PRINCIPLES AND PRACTICES OF MAKING AGRICULTURE SUSTAINABLE: CROP YIELD PREDICTION USING RANDOM FOREST

SYED MUZAMIL BASHA[*], DHARMENDRA SINGH RAJPUT[†], J JANET[‡], SOMULA RAMASUBBAREDDY[§] AND SAJEEV RAM[¶]

Abstract. Agriculture has advanced tremendously over the last 100 years. In fact it is been keeping up with food production at a very high rate. In fact, some scientists feel that agriculture already produces enough food to feed the world, but of course there are issues and problems with food availability, agricultural production practices, preservation and transportation, and probably more that one can think of that hinder many people in this world from getting adequate food. The basic challenge is to provide food for the needy people. This need can be fulfilled with the help of the farmers taking responsibility in increasing the food production by 50% by the year 2050. The objective of the present work is to increase this food production, protecting the environment with managing natural resources. Mainly focusing on water, nutrients and other inputs to produce foods without degrading the environment. The Goal is to develop the social, environmental, and the economic aspects of possible solutions to minimize the agricultural footprint, and become more sustainable. The dataset considered in our experiment is used in yield prediction based on historic yield and weather information. Implemented both the versions of Thomson model and compared the result with segmentation model, Random Forest (RF). Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are used as evaluation metrics in estimating the performance of models implements and stated that Random forest algorithm is providing 0.07(RMSE). The outcome of the present research work helps farmers in adopting best management practices and trying to give them the economical and technical support in making easier for them to adopt best management practices.

Key words: Food production, Sustainability, Thomson model, segmentation model, Random Forest, Root Mean Square Error, Mean Absolute Error.

AMS subject classifications. 97R40

1. Introduction. Farmers do indeed feed the world. Indian farmer is said to produce enough food for his or her family and for at least 150 other people in the country. Researchers estimates the need to increase food production by 50 to 70% along with maintaining sustainably. Getting pressure on agriculture on farmers to increase this food production and at the same time make sure that they're not doing any harm to the environment. More focus will be on the factors like nutrients, nitrogen and phosphorus and water quality. Increasing crop productivity through breeding and genetics, many such aspects has come from research and development that allow agriculture to persist and be efficient. The outcome of the proposal is to help farmers to adapt management strategies that will protect the environment. More scientists are focusing now on nutrient mass budgets for farms. In which the following questions can be answered. Agriculture has advanced tremendously over the last 100 years. In fact it's been keeping up with food production at a very high rate. In fact, some, some scientists feel that agriculture already produces enough food to feed the world, but of course, but of course there are issues and problems with food availability, food, low food or agricultural production practices, preservation and transportation, and probably more that one can think of that hinder many people in this world from getting adequate food. Getting pressure on agriculture on farmers to increase this food production and at the same time make sure that they're not doing any harm to the environment. To focus on nutrients, nitrogen and phosphorus and water quality. An attempt was made to improve nutrients in the form through

---

[*]Associate professor, Sri Krishna College of Engineering and Technology, Coimbatore-641008, India (muzamilbashas@skcet.ac.in)

[†]Associate professor, SCOPE, VIT, India (dharmendrasingh@vit.ac.in)

[‡]Professor, Sri Krishna College of Engineering and Technology, Coimbatore-641008, India (janet.fredrick@gmail.com)

[§]Assistant Professor, Department of Information Technology, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering &Technology, Hyderabad, India, (svramasubbareddy1219@gmail.com)

[¶]Associate professor, Sri Krishna College of Engineering and Technology, Coimbatore-641008, India (sajeevrama@skcet.ac.in)

Bio fortification process, which are useful to human beings. Also, deficiency of micro nutrients in plants is addressed using plant breeding works [17].

1.1. Challenges in making farmers to adopt new kinds of practices.

- To get water out of a river and raise it up and put it in a canal to help irrigate their crops in between floods, flood seasons.
- Land Management
- Competition for good form land
- Harvesting techniques have grown to include very, large, efficient machines to harvest large areas of produce.
- Irrigation is a very important aspect to sustainable crop production causing the leeching of nutrients.
- pest management, new varieties of crop plants have all played a role
- protecting soil and minimizing nutrient loses from farms.

