2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) | 979-8-3503-9926-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICACITE57410.2023.10182502

# Integration of Cloud Computing, IoT, and Big Data for the Development of a Novel Smart Agriculture Model

Molli Srinivasa Rao
*Department of CSE, Dadi Institute of Engineering & Technology*
Anakapalle, Visakhapatnam, India
drmollisrinivasarao@gmail.com

Sanjay Modi
*Lovely Professional University*
Phagwara, Punjab, India
sanjaymodi2234lpu@gmail.com

Rajat Singh
*Uttaranchal Institute of Technology Uttaranchal University*
India
rajat@uumail.in

K. Lakshmi Prasanna
*School of Agriculture, SR University,*
Warangal, Telangana.
k.lakshmiprasanna@sru.edu.in

Shakir Khan
*College of Computer and Information Sciences,*
*Imam Mohammad Ibn Saud Islamic University (IMSIU),*
Riyadh, Saudi Arabia
*University Centre for Research and Development,*
*Department of Computer Science and Engineering,*
Chandigarh University, Mohali 140413, India
shakirkhan2006@gmail.com
Orcid id 0000-0002-7925-9191

C. Ushapriya
*Assistant professor,*
*Department of CSE,*
*St. Martin's Engineering College,*
Secunderabad,
Telangana, India
cushapriyacse@smec.ac.in

***Abstract*-Since the traditional data paradigm cannot handle the volume of information generated by IoT (Internet - of - things) gadgets, cloud storage is now required. These data have been examined using big mining techniques. When evaluating the viability of smart agriculture, the Internet of Farmers must use technologies of information and communication (ICT) in their daily lives to get agricultural information. Crop growth observing, fertilizer classification and irrigation support systems use IoT. This article investigates and optimizes the large amounts of data produced during the farming process, but it seems to be analyzing data mining using the automated k-means system according to the maximum speed. The crop growth curve is intended to simulate the earliest K-means methodology. The experimental findings support the idea that clustering algorithms provide overall benefits in the F mutual information of 7.67% and just a decrease in the total period of 0.23 seconds is supported. Because it can more efficiently realize the operational processes of reliable information communication and data processing, the algorithm presented in this article has a major effect on advancing agronomic information technology.**

***Keywords: Cloud computing, IoT, Big Data, Agriculture***

## I. Introduction

Using cutting-edge latest technology like IoT devices, big data, the platform as a service, and mobile computing, many applications have been developed in recent years. At this time, the world is heading toward concepts like smartphones, intelligent buildings, and smart homes. Agriculture plays a significant role in India's economy. For the vast majority of Indian communities, it is still a way of life [1]. Ineffective crop maintenance results in a huge loss for just one farmer, which is why the concept of "smart agriculture" had been initiated. Precision agriculture increases crop production by utilizing cutting-edge technologies. It aids in system monitoring and provides data about various environmental aspects.

Monitoring the environmental factors is neither an adequate nor comprehensive method of boosting crop yield. Productivity is also greatly influenced by many other factors. One of these factors is insect and pest attacks by applying the appropriate pesticides and insecticides to the crop, which can be controlled. As soon as the crops are prepared for harvest, birds and certain other wild animals start stealing them. As a result, farmers encounter numerous difficulties during the planting and harvesting processes. All issues can be resolved by taking into account every aspect and building an integrated system with the aid of the technologies listed below.

### *A. IoT (Internet - of -Things)*

In the modern world, the Internet is crucial for all industries. The suggested approach is used in the arable domains to use the concept of the Internet of Things to screen the agricultural areas. Using sensor network wireless technology, sensors are used in the agricultural sector to analyze various parameters. The proposed method is used to gather the characteristics of the soil, which are then saved in a cloud server [2].

The amount and variety of agricultural data have grown alarmingly as a result of IOT's emergence in the agricultural sector. Agriculture's storage and analysis issues can be resolved using big data technologies as well as cloud computing. IoT has made it possible to apply agricultural production to big data in ever-deeper ways, and big data analysis as well provides technical support for agricultural computer technology. This allows farmers to continue to perceive as well as monitor the optimal time for planting crops. Big data is regarded as a breakthrough in the field of modern information technology, notably in the Iot and cloud, as it is the slashing technique for smart applications as well as mining. Big data background information and its internet connections of things were provided by ZF Sun et al (IoT), highlighted key technologies, examined the requirements for big data across the major application domains, and presented some recommendations and concepts for further growth in big data as well as precision farming. Big data-based solutions could increase supply chain efficiency and agricultural productivity.

