Irish Interdisciplinary Journal of Science & Research (IIJSR)
Vol.5, Iss.4, Pages 30-38, October-December 2021

An Efficient and Novel Crop Yield Prediction Method using Machine Learning Algorithm

Ashok Mishra

School of Computing, SRM Institute of Science and Technology, Tamil Nadu, India.
Email: ashokmishra_1982@gmail.com

Article Received: 29 September 2021 *Article Accepted: 27 November 2021* *Article Published: 23 December 2021*

ABSTRACT

The process of examining, filtering, and presenting data to obtain valuable information and make decisions is known as information analysis. Food resources are in high demand in countries like India, where they serve the population and help to secure the nation's security. Crop production is largely influenced by weather variations, soil quality, water availability, and fertilizer application, among other factors. The various types of soil play a significant effect in agricultural production. Recommending fertilizers to agriculturists may assist them in making better crop selection and maintenance decisions. Crop yield prediction can be done using a variety of studies using information and communication technology (ICT). Different sorts of mining techniques for data analysis and data acquisition can be widely used for a variety of purposes. Smart agriculture is a method of transmitting data from average farmers to skilled farmers.

Keywords: *Crop yield prediction, Random forest algorithm, Data preprocessing, Data mining, Environmental factors.*

1. Introduction

The most important sector of the Indian economy is agriculture. Agriculture in India accounts for 18 percent of the country's GDP and employs half of the population. However, recent research has revealed a steady drop in agriculture's contribution to the Indian economy, even though agriculture is India's most populous monetary zone and plays a substantial part in the country's usual socio-economic material [1]-[4]. Most farmers rely on their long-term experience in the arena on specific plants to predict a higher harvest in the following harvesting season, yet they still don't earn a fair value for their crops.

It occurs primarily as a result of improper irrigation or vegetation selection, or when crop yields are significantly lower than predicted [5],[6]. The majority of agricultural research relies on organic processes to predict crop growth and increase production, as agricultural researchers emphasize the necessity for a natural method to predict and improve crop growth. Crop output is mostly affected by crop variety, seed type, and environmental conditions like temperature, soil pH, water pH, rainfall, and humidity [8]-[10]. A high-quality crop may be cultivated with a greater crop yield and the entire agricultural output can be projected by researching the soil and environment in a specific area. Farmers will benefit from this forecast. Based on soil type, temperature, humidity, water level, spacing depth, soil PH, season, fertilizer, and months, select outstanding plants for their farm.

Data analysis is a tool for cleaning and modeling data to extract meaningful statistics and conclusions. It's a technique for analyzing, extracting, and forecasting significant data from enormous datasets to discover patterns [11]-[15]. This strategy is used by businesses to turn raw data from clients into usable data. Agriculture is also affected by this assessment [16]. Most farmers rely on their long-term experiences in the area with certain plants to forecast a larger yield in the upcoming harvesting season, but they still don't obtain a decent return on their crop investment [17]. It frequently occurs as a consequence of poor irrigation or plant selection, or when crop yields are low [18].

To estimate the yields of several plants in various states using a variety of technological factors and a newly constructed climate index as inputs. The regression and coefficient of dedication assessments, as well as the Average Error price, were used to create a reasonable contrast between our correct result, which is known as a goal, and the prediction model, which is a pleasant interface for farmers and provides the evaluation of rice manufacturing-based entirely on accessible data [19]. To optimize crop productivity, many data mining strategies have been employed to estimate crop yield. Accurate and timely monitoring of agricultural crop conditions, as well as estimating achievable crop yields, is critical operational approaches. Because crop yield prediction is so important, the goal of this lesson is to explore a variety of forecasting methodologies for evaluating crop yield estimates [20].

2. Related Works

Countless local weather elements have an impact on farm manufacturing. Precipitation parameters (rainfall, location-aware rainfall, irrigation, etc.) and soil parameters (humidity, wind speed, temperature, and moisture) are examples of metrological parameters (PH, natural carbon, phosphorus, fiber, etc.). And everything is fouled up due to the continual change in local weather conditions [21],[22].