Competition with urbanization will remove good agriculture land. Available productive land has grown only by 8% while food production doubled between 1967 and 2017. Land availability is going to be a challenge for producers. Obviously farmers compete with urban development and expansion. And so, the value of the land for those alternative uses is going to be very important. The average age of farmers in the India is approaching 60 years. And so often wonder how many farms will stay as family farms under the ownership as farmers become older. What happens to those farms in the years as farmers approach retirement. While the availability land, available availability of good agriculture production land has only increased about 8 to 10% globally.

1.2. Sustainable Development Goals.
The Sustainable Development Goals (SDGs) are a part of the 2030 agenda. This makes the SDGs relevant for every person, country, and company on Earth. The total interconnection between the goals and the necessity to move forward on all goals at the same time. But of course, there are companies who have special qualifications and special issues. Business community now arguing for higher standards and higher national ambitions in order to push forward the technologies and products. The overarching goal of the SDGs taken as a whole is human welfare. That is to improvement of the human condition. Indeed, the first eight goals, no poverty, zero hunger, good health and well-being, quality education, gender equality, clean water and sanitation, clean affordable energy, and decent work and economic growth, deal directly with humans and their immediate condition. The remaining nine goals address the living and non-living infrastructure or resources and social structures that one way or another, provide support for the maintenance and improvement of the human condition. Achieving any one of these 8 goals will have positive effects on most. All of these welfare goals are however also clearly interlinked with one or more of goals 9 to 17, which focus on environment and infrastructure issues. Climate change, goal 13 interacts with goal 1, poverty, as it is generally speaking, the poorest that are most impacted by climate change. Climate change makes it harder to escape poverty. Similarly, inequality plays a role in SDG 3, access to healthcare, SDG 4 access to education, SDG 5 gender equality, SDG 7 access to energy, SDG 8 access to decent work, and so on. Addressing inequality is however so essential to achieving sustainable development that inequality has earned an SDG in its own right. SDG 10, focuses on reducing inequality in all of its forms but where many of the indicators used to assess progress against the goal deal with income inequality. While SDG 10 focuses primarily on income inequality, which is the difference in annual income received by individuals or countries per year. The research happened on sustainable indicator and its advantages are presented in Table 1.1.

However, the direct relationship between income or total wealth and well-being breaks down at higher levels of income. In other words, income or wealth are not in themselves adequate metrics for assessing societal development or human well-being. This is, of course, one of the reasons why a target in SDG 17 is to develop metrics that can supplement traditional economic metrics for assessing societal development. Wealth inequality is largely driven by the unequal ownership of capital in its many forms, not just financial capital, but also human, natural, physical, and social capital all of which contribute to human opportunity and well-being. The role of business in the world achieving the SDGs are classified as three key roles is: First of course business has to be responsible, make sure that our own operations and our supply chain up their performance on all the important areas, from environmental issues to people related issues. The second is to innovate and deliver solutions to environmental or social problems. So that one can deliver a positive contribution to the world,

TABLE 1.1
Sustainable Indicators

Author Details	Problem Addressed	Names of the Indicators	Future Scope
[9]	Introduced monitoring tool on PRIMA research	Multidimensionality poverty Index, Agriculture value added, Crop water productivity, Amount of agriculture residuals used for energy purpose.	Make use of Technology in traditional agriculture practices.
[10]	Designed a conceptual framework for selecting appropriate indicators	Multidimensionality poverty Index, Agriculture value added, Crop water productivity	Need of integrating agriculture and policy maker for better decision making.

with high impact, at speed and scale.

The contributions made in the present work are:

1. Literature review on sustainable indicators and improvement in yield responses.
2. Implements both the versions of Thomson's model and compared the result with the segmentation, Random forest model.
3. Analyzed the model performance with respect to yield responses
4. Error (RMSE and MAE) in prediction yield is estimated and proved that Random forest algorithm is provided good results.

The organization of the present work is as follow: In Introduction, the challenges being faced in adopting sustainable agriculture is addressed along with the importance given to agriculture in achieving goals and indicators used. In Literature section, the past research work carried out in improving the methods of agriculture protecting the environment is discussed. In Methodology section, the work carried out in the present research work is described in detail. In Result and discussion, the experimental details and discussion on result obtained is described with comparison. Finally, the conclusion and future work section, describe the objective addressed in the present work and the steps followed in achieving the goal of the work as discussed in introduction section.