Resource sharing is offered by cloud computing at a price. The services are provided by providers of cloud computing at a reasonable price. It has served as a storage location for agricultural data. Along with IoT, it is utilized in the agricultural sector [3]. Big data is a sizable method of collecting data that has been amassed over time from many sources, such as sensor data, network security data, but also business data. Gathering, storing, analyzing, as well as searching the data is the main challenge. It is employed in business intelligence, data processing, and big data analytics to look for patterns in the data that are not yet known to science. Big data is utilized in the agricultural industry to monitor the entire chain of supply of goods and cut costs. Big data analytics is a method for fusing traditional and commercial analytics [4].

Techniques are applied to data analysis and prediction. The methods include association rule mining, clustering algorithms, and classification algorithms. A few good examples of algorithms include decision trees, k-nearest neighbor algorithms, C4.5, RepTree, J48, SVM, computer vision, naive Bayes K-means techniques for classification priority algorithms, as well as Fpgrowth algorithms. It has been used to categorize different types of soil by looking at their characteristics. It provides flexibility and is useful for both monitoring field data and controlling field operations. Big data technologies include Hdfs, Hadoop, Stream Processing, Pig, Apache Hadoop, Typhoons, Wild Mustang, and others. There is a vast and diverse amount of data available in the agricultural industry [5]. It has been suggested to use a mobile computing system to send farmers daily, seasonal messages with product and weather knowledge [6]. It was properly determined with an importance placed on information. The main focus is on the agricultural industry and employs a novel methodology to learn about design, conceptual framework, development, and evaluation.

## II. Literature Survey

In the field of agriculture, system models have a significant impact on the emergence of agroecological and socioeconomic conditions. We can determine appropriate and effective management strategies and provide information by making use of the field as well as farm experiments. It can assist with determining the monitoring to land users all through space and time as long as the required soil, strategic planning, climatic factors, and socioeconomic status information is available. Data are created using decision-support systems (DSSs) for the farm as well as pest management. Modern techniques are not being used in these processes to process the data. Therefore, use intelligent technique concepts to decide how to solve the issue.

It is crucial to the understanding of modern agricultural results, and more farmers are using them as decision-support tools. A brand-new method for categorizing and forecasting the weather has been developed by Sanjay D.Sawaitul and colleagues. The technique keeps track of weather variables like airspeed, wind direction, rainfall, temp, and climatic forecasting. The backpropagation algorithm for artificial neural networks is used to forecast the weather. For weather prediction, the author tested three models. The first framework is used to compile the weather forecasting techniques. Utilizing the third template, the Rear Algorithm is applied to various forecasting parameters, and also the new model is then used to start introducing the WSN toolbox for the collection of data. [7].

Authors in [8] proposed a method for crop yield prediction using fuzzified set theory and probabilistic reasoning. The climate conditions are used to predict crop yields based on the farmer's experience. The research project on data mining techniques for crop yield analysis in agriculture has been proposed by Ramesh D. et al. From 1965 to 2009, they gathered data in India's Andhra Pradesh region's East Godavari district. The information includes the year, rainfall, sowing area, and production. The crop yield is predicted based on rainfall using multiple regression (MLR) as well as K-Means techniques. This prototype is used to forecast yield production with high generality and high accuracy [9].

Using a decision tree algorithm, the author [10] projected her research findings in the agricultural sector. This methodology is used to boost the soybean crop's productivity. They gathered information on the climatic factors during the rabi and kharif seasons in the Bhopal District of Madhya Pradesh State. Predicated here on the Navie Bayesian classification method, they predict the paddy crop yield using the temperature and rainfall variables. Based here on climate conditions, the economic output of wheat production is determined.

Wen-Yaw Chung and colleagues have introduced their work in the area of agriculture that combines wireless sensor technology and the cloud. Temperature, humidity, pH level, and other information are monitored and collected in the agricultural environment using sensors. The sensors are employed to collect data much more quickly. The system consists of hardware as well as smart devices with adequate storage capacities that can be used to remotely access and monitor land information [11].