Farmers in India continue to use the traditional technology that their forefathers taught them. However, the problem is that at the outset, when the weather was healthy, everything happened on schedule [23]. However, due to global warming and a variety of other factors, most of the issues have changed. The fundamental hassle with agriculture in India is the lack of rainfall in seasonal times [24]-[28]. Humidity is additionally critical for crops, however, it has been excessive, and it additionally converts as a drawback. The winter season is been affected so Rabi plants are extensively affected. For a few years, the rainfall in the wintry weather season used to be excessive as expected [29],[30]. To overcome these above issues, we want to enhance a gadget that will in a position to locate the hidden information or results, patterns, and insights [31]. The farmer can predict which crop ought to sow so that he/she can get extra benefit. In the proposed gadget we are making use of statistical analytics methods on agriculture production- primarily based datasets and discover the insights so that it can assist the farmers and their selection making [32].

2.1. Objectives

[1] Provide the farmer with the yield of a crop based on land area, rainfall, temperature, and district using machine learning.

[2] Predict the future market price of crops by taking previous crop prices and predicted yield data into consideration.

[3] Compare the predicted result of the different algorithms and determine which approach is more suitable.

Handling in The Laboratory: As quickly as the samples are acquired at the soil checking out laboratory, they have to be checked with the accompanying records list. If the soil trying out laboratory personnel have accrued the samples themselves [33],[34]. Then the sufficient subject notes would possibly have been kept. All unidentified samples ought to be discarded. *Drying of Samples:* Samples acquired in the laboratory can

also be moist. These must be taken to ensure that each pattern is identified at all phases of the preparation process [35-39]. After drying, the samples are moved to a separate practice room from the main laboratory. The dirt mixture is beaten using a wooden pestle and mortar, but the soil particles are not decomposed.

3. Proposed Methodology

The proposed crop yield forecast methodology is available in this area. Crop yield prediction is used to forecast manufacturing in the agriculture sector to improve crop management and make strategic choices in the future. The existing mannequin can be combined with a decision support system (DSS) for precision agriculture. Predicting crop yields accurately in the development stage before harvest can help farmers and government agencies make better decisions about storage, selling, setting a minimal aid price, and importing/exporting.

Price prediction is surprisingly beneficial in agriculture for forecasting the market rate for the respective commodities and additionally beneficial for farmers to graph their crop cultivation things to do so that they ought to fetch extra fees in the market. Middle guys and brokers can be eradicated as the farmer himself is aware of the market charge beforehand. Consumers can use this fee prediction for their everyday way of life planning.

