International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6344
***Article Received**: 30 December 2022 **Revised**: 01 February 2023 **Accepted**: 20 February 2023*

An Intelligent Approach to Reducing Plant Disease and Enhancing Productivity Using Machine Learning

Saikat Banerjee[1], Dr. Abhoy Chand Mondal[2]
[1]State Aided College Teacher, The Department of Computer Application,
Vivekananda Mahavidyalaya, Haripal, Hooghly, The University of Burdwan.
The University of Burdwan, West Bengal, India
saikat.banerjee56@gmail.com
[2]Professor, The Department of Computer Science
The University of Burdwan, West Bengal, India
abhoy_mondal@yahoo.co.in

Abstract— Plant diseases are a normal part of the natural world, and they are one of the many ecological processes that work together to keep the vast number of living organisms in the world in a state of equilibrium with one another. Each plant cell has its own set of signalling pathways that help the plant fight off viruses, animals, and insects. Concerns have been raised about whether or not it is possible to use machine learning to make crop predictions based mostly on weather data. The goal of the research is to help users choose the right crop to grow so that they can maximise their yield and, as a result, the money they make from the project. In a rural area where almost half the people work in agriculture, one of the most important problems is when farmers can't use traditional or other non-scientific methods to choose a crop that will grow well in their soil. Researchers can't make use of case studies as well as they could because there isn't enough correct and up-to-date information available. With the resources at our disposal, we have proposed a model that makes use of random forests and the genetic algorithm. This model has the potential to solve this problem by providing predictive insights on the long-term viability of crops and recommendations based on machine-reading models that have been trained to take important environmental parameters into consideration..

Keywords- Crop; Plant Diseases; Machine Learning; Random Forest; Genetic Algorithm.

I. INTRODUCTION

Climate factors like rain, temperature, and humidity all play a big part in the production phase of agriculture. It is difficult for farmers to make decisions on how to prepare the soil, plant seeds, and harvest their crops when the climate continues to shift as a direct result of increasing deforestation and pollution. We recognise that our planet's climate is changing on a daily basis. For this reason, farmers do not cultivate their vegetation during the ideal times. Many researchers have begun to identify this problem in Indian agriculture and are dedicating their time and efforts to alleviating it. [8] Our purpose is to provide assistance to farmers and agencies. Most of the time, people have to deal with hard situations as they try to get access to and trust in educational outreach and education in order to learn how to improve agricultural yields and their financial situation. Because of the crucial nature of the troubles at stake and the hesitance to consider assistance from outside the network, any services or products meant to help should be carefully designed and tested with the intention to ensure high-quality consequences and a successful adoption. [4]. There is no modern software solution that recommends plants based entirely on more than one factor, such as soil type and elements related to the weather, including temperature and precipitation. The systems that already exist are primarily hardware-based, which makes them luxurious and difficult to hold. Extra functions covered in the system are pesticide prediction and online buying and selling based mostly on agricultural commodities [1–7]. It is critical to introduce technological advances in the fields of crop yield to countries such as India. In the essential sector of qualitative farming, research endeavours and a preliminary study procedure are aimed at enhancing yield and food crop quality at a low cost with a higher monetary consequence [5, 6]. The agricultural construction model is the outcome of a complex interlinking of soil, seeds, and growth-promoting chemicals. Fruits and vegetables are currently one of the most important agricultural products. Product value evaluation and improvement have always been crucial to the acquisition of surplus and effective products. Diseases are impairments to the plant's normal state that affect or prevent vital functions such as transpiration, photosynthesis, fertilisation, pollination, and germination [12, 13]. Pathogens such as fungus, bacteria, and viruses cause diseases as a result of poor environmental conditions. As a result, the preliminary stage of plant disease diagnosis is a critical

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6344
***Article Received**: 30 December 2022 **Revised**: 01 February 2023 **Accepted**: 20 February 2023*

effort. Farmers require professional monitoring regularly, which can be excessively expensive and time-consuming [2, 3]. As a result, finding faster, less expensive, and more precise methods of detecting illnesses from signs on the plant's leaf is critical. In our research, we propose a system that may be used to determine the specific disease that a crop may be suffering from. Identifying the type of illness that an essential crop like a tomato can have by adopting modern technologies like image recognition, which visually reflects the application's functioning, is a key problem, and it is also a major reason for the widespread adoption of digital technologies. In the sphere of agriculture, many people and technological groups are working to boost output and throughput [8–10]. In the past, several approaches have been used in an effort to find solutions to the challenges posed by the development of disease in plants. Because of this need, the technique known as "random forest" (RF) has attracted a lot of attention. The RF method surpasses other traditional techniques here for the categorization of remote sensing photos. This is due to the fact that it requires a lower number of parameters and a lower level of user interaction but produces a higher level of classification accuracy. In addition to this, it is able to process data with a high degree of dimension and deliver classification results rapidly [5, 19]. The RF algorithm is nothing more than a method for doing ensemble learning. The employment of a technique known as "ensemble pruning," which is often referred to as "ensemble selection," "selected ensemble," or "ensemble thinning," is an efficient approach that may be used to improve the performance and effectiveness of ensemble methods, including RF classification. The accuracy of the RF classifier has been significantly improved by the successful use of ensemble pruning procedures such as ranking and statistic-based algorithms. However, the optimizing-based techniques that have been employed in ensemble reduction for those other machine-learning techniques have not been applied in RF classification. This is because RF classification does not use ensemble pruning. Based on its effectiveness in random searching plus parameter optimization in this study, we first suggest using a genetic algorithm (GA) to improve RF classification. This is because GA excels at optimising parameters. A plant's capacity to endure, reproduce, compete, develop, or protect itself from herbivores and other parasites is one of the ways that fungus infections can change the fitness of the host. The results of a fungus-plant relationship can range from extremely parasitic, through communalistic, to mutualistic; therefore, infections do not always have to have consistently negative impacts [5, 11, 16]. The properties of the fungus and plant, along with the ecological circumstances, govern how the connection will turn out. Throughout the whole of agricultural history, one of the primary priorities of farmers has always been the pursuit of increased agricultural output. Which techniques give the most crops per acre? What elements have the greatest impact on crop yield? Recent years have seen a rise in this topic's significance as a direct result of the consistent growth in the global population. However, when new challenges for farmers appear, new methods and technological developments become necessary to handle those challenges. A large portion of the variability in disease incidence across places and times appears to be controlled by environmental and climatic factors. Non-systemic foliar diseases decrease plant reproduction, yet certain relationships may be almost coincidental due to healthy individuals' poor seed set. In situations where seed reproduction is not vital, the impacts of illness are thereby reduced. Pathogens that spread throughout a plant's body can change its capacity to compete in the environment, reproduce, grow, and resist herbivores. The connection can have both positive and negative impacts, and its overall result can range from extremely parasitic to absolutely mutualistic. Since the beginning of time, people have carefully chosen and raised plants for use as food, medicine, fibre, clothing, shelter, and ornamentation. When plants are removed from their natural surroundings and cultivated in pure stands under what are sometimes abnormal conditions, a number of dangers, including disease, must be taken into account. Due to their high susceptibility to disease and problematic survival in nature without human intervention, many important crop and ornamental plants are also particularly sensitive to illness. Compared to their wild ancestors, cultivated plants frequently have higher disease susceptibility [14, 16]. This is due to the fact that several individuals of the same species or variety, each with a distinct genetic background, are cultivated in close proximity, often spanning vast areas covering many thousands of square kilometres. In those circumstances, a disease may spread quickly. This study makes two distinct contributions. On the one hand, this study offers a plausible technique for RF classification parameter setting. However, in terms of application, this research provides a practical and successful method for long-term planning in agricultural change. The majority of insects get their food from plant life. Bedbugs can eat leaves, stems, roots, and flower petals. Chewing insects actually causes them to eat the affected parts [5, 11, 17]. Types of leaf-feeding with chewing bugs include hollow leaf litter with leaf beetles, beetles, and small caterpillars. Rare notches next to leaf blades are usually caused by numerous foxes, large caterpillars, locusts, and katydids. The best semi-circular reduction in leaf rates suggests the presence of bee-cutting bees. The term "mining" refers to the complete consumption of leaves. It's possible to classify leaf miners with insects, flies, sawflies, and moths [15].