Duncan Waga and colleagues describe their work. Analytics in cloud computing is crucial to agriculture. The capability to access the relevant data is also very advantageous to the framers. They store and retrieve data using a private cloud. For effective data collection and aggregation, distributed, flexible services are used, such as HDFS. Performance, capacity, and scaling challenges are met when using the cloud as well as big data. In this work, parameters like winds, temperature, and rainfall are used. Hadoop packages must be implemented in the innovative IT platform [12].

The Internet of Things and cloud computing, according to Rao et al., collaborate to address the Big Data issue. They consider cloud-based sensing services, which have a few particular uses, like farming and environmental monitoring. To provide the sensed data as a cloud service, they propose a brand-new prototype model. Applications and services are increasingly able to interact with the real world thanks to wireless sensor networks. The sensing network and such services might be on different sides of the Internet. With the aid of big data technologies and cloud services, the data can be used for storing and analysis. It has the potential to increase cloud services' scalability and usability [13].

Rajesh et al., present their work in the area of agriculture. Cloud computing and sensor data were combined. Here sensor node is integrated and managed using the service-oriented architecture. The technology of cloud computing allows for the use of storing and making data accessible to users. Both the new app and the new database are available from them. The cloud computing model and the internet incorporate sensor networks. The sensor network is used to record industrial processes. The information gathered is crucial for the sector and provides the data quickly [14].

## III. Proposed Methodology

### A. *IoT system for agriculture*

The agricultural industry is one of the areas of IOT development that is most urgent and significant. The production, processing, as well as distribution of agricultural products, are all impacted by the IOT application system for agriculture. Agriculture has gradually transitioned from a human-centered production model that relied on isolated machinery to information- as well as a software-centered production model that uses a lot of automated, smart, wireless router production equipment thanks to IOT technology [15]. To achieve the maximum earnings of agricultural output, operators, growers, along with managers, the agricultural IOT system's data can be used to change the planting plans of farmers. IOT for agricultural output is integrated into a hierarchy organization at the business level. There are a total of 5 layers, from lowest to highest.

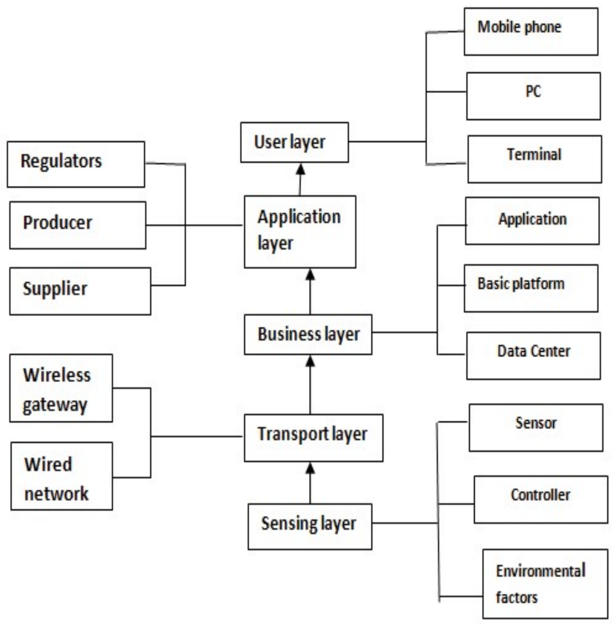


Fig. 1. Block diagram for Internet of Things Based Agricultural System

*1)* Sensor layers: It can gather fundamental data about the surroundings for growth, logistics, but also storage, including observing, trading, storage, but instead weather information.

*2)* Transport layers: Here communication's cabled and wireless components are combined at the transport layer. To ensure the consistency of data transmissions, the consistency of data, as well as the speed and simplicity of transmission and exchange, the backbone network mainly uses make a great contribution.

*3)* The business layer is used in quality certification bodies to track environmental variables, growth variables, and different kinds of data.

*4)* Application layers: Depending on the roles, it is split into various application portals., such as senior managers, producers, suppliers, broadcast providers, consumers, etc.

*5)* User layer: Using security systems and techniques to ensure secure access to that information as well as guard against malware attacks and data tampering. Figure 1 illustrates the fundamental structure.

### B. *Mining data*

Data mining is frequently used in business, manufacturing, research, and development, as research of sets of data, application value, as well as technical value in other industries [16].

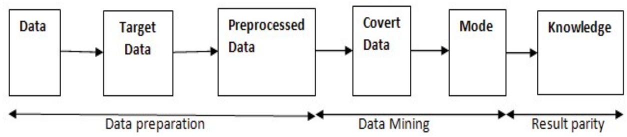


Fig. 2. Process flow of Data mining

*1) Establish the goals for the project.*
Setting goals will help you formulate your strategic goal more effectively and must consist of the project's final cost.