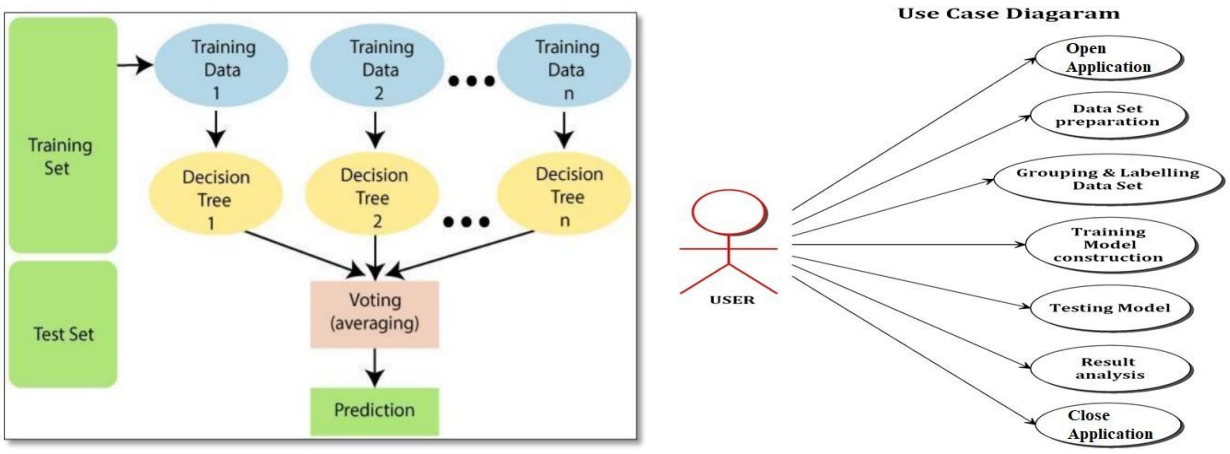

Fig.1. System architecture of Crop Yield prediction method

Since the dataset, we have taken is in unstructured form (noisy). Hence we have to bear a positive method referred to as facts pre-processing. Data pre-processing is a step the place the unstructured information is converted into a structured shape and feeder into the algorithm. As the records are structured, it is given as a Backpropagation and Random forest. Back Propagation algorithm is used to correctly instruct a neural community through an approach referred to as chain rule. Backpropagation executes a backward omission while modifying the model's parameters after each forward pass over a network (weights and biases). The supervised learning method Random Forest is well-known. It can be used to solve any classification or regression problem in machine learning. It is entirely based on ensemble learning theory, a way for combining numerous classifiers to tackle a complex problem and increase the model's overall performance. "Random Forest is a classifier that incorporates a range of decision timber on several subsets of the provided dataset and takes the common to improve the predicted accuracy of that dataset," as the

title suggests. Rather than depending on a single choice tree, the random forest collects predictions from all trees and forecasts the final result based only on the majority votes of predictions. The more bushes in the wooded region, the greater the accuracy and the less danger of overfitting. Both the above algorithms provide their character accuracy on the groundwork of their respective process. Finally, the accuracy of each algorithm is in contrast and the algorithm with the absolute best accuracy is selected.

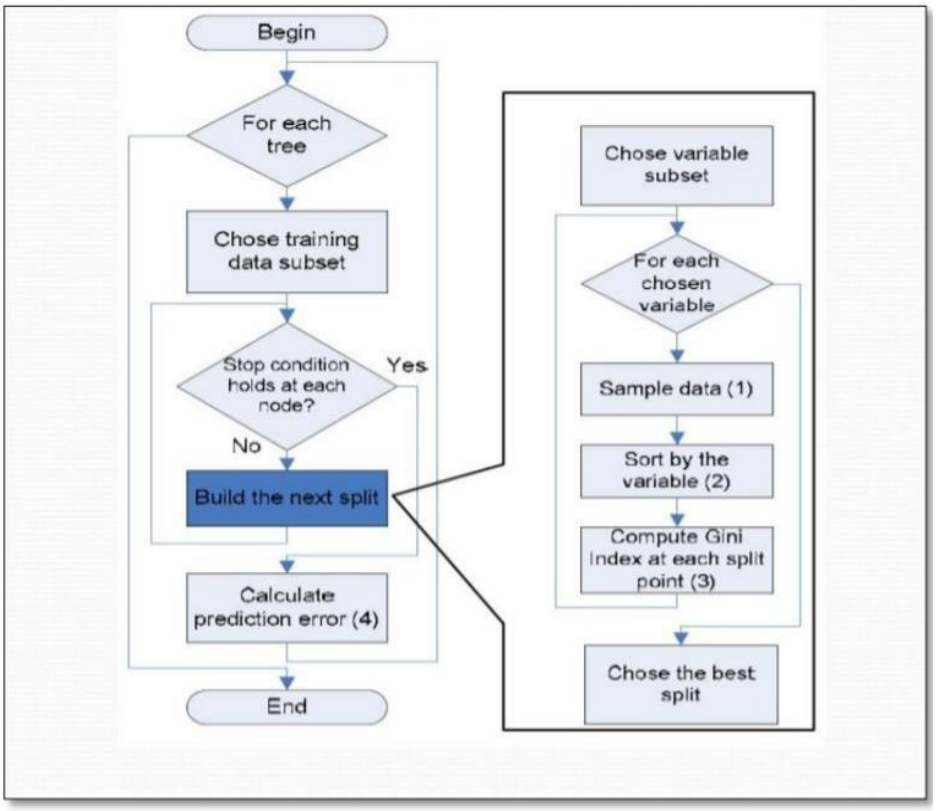

Fig.2. Flow diagram of the proposed method implementation

3.1. Fertilizer Prediction

In fertilizer prediction, we are offering soil parameters as a dataset from the Google app i.e the neck (nitrogen, phosphorus, potassium) value. These parameters are fed to the Backpropagation algorithm which produces the Fertilizer prediction.