2. Literature Study. Farmers understand that protecting the soil and conserving is critical to the success of the farm. So, soil conservation practices are extremely critical in that regard. sustainability and particularly sustainable agriculture can held with attention to both water quality and nutrients. Farmer need both of those inputs nutrients and chemicals that might impact water quality. The organic material on the immediate surface of the earth that serves as a natural medium for the growth of land plants [1]. The role of soil in environmental quality is high. Soil properties affecting water management (texture, hydraulic conductivity, water-holding capacity and natural drainage) [2] and nutrient management (organic matter content, soil PH, Cation exchange capacity and coatings on sand grains) [3]. Soil texture can be defined as relative proportion of sand, silt and clay in a mineral soil. Texture affects the amounts of water and nutrients a soil can hold [4]. Soil organic matter includes anything that was once alive from freshly deposited plant residues to highly decomposed organic matter. Can range from less than 1% of the soil by weight to nearly 100%. Mixture of living organisms recently dead, decomposing and stable materials. It adds water and nutrient holding capacity to the soil. It get lost rapidly by Oxidation during warm and humid climates. Oxidation also adds nutrients to the soil. Soils that tends to stay wet also tend to have more organic matter [5]. Hydraulic conductivity is the ability of the soil to transmit water when saturated. For sandy soil high hydraulic conductivity as a result they drain fast. Similarly, clayey soil have low hydraulic conductivity as a result water drains slowly and these soils stay wet longer than sandy soils [6]. Water Holding Capacity (WHC) is another important property of soil. The amount of water a soil can hold against gravity. Related to the proportion of silt, clay and organic matter in the soil. So, sandy soils have low WHC. It can affect irrigation management i.e., smaller irrigations on soils with low WHC [7]. Cation Exchange Capacity (CEC) is a measure of the ability of the soil to hold positively

charged nutrients called cations (calcium, magnesium, potassium and ammonium-N against leaching). It is largely imparted by the clay and organic matter particles in the soil [8]. Mostly sandy soils have low quantities of clay and organic matter. So, that would be expected to have low CEC. In general the higher the CEC means greater fertility. The types of soils that encounter can be classified based on soil orders.

Nutrient mass balance is an analysis of the quantities of nutrients brought onto the farm and those leaving the farm. Especially those that end up in the environment. To help the farmer maximize the efficiency of the nutrients that the farmer purchases and brings onto the farm to grow the crops. The goals here are to return on investment in fertilizer and minimize losses to the environment. Nutrients can be imported to the farm (fertilizer, manure, Animal feed and Inflow Rivers and streams). It can be recycled on the farm (or) urban landscape. Buildup of the same can be happen. They are lost to the environment different from export (Runoff, Leaching and Gaseous). Mass balance approach, aim is to quantify these pools and to find out how much nitrogen would be associated with the crop that's taken off the farm or fed to the animals. The amount of nutrients that might be returned to the soil, through crop refuse. The amounts of nutrients that might be applied, to the fields and fertilizer and manure, and quantify the nutrients that might potentially be lost [28, 29, 30].

2.1. Soil Management Techniques.

2.1.1. Soil tillage advantages.
1. Turn under, incorporate organic matter to decompose
2. Mix soil and its constituents, organic matter, nutrients.
3. Turn under weeds
4. Turn under disease organisms insects.
5. Help dry out the soil for earlier planting
6. Prepare the seed bed

There are also disadvantages:
1. Loosens the soil and exposes soil to drying, can lead to wind and water erosion
2. Exposes soil organic matter to oxidation can lead to soil loses organic matter content

Cover crops is the solution for the soil management. In which, one can reduce soil erosion (wind (or) runoff). It can Recover unused nutrients, add organic matter to the soil. Survey on Sustainable Agriculture and its practices is presented in Table 2.1.

3. Methodology. The dataset considered in our experiment is used in yield prediction based on historic yield and weather information [20]. Implemented both the versions of Thomson model and compared the result with segmentation model, Random forest [21]. The complete description of the dataset used in the project is described [22]. The steps carried out in the present work to achieve the outcome of the listed objectives are listed in Fig. 3.1.
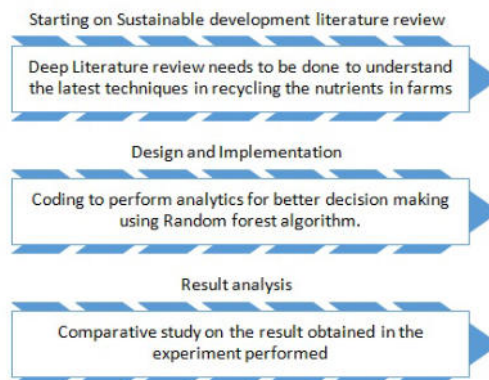