II. RELATED WORK

The identification of 26 diseases in 14 different crops using a total of 54,306 images and an approach utilising convolutional neural networks is described in the article [20]. They used these

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6344
***Article Received**: 30 December 2022 **Revised**: 01 February 2023 **Accepted**: 20 February 2023*

images to analyse the disease's symptoms. When there are 38 different possible crop-disease combinations, they assess the accuracy of their models based on how well they can predict which one will occur. They explored whether or not deep convolutional neural networks could be utilised to solve the classification problem that was presented earlier. They concentrated on two well-known frameworks, namely AlexNet and GoogleNet, both of which are widely used. This purpose has been accomplished inside the PlantVillage data set, which consists of 54,306 images and contains 38 classes of 14 agronomic traits and 26 diseases, as evidenced by the highest accuracy score of 99.35%. In [21], researchers talk about how machine learning (ML) and deep learning (DL) have helped scientists find plant diseases. In the event that the diseases really aren't identified, they will have an effect on the yield of the crop, which will ultimately lead to longer-term problems such as climate change and even famine. In addition, they talked about a few other topics, such as the database searches and size issue, the problems with the currently offered feature extraction methods, and the difficulties in the classification module. Looked at the maize plant for the purpose of analysing various diseases [22] that affect the leaves. They discussed the many categorization strategies that may be achieved by machine learning, including KNN Decision Tree, Naive Bayes, SVM, and RF. Image capture, image preprocessing, segmentation techniques, and edge detection are the approaches that are being considered for the categorization of diseases. The collection pertaining to maize plant diseases includes a total of 3,823 photos. They achieved an accuracy of 79.23% using the RF method, which is superior to previous methods. A machine learning algorithm-based rice leaf disease detection model that could help find diseases. [23] We used WEKA (Waikato Environmental for Knowledge Analysis), an open-source machine learning tool, to apply a variety of machine learning techniques in order to train our model. The data for this job was gathered from the Machine Learning Repository. K-means, logistic regression, naive bayes, and decision trees are the approaches that were used in this study. On the basis of the test results, it was discovered that the decision tree provided the highest accuracy with 97.9167%. Image capture was the first step that [24] took in developing a plant disease detection system. SVM, feature extraction, and CNN were the subsequent approaches that emerged after that. They gathered 12949 photos for CNN, which resulted in an accuracy of 97.71 percent, which was the highest possible. In [25], the authors undertook comparative research on the algorithms that were employed in a variety of past projects and then rationalised out all the ones that were best suited for the project based on their findings. They decided to use tomatoes for their paper since tomatoes are one of the most well-known vegetables all over the world and can be seen being used in a variety of different contexts, such as in salads, soups, and curries, among other things. The classification algorithms, such as SVM, CNN, and KNN, are fairly widely employed and are used in practically all applications that are conceptually comparable. They took 200 photographs of the leaf. First comes the preprocessing step, which includes scaling, smoothing, and noise filtering for the picture. Next comes feature extraction, followed by HOG and GLCM. They used the KNN classification algorithm because the KNN algorithm could be explained as supervised machine learning because it can be used to find solutions to both classification and regression problems. Also, the KNN algorithm can be applied to find solutions to problems involving clustering. Use accurate ways to classify leaves [26]. To figure out if a leaf has a disease or not, we use certain methods, such as pre-processing, which is used to reduce the size of the picture, and then a feature extract. This allows us to determine whether or not a leaf is infected. Using HOG, we are able to determine the nature of the features that are present in a picture (the histogram of oriented gradients). There are differences in texture between diseased and healthy leaves; to identify these differences, HOG makes use of the Haralick texture method, and for the purpose of analysing the colour scheme in the picture, HOG makes use of the colour histogram. Hu moments are used for highlighting the leaf, and the colour histogram is used for analysing the colour. The subsequent phases, after the process of obtaining the features, are learning the classifier and classifying the data. In order to do this, an algorithm that uses the RF classifier is put into action. A characteristic vector is created again for the training dataset that makes use of HOG, and this feature vector is then trained using a RF algorithm. We can tell when something is wrong with a plant's leaves by using all of these methods. Accuracy is at a level of 70% within this categorization model. use a deep learning algorithm [27] to provide an accurate and automated way to find out how bad a plant pathogen is. When trying to control plant diseases, ensure food security, and predict yield loss, it's important to think about how bad the diseases are. This characteristic is used to determine the severity of the condition. To determine which design produces the superior deep convolutional neural network, researchers will compare the following: I am constructing a shallow network from the ground up, one that is made up of layers and (ii) transfer learning, which involves using a limited number of data points in order to construct an effective classification network by perfectly aligning the parameters of the network. Botanists select the classes of certain healthy as well as apple black rot leaf photographs from the dataset provided by Plant Village. These classes indicate if the leaf is in the healthy stage, the early stage, the middle stage, or the final stage of its development. The number of samples used for each class has been set up so that eighty percent of the images are used as the training dataset for the beginning,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6344
***Article Received**: 30 December 2022 **Revised**: 01 February 2023 **Accepted**: 20 February 2023*

middle, and end stages, while the remaining twenty percent are used as hold-out test sets for balancing. According to the findings of a study of the networks VGG16, VGG19, Inception-V3, and ResNet50, network performance may be improved by the use of fine tuning on pretrained deep models. The VGG16 model delivers the greatest results in this regard, with a reliability of 90.4%. In order to gather more data for use in subsequent studies, flexible sensors such as thermal imaging and multispectral cameras are used. In the study [28], the authors used two different methods to segment the data: thresholding and the K-means clustering algorithm. They also used an artificial neural network to classify the data. The first picture is divided into sections using these two different segmentation algorithms. Then, useful features are taken from the sections to find the diseased parts of the plant leaf. Following that, the classification method was carried out by using the nf toolbox inside the MATLAB programme. Applications of picture thresholding, K-means clustering, and neural networks have been used to develop a method for grouping and classifying diseases of plant leaves (NN). Using a number of different algorithms, the effects of different diseases on the plants were tested. Based on the results of the experiments, which strongly suggest that the neural network gives the right answer in less time than other methods, we can say that it is the most accurate method.