3.2. Random forest algorithm

From a randomly selected subset of the training set, the random forest classifier builds a set of decision trees. The final class of the test item is calculated by combining the votes from various decision trees. Random Forest improves the algorithm's prediction power while reducing overfitting. The simplest and most used algorithm is random forest. Classification and regression are both possible with this method. It's a collection of random decision trees put together as a whole.

3.3. Working of random forest algorithm

The steps involved in implementing the Random Forest algorithm are

Step 1: Select the samples randomly from the data.

Step 2: The decision tree is created for each sample, and then the forecast result is obtained.

Step 3: Appointed by each anticipated outcome in this step.

Step 4: Select the forecast result with the most views as the final estimate result.

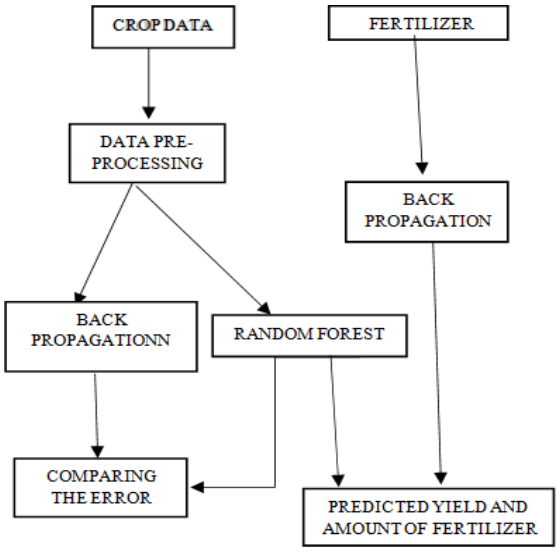

Fig.3. Flow diagram of the Random forest algorithm

4. Simulation Results and Discussion

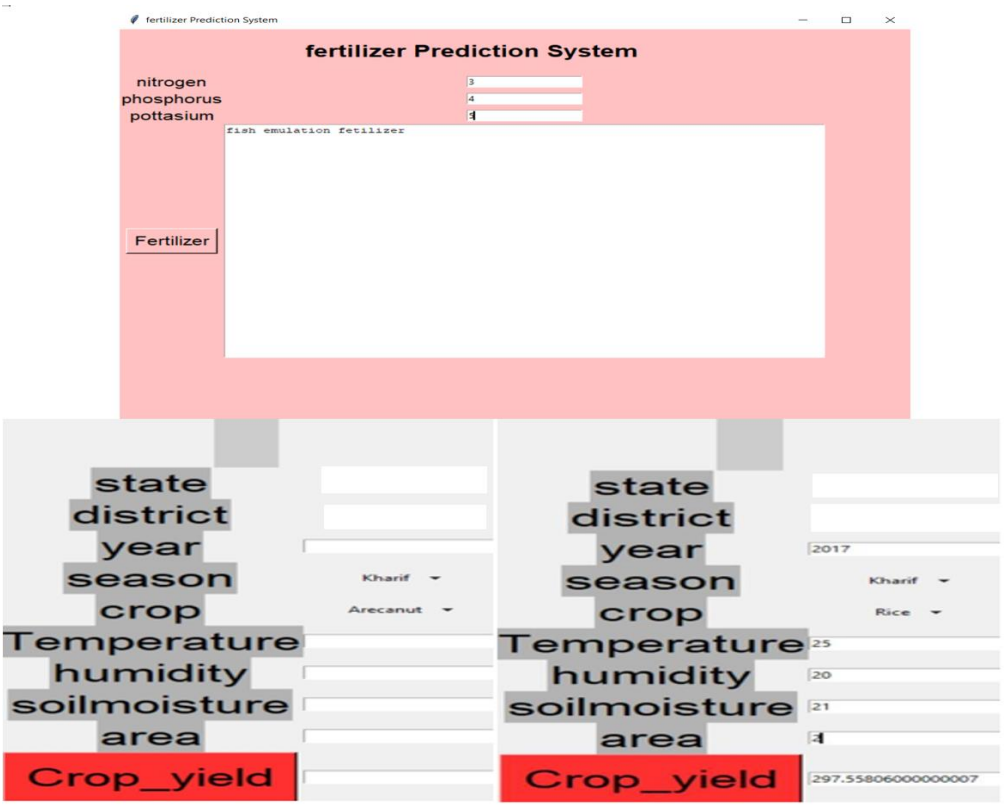

Fig.4. Snapshot result of proposed Fertilizer Prediction system

This research proposes a system that will assist farmers in estimating yields based on meteorological factors and cultivated areas. With this information, a farmer can decide whether to produce a specific crop or switch

to a different crop if yield projections aren't good. By which farmers can make the decision and they can get more benefit.

5. Conclusion

In countries like India, agriculture is the backbone. However, as we get closer to eliminating agriculture, the use of technological know-how in the direction of agriculture must be prioritized. This research presents a machine that would aid farmers in estimating yields based on climate characteristics and the location of crops under cultivation. Using this information, a farmer can decide whether to cultivate a specific crop or switch to a different crop if yield projections aren't favorable. This lookup work can be bettering to the subsequent level. We can construct a recommender machine of agriculture manufacturing and distribution for the farmer. By which farmers can make a selection in which season which crop have to sow so that they can get greater benefit. This gadget works for the structured dataset. In the future, we can put in force records impartial gadgets also. It is the ability layout of statistics whatever, our device must work with equal efficiency.