*2) Gathering of data.*
Data gathering involves gathering as well as analyzing relevant decisions making in a given situation [17]. Data collection methods include business surveys, network crawling, and data extraction from existing data systems.

*3) Processing data in advance.*
The collected data are "formatted" during this procedure in preparation for model training but instead statistics. The gathered Data that have not undergone data preprocessing and other steps may yield inaccurate results during the mining process.

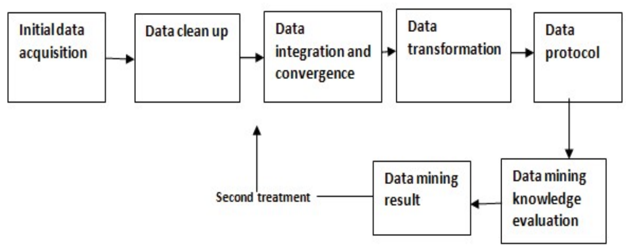


Fig. 3. Flow of Preprocessing

Data preprocessing should be responsible for more than 80% of the students who are working during the entire data mining process.

*4) Analyzing data*
The method of converting useless data into knowledge is known as data mining. Knowledge discovery is the process of identifying patterns in a sizable amount of information and using an analysis technique or an algorithm. Finally, through analysis as well as result comparison, we should select the most suitable mining algorithm.

*5) Model evaluation*

Model evaluation is an approach for the in-depth examination of data mining results. This project's main objective is to assess the data mining findings. Analyzing the root of the error is necessary if the data mining outcome significantly differs from the user's intended outcome [18]. The most popular and efficient technology in use today for processing agricultural information data is data mining. Increasing crop yield requires the use of data mining techniques to mine agro-based information data and improve crop productivity to give users timely agricultural planting information.

*6) Data mining software*

The primary goal of clustering analysis, which is a crucial part of data mining, is to organize various similar items into clusters or clusters by the attribute of the objects they describe, allowing for the placement of highly similar objects in the same clusters and the best possible differentiation between objects from different clusters. Data aggregation analysis is a common method for data reduction in statistical analysis.

### C. K-means algorithm

The K-means algorithm is a popular clustering algorithm because it is both effective and straightforward. Using a random number generator, determine the Euclidean distance between each sample point and the clustering center, and then use the nearest criterion to assign each sample point to the classes that share the most similarities with the clustering center [19]. The average values of all sample points within every category are computed while the segmentation core is refreshed until the impartial criterion function converges. This is the precise procedure:

- The database containing N objects and the K clusters' total number is the input data.
- K clusters in the output data, attempting to make the MSE stack this same smallest.

### D. Recommend Exercise

The chosen data are produced following the growers' data requirements in this article. The following are the primary requirements. By obtaining a large amount of scientific information, which includes soil temperature, ambient temperature, air humidity, and soil moisture, one can determine the growth environmental curve of young plants in each fixed cycle. The sensors measure the soil, air, heating rate, and some other relevant data in just a few minutes, producing a significant amount of data in just one hour. In this article, data mining is done using the k-means method of analysis based on the most significant distance. This is the precise procedure:

1. Select a point at random to act as the through one center
2. Constantly update the shortest route between clustering centers and all other points that aren't cluster centers.
3. If the amount of grouping anchor nodes is lower than k, move on to the following step; otherwise, the selection is complete.
4. Continue with the previous step after updating the least distance between the set of anchor nodes and all parallel nodes in the network, determining the greatest benefit inside the shortest path array, and designating the data points that correspond to the true worth as the centroids.

Compared to the K-means algorithm, a more careful selection of the nearest centroid is made, resulting in less iteration. The study's findings and analysis. In this article, MATLAB is used to compare the effectiveness of the K-means algorithm as well as the optimum range K-means algorithm. After the information is clustered that uses the K-means algorithm, quantities are selected to establish the clustering infrastructure for iteration. Then, K-means automated system is based on the best distance used to cluster data.

## IV. Result and Discussion

The F assessments of the interactions determine algorithms are compared in Figures 4 as well as in Figure 5. Comparing the improved K-means graph-based method to the existing K-means cluster analysis technology; the F measure is increased by 7.67%. While decreasing overall time consumption by an average of 0.23 seconds.