FIG. 3.1. *Steps followed to achieve the outcome*

TABLE 2.1
Survey on Sustainable Agriculture and its practices

Author Details	Contribution	Features	Gaps
[11]	Suggested three principles to be used in Circular Economy1. Waste is Food2. Make use of renewable energy3. Taking inspiration from nature	1. Food Production2. Food Distribution3. Food Consumption	Need to redesign food systems in line with circular principles to present a feasible solution.
[12]	Provides overview on the characteristics of regulations of organic farming and agroecology	1. Soil tillage2. Soil fertility3. Crop and cultivar choice4. Crop rotation5. Intercropping6. Management of landscape elements and habitats7. Pest, disease and weed management8. Water quantity and quality9. Agroforestry.	Focus on the restriction of external inputs and the limitation of chemical inputs
[13]	Describing the need for a link between agronomy and education in sustainability	To overcome epistemological boundaries between the natural and social sciences	Strong contrast between sustainable intensification in high-external-input agriculture
[14]	Addressees the possible ways to improve natural capital by generating more food.	1. Management of application of pesticides2.Agroecological system and habitat redesign	There are some regional scale exemplars of positive policy practice.
[15]	Cropland monitoring using time series analysis on satellite image	Random forest algorithm is used in image classification and achieved 95% accuracy	Make use of Google Earth Engine (GEE) platform to handle peta bytes of data.
[16]	Focused on developing precision agriculture and precision conservation to maintain sustainable development of agriculture at field level. Developed Geo-informatics sustainable agriculture framework	1.Modern form Management techniques2. Natural resources conservation service.3. Both agriculture people and policy makers are considers in decision making	Deployment of new virtual agriculture practices in finding out better environment outcome with the help of IOT, drones and Data servers.
[23]	Focused on crop management	Crop water productivity	Implications and roles of public and private sectors
[24]	Listed out Plant Growth Promoting Microorganisms (PGPM)	Efficient delivery system of PGPM	Amount of agriculture residuals used for energy purpose
[25]	Addressed key precision agriculture milestones like Global Navigation Satellite Systems (GNSS), Global Positioning System (GPS), Variable Rate Technology (VRT).	Amount of agriculture residuals used for energy purpose	Estimates the VRT adoption may exceed 50%

The reason for selecting random forest algorithm in predicting the yield production value is, as this algorithm allows the developer to incorporate the human drafted rules in most easier way. The implementation details of the thomson model is compared with the recent study made by [22]. In the research work [26], [27] the author had used optimization algorithm to improve the weights of each attribute in improving the accuracy of classification algorithms.
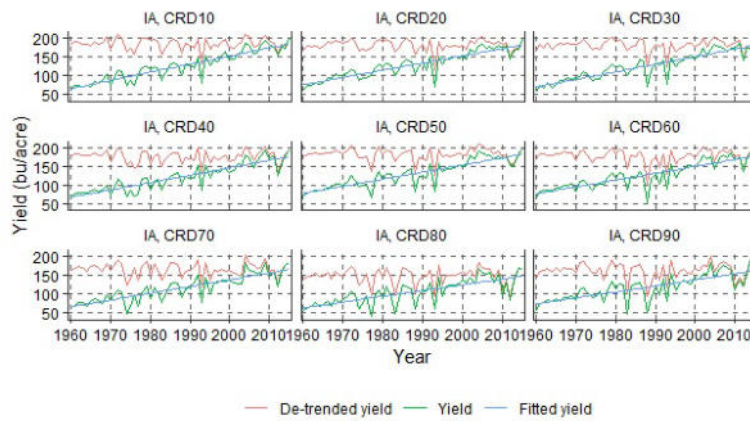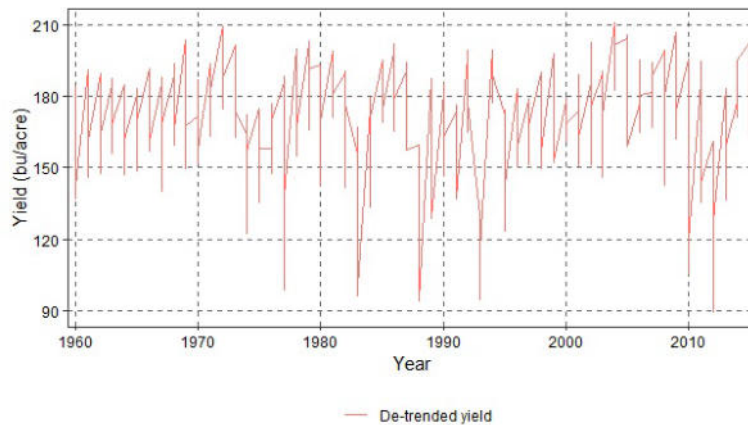
FIG. 4.1. *Plot on different types of yield*