III. METHOD

The architecture of a system is indeed a conceptual model that allows us to characterise the behaviour and structure of the proposed method using that model. The proposed model based on a machine learning algorithm is shown in Figure 1. The architecture of the system that our method uses is shown in Figure 2. It is a formal depiction of a system. The term "system architecture" has the potential to refer to both a model that is used to explain the system and a technique that is used to construct the system, depending on the context in which it is used. Constructing an appropriate architecture for the system is beneficial to evaluating the proposed work, particularly in the beginning phases. In a more technical sense, it is an ensemble strategy (with the divide-and-conquer method as the main focus) of decision trees that are built on a dataset that has been arbitrarily split up. The term "forest" may also be used to refer to this set of decision tree classifiers. Before each decision tree is built, each trait is given an attribute selection indicator, such as extracted features, the gain ratio, or the Gini index. A completely independent random sample is used for each individual tree. When it comes to a problem requiring classification, each individual tree gets a vote, as well as the category itself, which provides the answer that receives the greatest support. The ultimate outcome of the regression analysis is found by averaging all of the findings that were received from he trees, and this is how the result is decided. Compared to some other non-linear classification algorithms, this one is not only easier to use, but it also gives better results. The procedure will be described in detail below, step by step. Charles Darwin's theory of evolution helped shape GA, which are a type of natural search strategy. The flowchart of the GA is shown in Figure 2. By simulating the processes of biological evolution, reproduction, and mutation, GA are able to provide high-quality solutions to a wide range of issues. These challenges include search and optimization. By properly applying Darwin's theory of evolution, GA are able to tackle problems that traditional algorithms are unable to. According to Darwin's theory of evolution, evolution ensures that the population will always include individuals who are distinct from each other (variation). Those individuals who are better able to adjust to their environment have a greater likelihood of surviving, propagating, and passing on their characteristics to descendants (survival of the fittest) [18, 19].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6344
***Article Received**: 30 December 2022 **Revised**: 01 February 2023 **Accepted**: 20 February 2023*

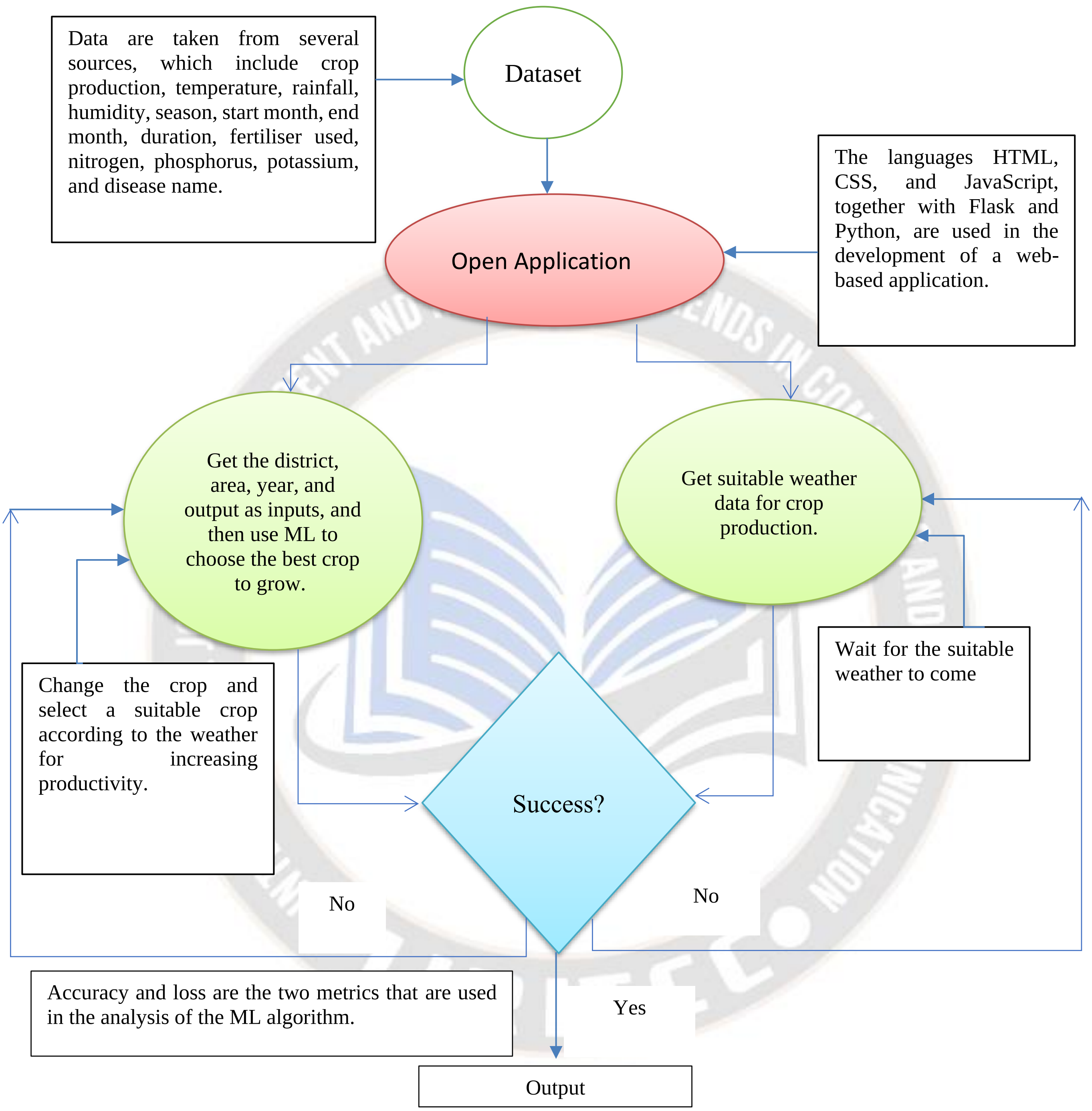

Figure 1 Proposed Model Flowchart

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6344
***Article Received**: 30 December 2022 **Revised**: 01 February 2023 **Accepted**: 20 February 2023*

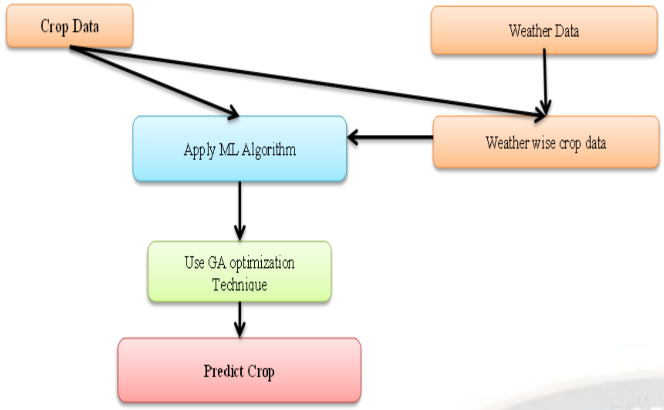

Figure 2. System Architecture

Step By Step Workflow
1. Using the dataset provided, choose several samples at random.
2. Generate a decision tree for all of the samples, and then get a prediction from each of the decision trees.
3. Hold a vote to determine the outcome of each prediction.
4. Choose the outcome of the prediction that got the most votes as the final prediction.