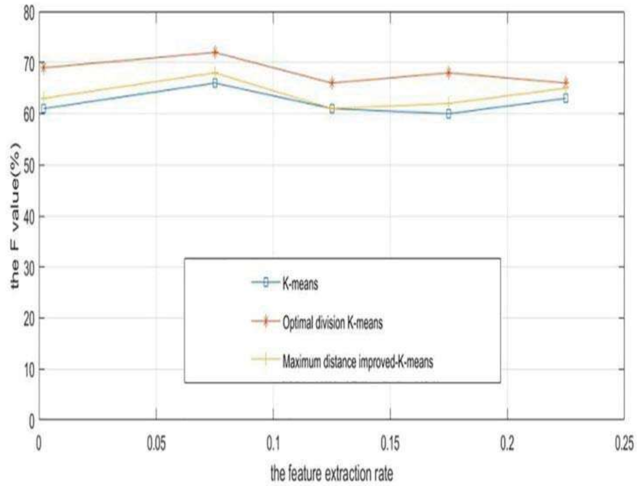


Fig. 4. Clustering algorithm F-value comparison

The aforementioned experimental results demonstrate that the improved algorithm has the best clustering effect as well as time performance.

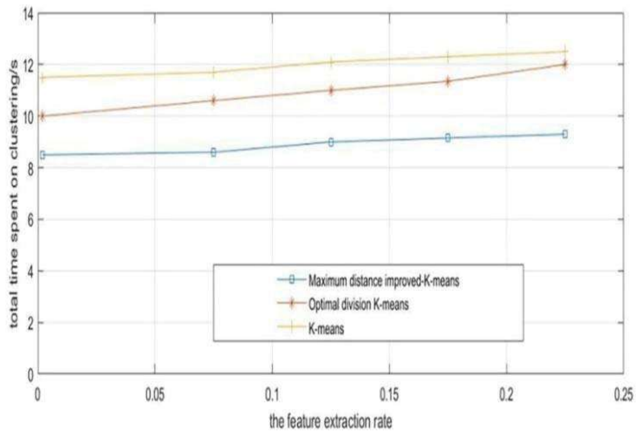


Fig. 5. Comparison in terms of time consumption

Data management and data mineral extraction modules served as the foundation for the data analysis and data quarrying modules used in the simulation experiments for this study. It was found that the initial cluster center was chosen at random, which accounts for the quick initialization speed of the K-means algorithm. However, because the K-means clustering algorithm selects cluster centers at random, two data that should only form one cluster can easily be mistakenly counted as two clusters, trying to increase the clustering algorithm's overall repetition count. As a result, clustering takes longer overall, and it is simple to reach a local optimum. The improved algorithm chooses a highly distinct initial cluster center. On the one hand, it minimizes the number of variations and prevents reaching a local optimum; on the other, it enhances the cluster's Fvalue and exhibits strong stability.

## V. Conclusion and Future Scope

This article uses IOT technology and data information extraction from big data to crop results, putting the focus on the corporation's priorities of modern agricultural development, to create a smart farming landscape which thus replicates the good control and management of agricultural greenhouses. In light of the thorough analysis and appreciation of sustainable farming utilizing the Internet of Things and big data technology, the K-means edge detection using the legal maximum method for selecting the initial cluster is proposed. Additionally, the experimental simulation is used to study and improve the data memory, processing, as well as mineral extraction in the agro-based big data generated by IOT technology.Based on a thorough analysis as well as consideration of agricultural production based on IOT as well Using big data, this same maximum allowable K-means onset procedure is proposed to select the ideal parameters. Through experimental simulation, the data manufacturing, storage, and mining in the IoT device’s big data for agricultural production are also analyzed and optimized. In the future, neural networks can be used to analyze.

## References


[1] M. Kaur, H. Gulati, and H. Kundra, "Data mining in agriculture on crop price prediction: Techniques and applications," International Journal of Computer Applications, vol. 99, no. 12, pp. 29-34, August 2014.

[2] S. H. Shah, F. Khan, W. Ali, and J. Khan, "A new framework to integrate wireless sensor networks with cloud computing," in IEEE Aerospace Conference, Big Sky, MT, USA, Mar. 2-9, 2013, pp. 1-8.

[3] V. C. Patil, K. A. Al-Gaadi, D. P. Biradar, and M. Rangaswamy, "Internet of Things (IoT) and cloud computing for agriculture: An overview," in Proceedings of AIPA 2012, India.