FIG. 4.2. *Effect of yielding from 1960 to 2010.*

4. Result and Discussion. To better understand the behavior of the attributes and its distribution the is plotted in Fig. 4.1 [18].

Van Eeuwijk et al. made in 2019 a similar kind of study from the year 1960 to 2000. The dataset is updated using cross validation technique with 10 folds and thereby improved the number of instances from the dataset collected. Fig. 4.2 helps in interpreting the impact of yield from 1960 to 2010.

The mean calculated mean value of the Temperature year wise is plotted to understand the impact of temperature on yielding crops like corn and soybeans Fig. 4.3. The performance of Random forest using Mean Square Error is estimated and plotted in Fig. 4.4. The comparative performance of the models implemented in the present work is plotted in Fig. 4.5. The Error (MSE) in prediction and its interpretation is done with the help of Fig. 4.6.

5. Conclusion and Future scope. The Goal is to develop the social, environmental, and the economic aspects of possible solutions to minimize the agricultural footprint, and become more sustainable. . Implemented both the versions of Thomson model and compared the result with segmentation model, Random forest. Root Mean Square Error and Mean Absolute Error are used as evaluation metrics in estimating the performance of models implements and stated that Random forest algorithm is providing 0.07 (RMSE). The outcome of the present research work helps farmers in adopting best management practices and trying to give
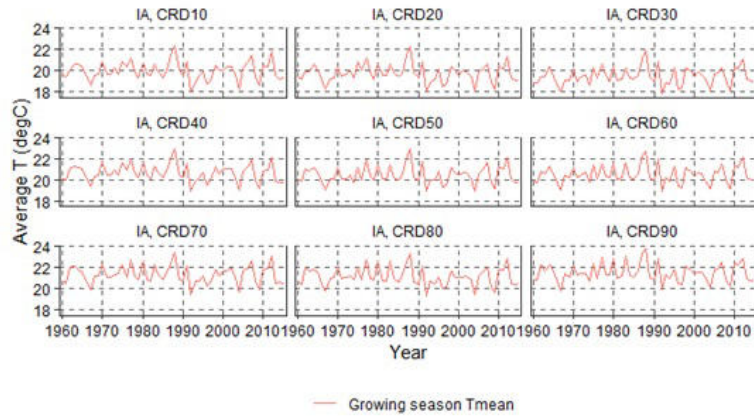
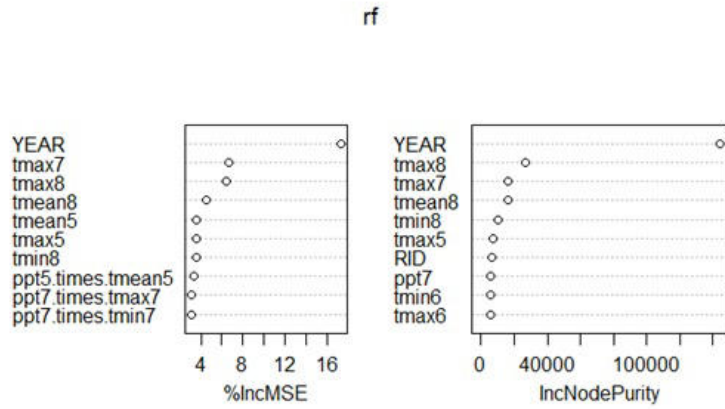FIG. 4.3. *Plot on Season factor on Temperature.*
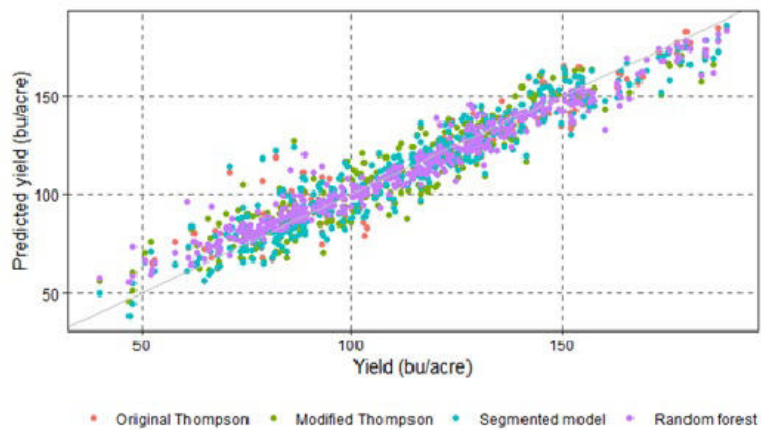