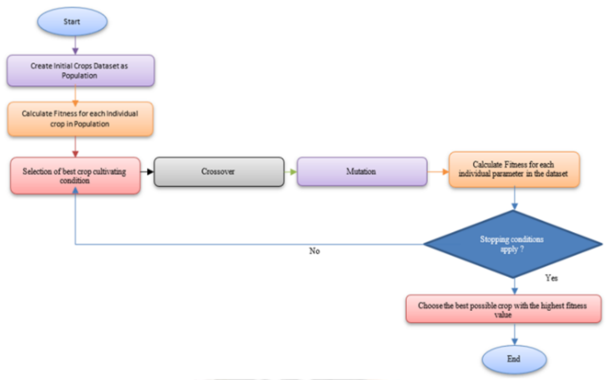

Figure 3 Flowchart of Genetic Algorithms

Chromosome/Individual: Genes are organised into collections known as chromosomes. For example, a chromosome may be portrayed as a binary string, and each bit in that string can stand in for a gene. Each kind of chromosome may be thought of as a unit. Because each person is equivalent to one chromosome, a population may be thought of as a collection of these chromosomes. Make a dataset called "Initial Crops" and call it the population. The fitness function is in charge of figuring out each person's fitness score and using that score as the basis for judging each person at each step of the iteration. People with higher fitness scores are thought to have better solutions, and they have a better chance of being chosen to cross over and be passed on to the next generation. We do fitness analysis on each unique crop that is included in the dataset. The suggested sequence diagram is an interaction diagram that shows how methods work with each other and what order they should be done in. This diagram is represented in Figure 3. It's a component of a data collection chart. The interactions between objects are shown using a chain diagram and are ordered chronologically. It illustrates the items and instructions that were engaged in the circumstance, as well as the sequence of information that was passed between the items in order for them to carry out the functionality of the state of things. Inside the operational view of the system that is currently being developed, collection diagrams are often related to use-case realisations. The diagrams that represent a series of events are often referred to as event schematics or occasion eventualities. A sequential diagram is a simple and straightforward method of characterising the behaviour of an apparatus by looking at the interaction between the apparatus and its surrounding environment. An interaction shown as a time-ordered sequence of events is represented by a chain diagram. A sequence diagram includes dimensions: the vertical measurement reflects the passage of time, and the horizontal measurement indicates the many states that an object may be in throughout the course of the interaction.

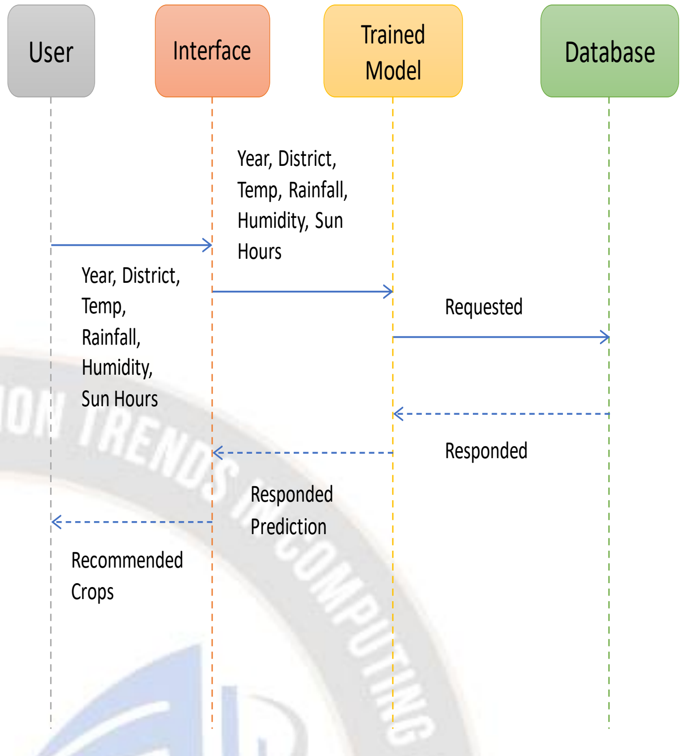

Figure 4. Proposed Sequence Diagram

IV. IMPLEMENTATION

A. *Data Analysis*

One of the first things we do when putting something in place is to evaluate the information. This was carried out without our participation, and we then investigated whether or not there were any connections between the many qualities that were already present in the databases. Table 1 shows the dataset of our model

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6344
***Article Received**: 30 December 2022 **Revised**: 01 February 2023 **Accepted**: 20 February 2023*

that we used in the model. Acquisition of education datasets: The accuracy of any machine studying a set of rules depends on the number of parameters and the correctness of the education dataset. In this challenge, we analysed more than one dataset amassed from government websites and Kaggle and thoroughly selected the parameters that could provide satisfactory results. Many studies in this field have found that environmental parameters are important for crop sustainability; some have concentrated on yield, while others have concentrated on the most fundamental economic elements. We've tried to combine both environmental parameters like rainfall, temperature, humidity, and sun hours with monetary parameters like manufacturing and proximity to offer correct and dependable advice to the farmer on which crop could be most appropriate for his land.

B. Data Preprocessing

The following phase, which comes after evaluating and displaying the data, is called pre-processing. The phase of data preparation is crucial since it helps clean the data and makes it appropriate for use within machine learning algorithms. This makes it an essential step. During the pre-processing stage, the majority of attention is paid to the elimination of any outliers or incorrect data, as well as the management of any missing values.