Declarations

Source of Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing Interests Statement

The author declares no competing financial, professional and personal interests.

Consent for publication

Author declares that he/she consented for the publication of this research work.

References

[1] Ponce-Guevara, K. L., Palacios-Echeverria, J. A., Maya-Olalla, E.,DominguezLimaico, H. M., Suarez-Zambrano, L. E., Rosero-Montalvo, P.D., Alvarado-Perez, J. C. (2017). GreenFarm-DM: A tool for analyzing vegetable crops data from a greenhouse using data mining techniques (First trial). 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM).

[2] Venkatesan, C., Karthigaikumar, P., & Varatharajan, R. (2018). FPGA implementation of modified Error Normalized LMS adaptive filter for ECG noise removal. Cluster Computing, 22(S5), 12233–12241. https://doi.org/10.1007/s10586-017-1602-0.

[3] Shakoor, M. T., Rahman, K., Rayta, S. N., Chakrabarty, A. (2017). Agricultural production output prediction using Supervised Machine Learning techniques.2017 1st International Conference on Next Generation Computing Applications (NextComp).

[4] T.Thamaraimanalan, S.P.Vivekk, G.Satheeshkumar and P.Saravanan (2018), Smart Garden Monitoring System Using IOT, Asian Journal of Applied Science and Technology, Vol 2 (2), Pg 186-192.

[5] Rajagopal, L. (2016). Power and area efficient decimation filter architectures of wireless receivers. Proceedings of the National Academy of Sciences, India Section A: Physical Sciences, 87(1), 83–96. https://doi.org/10.1007/s40010-016-0314-9.

[6] Venkatesan, C., Karthigaikumar, P., & Satheeskumaran, S. (2018). Mobile cloud computing for ECG telemonitoring and real-time coronary heart disease risk detection. Biomedical Signal Processing and Control, 44, 138–145. https://doi.org/10.1016/j.bspc.2018.04.013.

[7] Niranjan, L., Venkatesan, C., Suhas, A. R., Satheeskumaran, S., & Nawaz, S. A. (2021). Design and implementation of chicken egg incubator for hatching using IoT. International Journal of Computational Science and Engineering, 24(4), 363. https://doi.org/10.1504/ijcse.2021.117018.

[8] Arun, T. V., Lathesh, L., & Suhas, A. R. (2016). Substation automation system. International Journal of Scientific & Engineering Research, 7(5), 215–218.

