testing for leaf identification system. Similarly for disease identification 80% data is used for training and 20% for testing. The proposed system can classify 35 Ayurveda spices and also classify 25 different disease based on the values of 35 symptoms. Table 4 shows performances of model by considering four different pre-trained model with highest accuracy gain by CNN i.e 81.28%. Table 5 shows performance of disease identification using seven pre-trained model with highest accuracy gain by KNN i.e 91.06% .After identification of disease our system will also generates the home remedies for that particular disease.

Table IV: four pre-trained models for leaf identification

Modes	Accuracy	Precision	Recall	F1-Score
CNN	0.812	0.812	0.812	0.812
VGG-16	0.615	0.615	0.615	0.615
MobileNet	0.816	0.816	0.816	0.816
Inception	0.812	0.824	0.812	0.810

Table V: Seven pre-trained models for leaf identification showing training and validation accuracy

Sr.No	Classifier	Accuracy	Validation accuracy
1	MNB	84.95	84.50
2	RF	89.70	86.80
3	KNN	91.06	85.25
4	LR	90.16	89.19
5	SVM	89.48	88.62
6	DT	90.95	83.62
7	MLP	90.38	86.61

The output for leaf and disease identification on anroid based platform is depicted in

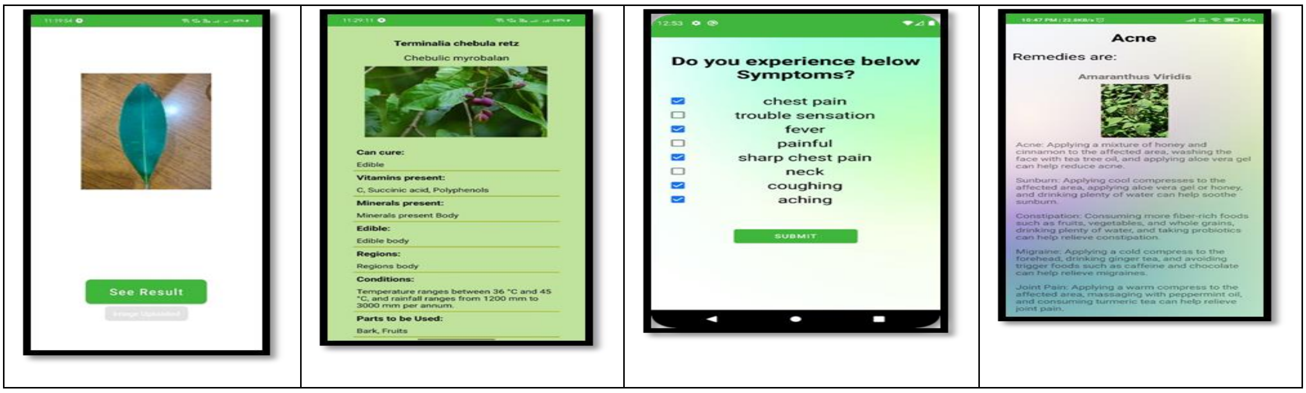


Fig 3.Screen shots for leaf and disease identification system.

6. Conclusion

Disease prediction by symptoms using machine learning is an exciting and rapidly evolving field in healthcare. While there are still many challenges to be overcome, the potential benefits of this approach are enormous. As the field continues to advance, we can expect to see even more accurate, reliable, and personalized disease predictions that can help healthcare professionals to improve patient outcomes and ultimately save lives. This research work is an application which can diagnose disease based on symptoms of patients and recommend ayurvedic remedies for the disease and also identify the type and use of ayurvedic leaf..

References

[1] Pulok Kumar Mukherjee, Subhadip Banerjee, Barun Das Gupta, Amit Kar,Chapter 1 - Evidence-based validation of herbal medicine: Translational approach,Editor(s): Pulok K. Mukherjee,Evidence-Based Validation of Herbal Medicine (Second Edition),Elsevier,2022,Pages 1-41,ISBN 780323855426.

[2] Pushpa BR, Anand C and Mithun Nambiar P" Ayurvedic Plant Species Recognition using Statistical Parameters on Leaf Images": International Journal of Applied Engineering Research ISSN 2016.

[3] S. U. Habiba, M. K. Islam and S. M. M. Ahsan, "Bangladeshi Plant Recognition using Deep Learning based Leaf Classification," 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), 2019, pp. 1-4, doi: 10.1109/IC4ME247184.2019.9036515.

[4] Review of Machine Learning Herbal Plant Recognition System by Priya Pinder Kaur, Sukhdev Singh, Monika Pathak in 2020

[5] Sethulekshmi A V and Sreekumar K "Ayurvedic leaf recognition for Plant Classification" International Journal of Computer Science and Information Technologies in 2014.

[6] Ahila Priyadharshini, R., Arivazhagan, S., Arun, M. (2021 "Ayurvedic Medicinal Plants Identification: A Comparative Study on Feature Extraction Methods" .Computer Vision and Image Processing. CVIP 2020. Communications in Computer and Information Science, vol 1377. Springer, Singapore. https://doi.org/10.1007/978-981-16-1092-9_23

[7] M. Bandara and L. Ranathunga, "Texture Dominant Approach for Identifying Ayurveda Herbal Species using Flowers," 2019 Moratuwa Engineering Research Conference (MERCon), 2019, pp. 117-122, doi: 10.1109/MERCon.2019.8818944.

[8] M. Raghukumar and G. Narayanan, "Comparison Of Machine Learning Algorithms For Detection Of Medicinal Plants," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 56-60, doi: 10.1109/ICCMC48092.2020.ICCMC-00010.

[9] C K Gomathy. "The Prediction of disease using machine learning". https://www.researchgate.net/publication/357449131_THE_PREDICTION_OF_DISEASE_USING_MACHINE_LEARNING

[10] Gomathy, C K. (2021). the prediction of disease using machine learning..https://www.researchgate.net/publication/357449131

[11] Maniruzzaman, M., Rahman, M.J., Ahammed, B. et al. Classification and prediction of diabetes disease using machine learning paradigm. Health Inf Sci Syst 8, 7 (2020). https://doi.org/10.1007/s13755-019-0095-z

[12] M. Chen, Y. Hao, K. Hwang, L. Wang and L. Wang, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities," in IEEE

Access, vol. 5, pp. 8869-8879, 2017, doi: 10.1109/ACCESS.2017.2694446.

[13] Amin Ul Haq, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir, Ruinan Sun, "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", Mobile Information Systems, vol. 2018, Article ID 3860146, 21 pages, 2018. https://doi.org/10.1155/2018/3860146

[14] S. G. Kale et al., "Identification of Ayurvedic Leaves using Deep Learning," 2023 International Conference on Communication, Circuits, and Systems (IC3S), BHUBANESWAR, India, 2023, pp. 1-6, doi: 10.1109/IC3S57698.2023.10169388.

[15] Tonk, A., Dhabliya, D., Sheril, S., Abbas, A.H.R., Dilsora, A. Intelligent Robotics: Navigation, Planning, and Human-Robot Interaction (2023) E3S Web of Conferences, 399, art. no. 04044, .