There are two approaches that may be used to deal with missing data. The first approach is to just get rid of the whole row that has the incorrect or missing data in it. This is the simplest solution. Although it is a simple procedure, you should use it only for very large datasets if at all possible. When used with relatively small datasets, this strategy may result in an excessively reduced dataset size, which is particularly likely when there are a significant number of variables that are missing. It's possible that this will have a substantial effect on how reliable the conclusion is. The best way to gather and measure data from different sources, such as government websites, the website for the Kisaan Helpline, the website for the India Meteorological Department, and so on, to get an approximate dataset for the system is to collect the data. The following characteristics need to be included in this dataset. For crop prediction, the following characteristics will be taken into consideration: I: Soil PH; II: Temperature; III: Humidity; IV: Rainfall; V: Crop Data; and VI: forecast, we gather rainfall data from the years before. There are a number of different things that might have an influence on the yield of a crop as well as its overall productivity. In their most basic form, these are the factors that contribute to an accurate forecast of the annual yield of any given crop. In this study, we take into account a variety of variables, including temperature, rainfall, land area, humidity, and wind speed. The total number of stocks included in this dataset is 88138, with 15 columns. There are nine columns for numerical data and six columns for categories. Table 1 shows part of our dataset.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6344
***Article Received**: 30 December 2022 **Revised**: 01 February 2023 **Accepted**: 20 February 2023*

Table 1 Dataset

Crop	Start (Month Name)	End(Month Name)	Season	Duration	Production(Metric-ton)	Temparetue (Celcius)	Rainfall (mm)	Humidity (%)	Sun hours (hours)	Fertilizer Name	Disease Name	N %	P %	K %
Rice	Dec	Feb	Winter	3	1231421	19.1	46	63	9.1	Ammonium Sulfate	Bacterial blight	50	12	12
Urad	Oct	Feb	Rabi	5	6	20.3	171	66	9	Urea	Cercospora leaf spot	20	40	40
Soyabean	Jun	Oct	Kharif	5	2	26.9	1189	80	8.2	Phosphorus	Frogeye leaf spot	20	40	20
Wheat	Oct	Feb	Rabi	5	10115	20.3	171	66	9	Nitrogen	Powdery Mildew	80	40	40
Mesta	Jun	Oct	Kharif	5	46	26.9	1189	80	8.2	NPK	Foot and stem rot	40	20	20
Tomato	Mar	Jun	Summer	4	7512	29.6	247	55	9.9	NPK	Damping off	3	4	6
Maize	June	Aug	Kharif	3	3451	27.4	1204	81	7.9	NPK	Maydis leaf blight	60	30	30
Bajra	Jun	Jul	Kharif	2	2574	29.7	842	83	8.4	NPK	Phyllosticta leaf blight	40	20	20
Urad	Jun	Oct	Kharif	5	155	26.9	1189	80	8.2	Urea	Cercospora leaf spot	20	40	40
Chickpea	Oct	Nov	Rabi	2	18420	20.4	614	79	8.3	NPK	Root Rot	20	40	
Green Pea	Oct	Nov	Rabi	2	12046	21.3	451	81	8.1	Phosphorus	Pea rust	20	60	40
Cauliflower	Sep	Dec	Winter	4	1640	22.1	341	81	8.1	NPK	Downy Mildew	12	24	12
Groundnut	Jan	Dec	Annual	12	3619	25.71	1474	78.38	8.9	Nitrogen	Botrytis blight Botrytis cinerea	50	25	25
Peas & beans (Pulses)	Oct	Feb	Rabi	5	2	20.3	171	66	9	NPK	Downy Mildew	10	10	10

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6344
***Article Received**: 30 December 2022 **Revised**: 01 February 2023 **Accepted**: 20 February 2023*

C. Application Setup

The implementation of the system can be divided into two phases, i.e., frontend and backend implementation. The flowchart of our application setup is shown in Figure 4. The frontend is implemented using web technologies—HTML, CSS, and JavaScript. Here, we used Python in the backend. Figure 5 displays the source code that examines the input and makes an effort to anticipate the result. There is a possibility that the volume of the tree-based forest will have an impact on the precision of the forecasts. The reliability of the forecast is improved as a direct result of expanding the tree-based forest.After implementing the frontend and backend, we use Flask to make a web-based application. A demo of our application is shown in Figure 6.

Figure 5 Application Setup

Hardware System Configuration:

Processor: Intel(R) Core (TM) i7-10750H CPU @ 2.60GHz 2.59 GHz

RAM: 16 GB

Hard disk space: 1TB

SSD: 512 GB

GPU: 4 GB (Nvidia)

V. RESULTS AND PERFORMANCE ANALYSIS

In this study, the RF classification algorithm was better at making accurate predictions than the single-tree model. However, the RF classification algorithm is a black-box model, so it can't be used to see the decision tree forest. The accuracy of the predictions may depend on the size of the decision tree forest. The accuracy of a prediction gets better as a decision tree forest grows. There are two different kinds of size control mechanisms associated with the RF. These are the total number of trees present inside the RF and the height of the tree in question. Some research suggests that it is best to cultivate trees that are very large in size; hence, the maximum level would represent very large trees, and indeed, the minimum node length control should be set to a value that restricts the dimensions of the trees. After using the model that was presented, an increase in productivity can be shown in Figure 10.Accuracy ratings for each of the two techniques shown in Figure 8 After the training dataset is managed with the RF algorithm and the GA Process, we predict the harvest using the value of weather data such as temperature, rainfall, humidity, and sun hours and get an accuracy of 94%. This is a significant improvement over the previous method. When we employ simply RF, on the other hand, we get an accuracy of 87%. Figure 5 illustrates the many parameters that are used as input, while Figure 9 displays the page that is produced.

In this research, a RF prediction model outperformed the model in terms of making accurate predictions. The RF classification approach, on the other hand, is a black-box model, which indicates that it is not feasible to access its structure's actual decision tree behaviour by utilising it. This is because the algorithm is used to classify data. There are two distinct types of procedures for controlling the size of the RF's trees. These values represent both the total quantity of trees that are discovered inside the RF as well as the volume of the tree that is being discussed. The contrast between the GA and RF approaches is seen in Figures 11 and 12. Figure 11 shows India's total crop production during 2022 in selected crops such as rice, wheat, and etc. in comparison to the predicted total crop using GA and RF in the said crops. As a result, the highest production should reflect very large scale, and the minimum node volume control must be set to a value that inhibits the magnitude of the graph A higher crop production output value suggests a higher level of accuracy in the prediction. The results of an analysis of the quality metrics parameters using various parameter approaches are shown in Figures 7 and 8. Fig. 9 shows accuracy of Ga and RF in their respective forms. The graph shows that the performance of GA is much better than other

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6344
***Article Received**: 30 December 2022 **Revised**: 01 February 2023 **Accepted**: 20 February 2023*