[9] Niranjan, L., Suhas, A. R., & Chandrakumar, H. S. (2018). Design and implementation of self-balanced robot using proteus design tool and arduino-uno. Indian J. Sci. Res, 17(2), 556–563.

[10] Jubairahmed, L., Satheeskumaran, S., & Venkatesan, C. (2017). Contourlet transform based adaptive nonlinear Diffusion filtering for speckle noise removal in ultrasound images. Cluster Computing, 22(S5), 11237–11246. https://doi.org/10.1007/s10586-017-1370-x.

[11] Thamaraimanalan, T., RA, L., & RM, K. (2021). Multi Biometric Authentication using SVM and ANN Classifiers. Irish Interdisciplinary Journal of Science & Research (IIJSR).

[12] N, S., L, N., N, G., Jahagirdar, S., R, S. A., & N, S. (2021). Efficient usage of water for smart irrigation system using Arduino and proteus design tool. 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), 54–61. https://doi.org/10.1109/icosec51865.2021.9591709.

[13] Asha, K. K., & Suhas, A. R. (2015). Multi-Stage Auv-Based Localization for Underwater Wireless Sensor Network. JEST-M, 4(4), 14–19.

[14] L , N., B , S., & A R, S. (2014). Advanced System for Driving Assistance in Multi-Zones using RF and GSM Technology. International Journal of Engineering Research & Technology (IJERT), 3(6), 1884–1889.

[15] Venkatesan, C., Karthigaikumar, P., & Varatharajan, R. (2018). A novel LMS algorithm for ECG Signal preprocessing and KNN classifier Based abnormality detection. Multimedia Tools and Applications, 77(8), 10365–10374. https://doi.org/10.1007/s11042-018-5762-6.

[16] Suhas, A. R., & Viswanath, H. L. (2014). A Comparative Study between Extended Kalman Filter and Unscented Kalman Filter for Traffic State Estimation. International Journal of Innovations in Engineering and Technology (IJIET), 4(1), 188–197.

[17] Thamaraimanalan T, Sampath P (2019) A low power fuzzy logic based variable resolution ADC for wireless ECG monitoring systems. Cogn Syst Res 57:236–245. https://doi.org/10.1016/j.cogsys.2018.10.033.

[18] Balachander, K., Venkatesan, C. and R., K. (2021), "Safety driven intelligent autonomous vehicle for smart cities using IoT", International Journal of Pervasive Computing and Communications, Vol. 17 No. 5, pp. 563-582. https://doi.org/10.1108/IJPCC-10-2020-0182.

[19] L., N., & Priyatham M., M. (2021). Lifetime ratio improvement technique using special fixed sensing points in Wireless Sensor Network. International Journal of Pervasive Computing and Communications, ahead-of-print(ahead-of-print). https://doi.org/10.1108/ijpcc-10-2020-0165.

[20] Niranjan, L., Suhas, A. R., & Sreekanth, B. (2018). Design and implementation of robotic arm using proteus design tool and arduino-uno. Indian J Sci Res, 17(2), 126–131.

[21] Chellappan, R., Satheeskumaran, S., Venkatesan, C., & Saravanan, S. (2021). Discrete stationary wavelet transform and SVD-based digital image watermarking for improved security. International Journal of Computational Science and Engineering, 24(4), 354. https://doi.org/10.1504/ijcse.2021.117016.

[22] T. Thamaraimanalan and P. Sampath, "Leakage Power Reduction in Deep Submicron VLSI Circuits using Delay based Power Gating", National Academy Science Letters, Vol. 43, No. 3, pp. 229-232, 2019.

[23] L., N., & Priyatham M., M. (2019). An energy efficient and lifetime ratio improvement methods based on energy balancing. International Journal of Engineering and Advanced Technology, 9(1S6), 52–61. https://doi.org/10.35940/ijeat.a1012.1291s619.

[24] Venkatesan, C., & Karthigaikumar, P. (2018). An efficient noise removal technique using modified error normalized LMS algorithm. National Academy Science Letters, 41(3), 155–159. https://doi.org/10.1007/s40009-018-0635-0.