[4] F. Diebold, "Big data," Pier working paper archive, Penn Institute for Economic Research, 2012.

[5] D. Laney, "3-D data management: Controlling data volume, velocity, and variety," META Group Research Note, 2001.

[6] K. Ahmed and M. A. Gregory, "Integrating wireless sensor networks with cloud computing," in 7th International Conference on Mobile Ad-hoc and Sensor Networks (MSN), Beijing, China, Dec. 16-18, 2011, pp. 240-246.

[7] S. D. Sawaitul, K. P. Wagh, and P. N. Chatur, "Classification and prediction of future weather by using backpropagation algorithm - An approach," International Journal of Emerging Technology and Advanced Engineering, vol. 2, no. 1, pp. 110-113, January 2012.

[8] I. Jagielska, C. Matthews, and T. Whitfort, "An investigation into the application of neural networks, fuzzy logic, genetic algorithms, and rough sets to automated knowledge acquisition for classification problems," Neurocomputing, vol. 24, pp. 49-62, 2012.

[9] D. Ramesh and B. V. Vardhan, "Analysis of crop yield prediction using data mining techniques," International Journal of Research in Engineering and Technology, vol. 4, no. 2, pp. 118-121, 2015.

[10] S. Veenadhari, B. Misra, and C. D. Singh, "Data mining techniques for predicting crop productivity - A review article," International Journal of Computer Science and Technology (IJCST), vol. 2, no. 1, pp. 1-7, March 2011.Wen-Yaw Chung, Pei-Shan Yu, Chao-Jen Huang, “Cloud Computing System based on Wireless Sensor Network”, Federated Conference on Computer Science and Information Systems, 8-11 Sept 2013, pp 877-880.

[11] D. Waga, et al., “Environmental Conditions’ Big Data Management and Cloud Computing Analytics for Sustainable Agriculture,” World Journal of Computer Application and Technology, pp. 73-81, 2014.

[12] R. Piyare, S. Park, S. YeongMaeng, S. C. Oh, S. G. Choi, H. S. Choi, and S. R. Lee, “Integrating Wireless Sensor Network into Cloud Services for Real-time Data Collection,” in International Conference on ICT Convergence (ICTC), Jeju, Oct. 2013, pp. 752-756.

[13] V. Rajesh, J. M. Gnanasekar, R. S. Ponmagal, and P. Anbalagan, “Integration of Wireless Sensor Network with Cloud,” in International Conference on Recent Trends in Information, Telecommunication and Computing, Kochi, Mar. 2010.

[14] V. Panwar, D. K. Sharma, K. V. P. Kumar, A. Jain, and C. Thakar, “Experimental Investigations And Optimization Of Surface Roughness In Turning Of EN 36 Alloy Steel Using Response Surface Methodology And Genetic Algorithm,” Materials Today: Proceedings, vol. 42, pp. 315-320, 2021.

[15] K. Banerjee and S. Dua, “Astoundingly Smart System Furnishing Ranking of Big Data in Search Engines”, J. Comp. Mech. and Mgmt., vol. 1, no. 1, pp. 19–29, Sep. 2022.

[16] U. R. Saxena, P. Sharma, and G. Gupta, “Comprehensive Study of Machine learning Algorithms for Stock Market Prediction during COVID-19”, J. Comp. Mech. and Mgmt., vol. 1, no. 2, pp. 14–21, Dec. 2022..

[17] S. Kumar, “Reviewing Software Testing Models and Optimization Techniques: An Analysis of Efficiency and Advancement Needs”, J. Comp. Mech. and Mgmt., vol. 2, no. 1, pp. 32–46, Feb. 2023.

[18] V. Panwar, D. K. Sharma, K. V. P. Kumar, A. Jain, and C. Thakar, “Experimental Investigations And Optimization Of Surface Roughness In Turning Of EN 36 Alloy Steel Using Response Surface Methodology And Genetic Algorithm,” Materials Today: Proceedings, vol. 42, pp. 315-320, 2021.

[19] M. Shah, N. Naik, B. K. Somani and B. Z. Hameed, "Artificial intelligence (AI) in Urology-Current use and future directions: An iTRUE study," in Turkish Journal of Urology, vol. 46, no. Suppl 1, pp. S27-S32, Nov. 2020, doi: 10.5152/tud.2020.20291.