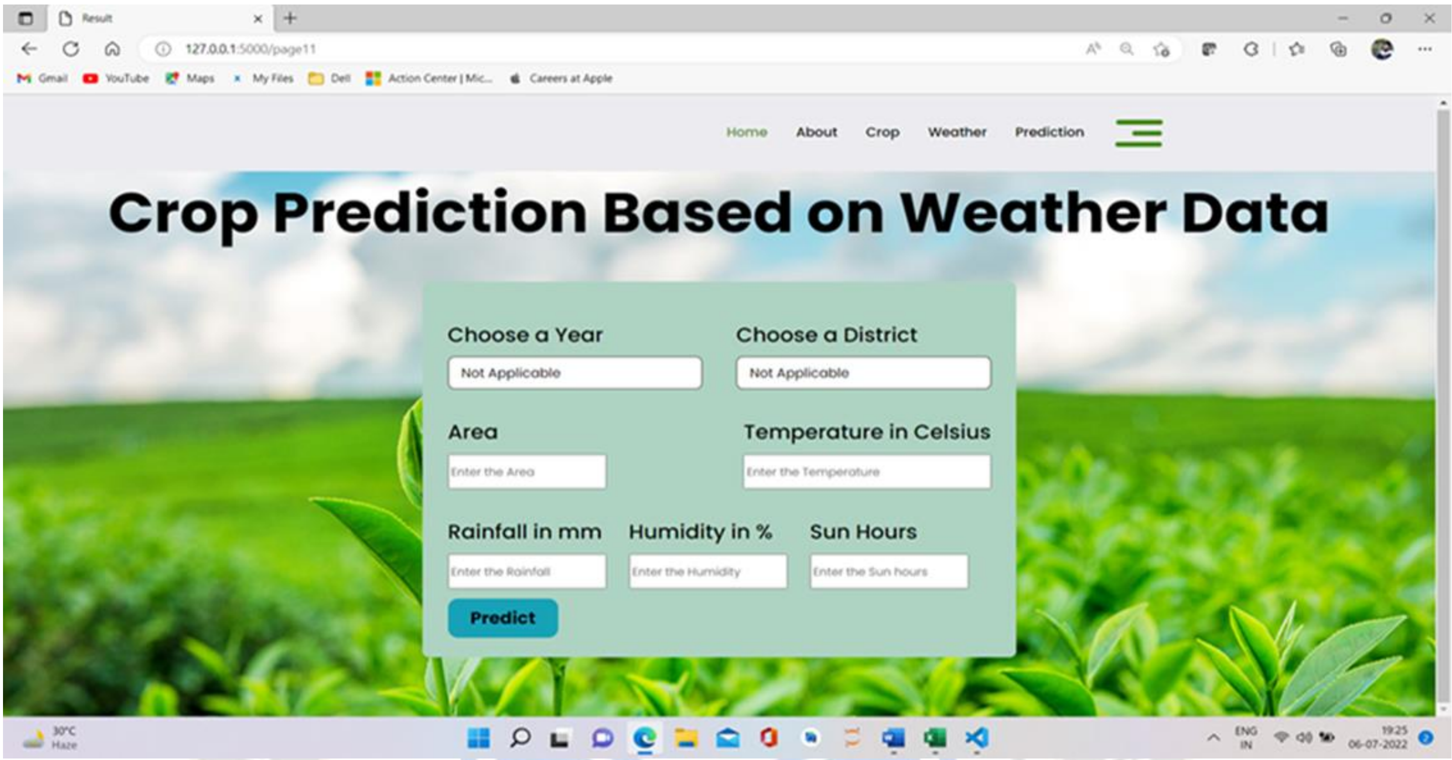

Figure 6 Application Input Page

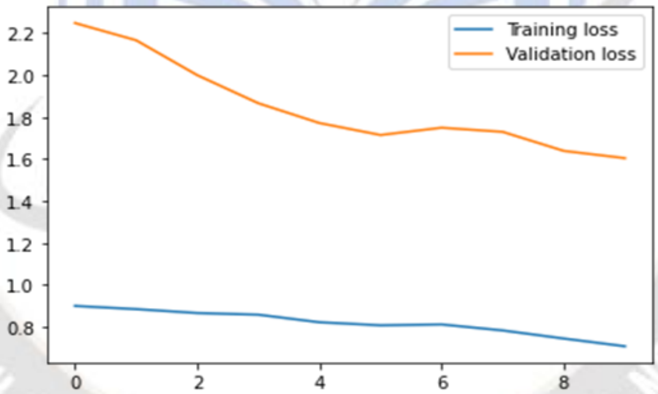

Figure 7 Training and Validation loss of Random forest

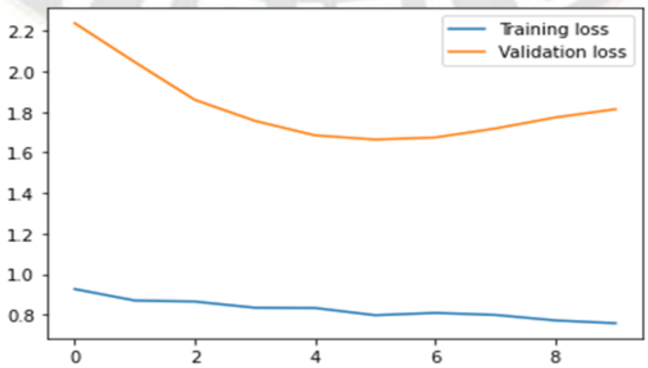

Figure 8. Training and Validation loss of Genetic Algorithm

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6344
***Article Received**: 30 December 2022 **Revised**: 01 February 2023 **Accepted**: 20 February 2023*