[25] AR, S., & Priyatham M., M. (2019). Lifetime and energy efficiency improvement techniques for hierarchical networks. International Journal of Engineering and Advanced Technology, 9 (1S6), 62–72. https://doi.org/10.35940/ijeat.a1013.1291s619.

[26] T, Thamaraimanalan & Mohankumar, M. & Dhanasekaran, S. & Haldorai, Anandakumar. (2021). Experimental analysis of intelligent vehicle monitoring system using Internet of Things (IoT). EAI Endorsed Transactions on Energy Web. 8. 169336. 10.4108/eai.16-4-2021.169336.

[27] AR, S., & Priyatham M., M. (2021). Heal nodes specification improvement using modified chef method for group Based Detection Point Network. International Journal of Pervasive Computing and Communications, ahead-of-print (ahead-of-print). https://doi.org/10.1108/ijpcc-10-2020-0170.

[28] K Kavitha, T Thamaraimanalan, M Suresh Kumar (2014), An Optimized Heal Algorithm for Hole Detection and Healing in Wireless Sensor Networks International Journal of Advanced Engineering Research and Technology (IJAERT), Vol 2, Issue 3, pp 243-249.

[29] Venkatesan, C., Saravanan, S., & Satheeskumaran, S. (2021). Real-time ECG Signal pre-processing and Neuro fuzzy-based CHD risk prediction. International Journal of Computational Science and Engineering, 24(4), 323. https://doi.org/10.1504/ijcse.2021.10039962.

[30] T.Thamaraimanalan, D.Naveena, M.Ramya & M.Madhubala (2020), "Prediction and Classification of Fouls in Soccer Game using Deep Learning", Irish Interdisciplinary Journal of Science & Research (IIJSR) Vol.4, Iss.3, Pages 66-78.

[31] Venkatesan, C., Karthigaikumar, P., Paul, A., Satheeskumaran, S., & Kumar, R. (2018). ECG signal preprocessing and SVM Classifier-based abnormality detection in Remote healthcare applications. IEEE Access, 6, 9767–9773. https://doi.org/10.1109/access.2018.2794346.

[32] R. Santhiya and M. T. Thamaraimanalan, "Power Gating Based Low Power 32 Bit BCD Adder using DVT," Int. J. Sci. Res. Dev., vol. 3, no. 02, pp. 802–805, 2015.

[33] Latha, R., & Vanathi, P. T. (2015). Design of digital filters for multi-standard transceivers. International Journal on Electrical Engineering and Informatics, 7(3), 517–530. https://doi.org/10.15676/ijeei.2015.7.3.13.

[34] Madasamy Raja, G., Thaha, M., Latha, R., & Karthikeyan, A. (2019). Texture classification using optimized local ternary patterns with nonlinear diffusion as pre-processing. Multimedia Tools and Applications, 79(5-6), 3831–3846. https://doi.org/10.1007/s11042-019-7197-0.

[35] Mohanapriya, C., Bharathi, K. M., Aravinth, S. S., Gowrishankar, J., & Ramkumar, M. (2014). A trusted data governance model for big data analytics. International Journal for Innovative Research in Science and Technology, 1(7).

[36] Narmadha, B., Ramkumar, M., Vengatesan, K., & Srinivasan, M. (2017). Household Safety based on IoT. International Journal of Engineering Development and Research (IJEDR), ISSN, 2321-9939.

[37] Sara, S. B. V., Anand, M., Priscila, S. S., Yuvaraj, N., Manikandan, R., & Ramkumar, M. (2021). Design of autonomous production using deep neural network for a complex job. Materials Today: Proceedings.

[38] Ramkumar, M., Manikandan, M., Kumar, K. S., & Kumar, R. K. (2019). Intrusion detection in manets using support vector machine with ant colony optimization. ICTACT journals on data science and machine learning, 1(1).

[39] Vengatesan, K., Padmanaban, S., Singh, R. P., Ravishankar, T. N., Bhaskar, M. S., & Ramkumar, M. (2017, July). Performance analysis of gene expression data using biclustering iterative signature algorithm. In 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (pp. 7-11). IEEE.