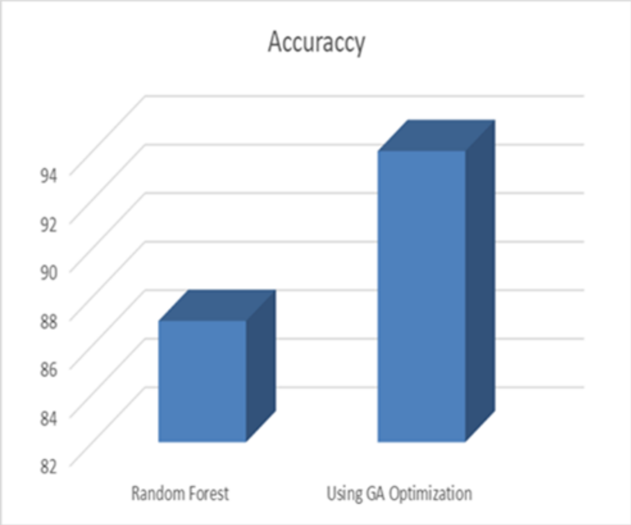

Figure 9 Accuracy

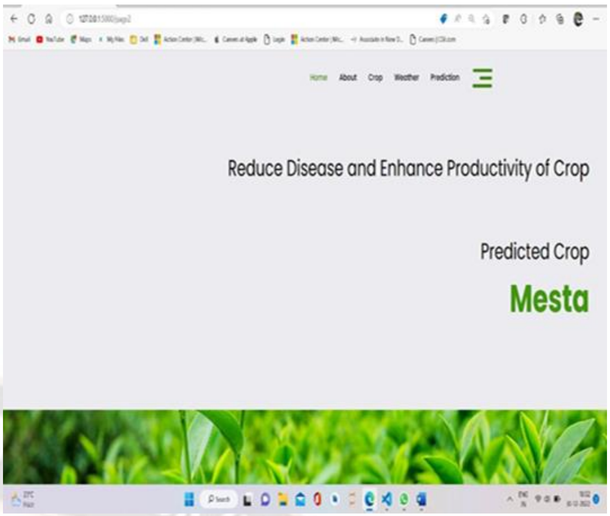

Figure 10. Application Output Page

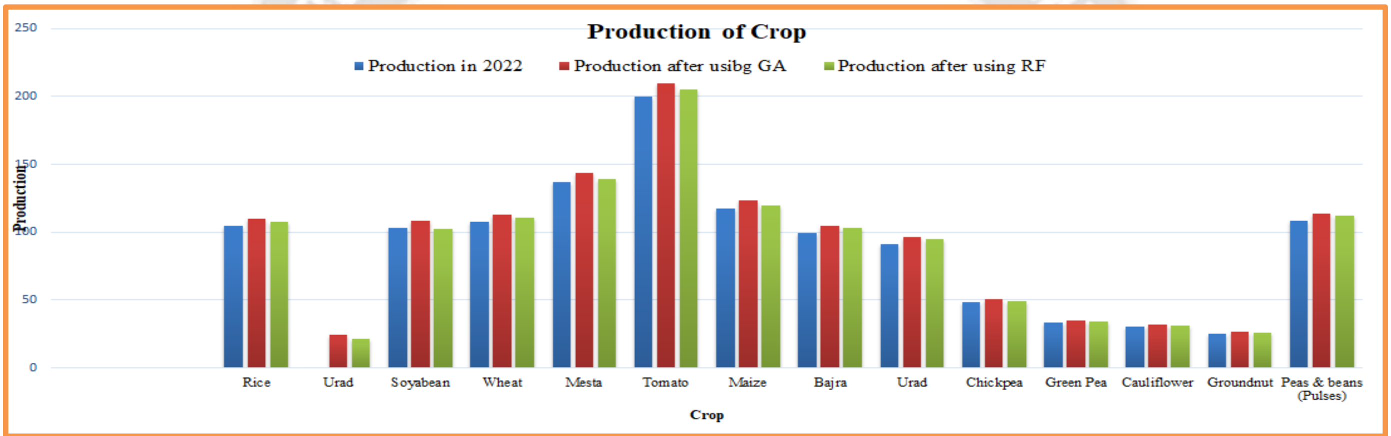

Figure 11. crop production on the year 2022 and predicted production on proposed model.

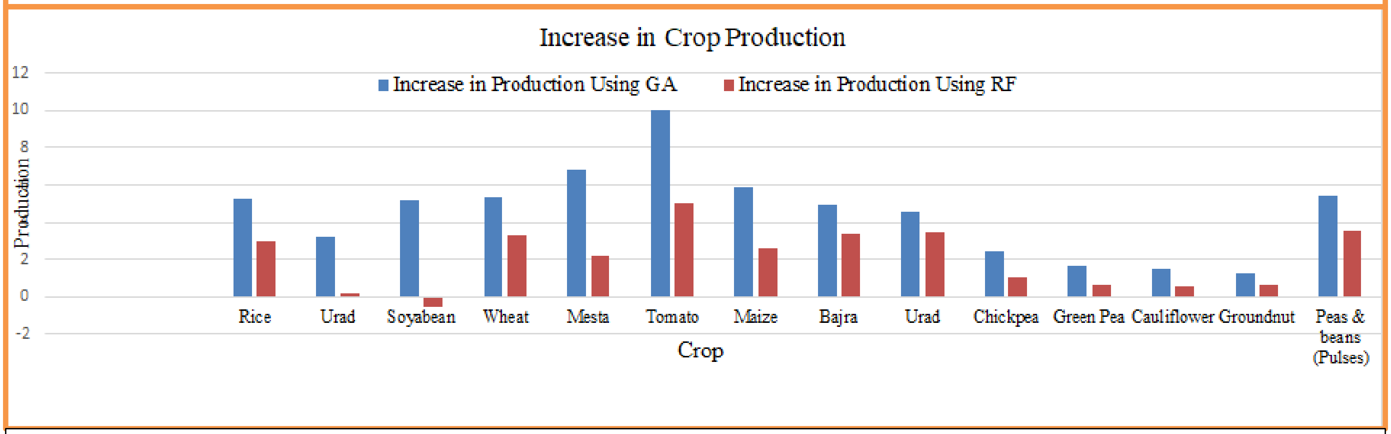

Figure 12. Increase in crop production based on proposed model.

VI. CONCLUSION

The evolutionary algorithms are a powerful tool that can be utilised for the purpose of finding solutions to a broad range of issues. It is able to identify solutions by using evolutionary processes, even though it does not fully comprehend the complexities of the issue. Seasonal factors that have an effect on farmers' ability to secure their food and livelihoods are generally not systematically analysed and integrated into development programming, which results in programmes that are not sustainable and do not meet the needs of farm families. This research is a qualitative study that analyses how climatic unpredictability is affecting current seasonal challenges with food availability in three indigenous communities. The concerns being studied are related to the food supply. Gaining an understanding of these processes and their effects on the stability of one's food supply and means of subsistence provides crucial background information that can be used to assist in the decision-making process regarding viable and long-term solutions that are focused on improving one's standard of living. The search space used by the RF approach is both reduced and optimised so that it may take advantage of the benefits offered by GA, as described above. The GA is used to "evolve" the decision tree in a RF environment. An appropriate combination of decision trees has led to the development of a method known as RF, which is based on genetic optimization. Because it uses integrated learning, this method beats other classification algorithms in terms of performance. In addition to this, it uses a GA to optimise the parameters, which leads to a gain in operating efficiency as well as a quicker search for the parameters that are optimal. The experiment demonstrates that the algorithm's impact on a real data set of the crop is more obvious and that it successfully predicts and analyses the crop, enabling it to be predicted early on. This is shown by the fact that the experiment was successful. The further effort will centre on providing the crop sequence that should be produced depending on the characteristics of the soil and the weather, as well as continually updating the information in order to provide accurate forecasts. The objective of the Future Work is to design an IoT-based platform that is capable of collecting accurate weather data in a totally automated manner.

REFERENCES

[1] AlaaBadrEysa; Asst Prof. Dr.Sefer Kurnaz, "Diabetes Diagnosis using Machine Learning|", IJCSMC vol. 8, Issue. 3, March 2019, pg.36-41.

[2] Bishwajit Purkaystha, Tapos Datta, MdSaiful Islam, "Bengali Handwritten Character Recognition using Deep Convolutional Neural Network", 20th ICCIT, 22-24 December 2017.Online Available-https://ieeexplore.ieee.org/document/8281853

[3] Ms Kavita, Pratistha Mathur, "Crop Yield Estimation in India Using Machine Learning", 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA) Galgotias University, Greater Noida, UP, India. Oct 30-31, 2020.

[4] Kento Morita, Atsuki Tashita, Manabu Nll,Syoji Kobashi, "Computer-Aided Diagnosis System for Rheumatoid Arthritis using Machine Learning", International conference of Machine Learning and Cybernatics, Ningbo, China, 9-12 July2017".Online Available-https://ieeexplore.ieee.org/document/8108947.

[5] Dr. Y. Jeevan Nagendra Kumar, V. Spandana, V.S. Vaishnavi, K. Neha, V.G.R.R. Devi, "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector‖", Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020) IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1.

[6] Jarrett Booz, Wei Yu, Guobin Xu, David Griffith, Nada Golmie, "A Deep Learning-Based Weather Forecast System for Data Volume and Recency Analysis", 2019 International Conference on Computing, Networking and Communications (ICNC): Green Computing, Networking, and Communications.

[7] Anakha Venugopal, Aparna S, Jinsu Mani, Rima Mathew, Prof. Vinu Williams," Crop Yield Prediction using Machine Learning Algorithms", International Journal of Engineering Research & Technology ISSN: 2278-0181 Published by, www.ijert.org NCREIS - 2021 Conference Proceedings.

[8] Shook J, Gangopadhyay T, Wu L,Ganapathysubramanian B, Sarkar S, Singh AK, ―Crop yield prediction integrating genotype and weather variables using deep learning‖ , PLoSONE 16(6): e0252402. https://doi.org/10.1371/journal.pone.0252402, June 17, 2021.

[9] Kazuo Yonekura, Hitoshi Hattori and Taiji Suzuki, "Short-term local weather forecast using dense weather station by deep neural network‖", IEEE International Conference on Big Data, IEEE 2018.

[10] Khaki S and Wang L, ―Crop Yield Prediction Using Deep Neural Networks‖, Front. Plant Sci. 10:621.doi: 10.3389/fpls.2019.00621, 22 May 2019.

[11] Sonal Agarwal, Sandhya Tarar, "A HYBRID APPROACH FOR CROP YIELD PREDICTION USING MACHINE LEARNING AND DEEP LEARNING ALGORITHMS", CONSILIO 2020 Journal of Physics: Conference Series, 1714 (2021) 012012 IOP Publishing doi:10.1088/1742-6596/1714/1/012012.

[12] Rishi Gupta, Akhilesh Kumar Sharma, Oorja Garg, Krishna Modi, Shahreen Kasim, Zirawani Baharum, Hairulnizam Mahdin, Salama A. Mostafa, ―WB-CPI: Weather Based Crop Prediction in India Using Big Data Analytics ‖ -VOLUME 9, 2021, IEEE Access.

[13] Thayakaran Selvanayagam, Suganya S, Puvipavan Palendrarajah Mithun Paresith Manogarathash, Anjalie Gamage, Dharshana Kasthurirathna , "Crop Prediction using Machine Learning", Volume 4, Issue 10, International Journal of Innovative Science and Research Technology, 2019.

[14] Sonal Jain, Dharavath Ramesh, "Machine Learning convergence for weather based crop selection", 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
DOI: https://doi.org/10.17762/ijritcc.v11i3.6344
***Article Received**: 30 December 2022 **Revised**: 01 February 2023 **Accepted**: 20 February 2023*

[15] Shubham Madan1, Praveen Kumar, Seema Rawat, Tanupriya Choudhury, "Analysis of Weather Prediction using Machine Learning & Big Data‖", ICACCE-2018, Paris, France 22-23 June 2018.

[16] Urmay Shah, Sanjay Garg, NehaSisodiya, Nitant Dube, Shashikant Sharma, "Rainfall Prediction: Accuracy Enhancement Using Machine Learning and Forecasting Techniques", 5th IEEE International Conference on Parallel, Distributed and Grid Computing(PDGC-2018), Solan, India, 20-22 Dec, IEEE 2018.

[17] Aruvansh Nigam, Saksham Garg, Archit Agrawal, Parul Agrawal, ―"Crop Yield Prediction Using Machine Learning Algorithms", Fifth International Conference on Image Information Processing, 2019.

[18] Yi-Chieh Shen, Chi-Hang Hsu, Sung-Huai Hsieh, "Integrated Genetic Algorithms and Cloud Technology to Solve Travelling Salesman Problem on Hadoop", 2012, IEEE 4th International Conference on Cloud Computing Technology and Science.

[19] Rubén Iván Bolaños, Mauricio Granada Echeverry , John Willmer Escobar, "A multiobjective non- dominated sorting genetic algorithm (NSGA-II) for the Multiple Traveling Salesman Problem", Decision Science Letters 4 (2015) 559–568 © 2015 Growing Science Ltd. All rights reserved.

[20] Sharada P. Mohanty, David P. Hughes , and Marcel Salathé ," Using Deep Learning for Image-Based Plant Disease Detection.", Front. Plant Sci. 7:1419. doi: 10.3389/fpls.2016.01419.

[21] Rehan Ullah Khan, Khalil Khan, Waleed Albattah and Ali Mustafa Qamar, "Image-Based Detection of Plant Diseases: From Classical Machine Learning to Deep Learning Journey",. Hindawi, Wireless Communications and Mobile Computing Volume 2021, Article ID 5541859, 13 pages. https://doi.org/10.1155/2021/5541859.

[22] Kshyanaprava Panda Panigrahi, Himansu Das, Abhaya Kumar Sahoo and Suresh Chandra Moharan," Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms",. Springer Nature Singapore Pte Ltd. 2020 H. Das et al. (eds.), Progress in Computing, Analytics and Networking, Advances in Intelligent Systems and Computing 1119. https://doi.org/10.1007/978-981-15-2414-1_66.

[23] Kawcher Ahmed, Tasmia Rahman Shahidi, Syed Md. Irfanul Alam and Sifat Momen, "Rice Leaf Disease Detection Using Machine Learning Techniques.", 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), 24-25 December, Dhaka

[24] Utkarsha N. Fulari, Rajveer K. Shastri, Anuj N. Fulari, "Leaf Disease Detection Using Machine Learning.", Journal of Seybold Report ISSN NO: 1533-9211

[25] G. Geetha , S.Samundeswari , G.Saranya ,K.Meenakshi and M. Nithya, "Plant Leaf Disease Classification And Detection System Using Machine Learning.", ICCPET 2020 Journal of Physics: Conference Series 1712 (2020) 012012 , IOP Publishing doi:10.1088/1742-6596/1712/1/012012.

[26] Shima Ramesh, Niveditha M, Pooja R, Prasad Bhat N, Shashank N, Mr. Ramachandra Hebbar & Mr. P V Vinod, "Plant Disease Detection Using Machine Learning.", International Conference on Design Innovations for 3Cs Compute Communicate Control., DOI 10.1109/ICDI3C.2018.00017.2018.

[27] Guan Wang, Yu Sun and Jianxin Wang, "Automatic Image-Based plant disease severity estimation using Deep Learning.", Volume 2017| Article ID 2917536., https://doi.org/10.1155/2017/2917536.

[28] Trimi Neha Tete and Sushma Kamlu, "Plant Disease Detection Using Different Algorithms",. Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering,.pp.103–106, DOI: 10.15439/2017R24, ACSIS, Vol. 10 ISSN 2300-5963