FIG. 4.4. *Performance of Random Forest*

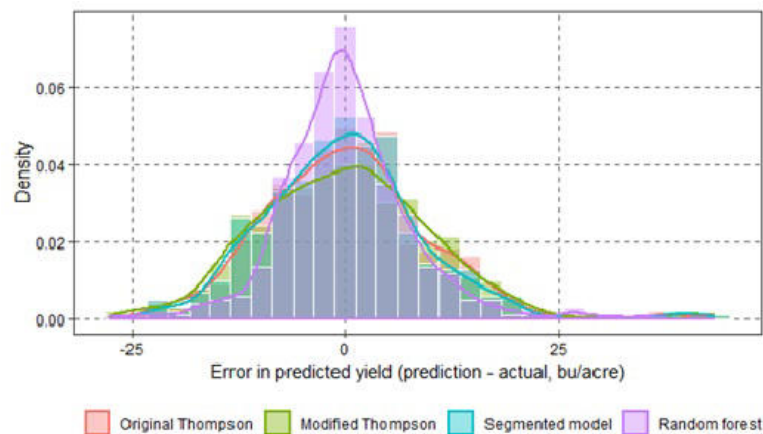FIG. 4.5. *Comparison on yield prediction Level*

FIG. 4.6. *Comparison on Error in yield prediction*

them the economical and technical support in making easier for them to adopt best management practices. The outcome of the present work helps farmers to manage/monitor more profit and production without effecting the environment. The farmer can maintain the cycle of nutrients with the help of environment. To put forward sustainable use of natural resources and inputs. In future, the parameters like Soil Texture, Hydraulic Conductivity, Water Holding Capacity, Cation Exchange Capacity to be considered in our experiments to support sustainable agriculture.

REFERENCES

[1] BISWAS, H., RAIZADA, A., MANDAL, D., KUMAR, S., SRINIVAS, S., & MISHRA, P. K. *Identification of areas vulnerable to soil erosion risk in India using GIS methods* Solid Earth, 6(4), 1247, 2015

[2] OMONDI, M. O., XIA, X., NAHAYO, A., LIU, X., KORAI, P. K., & PAN, G., *Quantification of biochar effects on soil hydrological properties using meta-analysis of literature data* Geoderma, 274, 28-34, 2016.

[3] OLORUNFEMI, I. E., &FASINMIRIN, J. T. *Land use management effects on soil hydrophobicity and hydraulic properties in Ekiti State, forest vegetative zone of Nigeria* Catena, 155, 170-182, 2017.

[4] KOHLER, J., ROLDÁN, A., CAMPOY, M., &CARAVACA, F. *Unraveling the role of hyphal networks from arbuscularmycorrhizal fungi in aggregate stabilization of semiarid soils with different textures and carbonate contents* Plant and Soil, 410(1-2), 273-281, 2017.

[5] STIRLING, G., HAYDEN, H., PATTISON, T., &STIRLING, M. *Soil health, soil biology, soilborne diseases and sustainable agriculture: A Guide* Csiro Publishing, 2016.

[6] NAVARRETE, J. L. *Evaluation Of Recycled Gypsum Application Dosages To Enhance The Water Infiltration Rate At Water Retention Ponds*, 2018.

[7] MOHAMED, B. A., ELLIS, N., KIM, C. S., BI, X., &EMAM, A. E. R. *Engineered biochar from microwave-assisted catalytic pyrolysis of switchgrass for increasing water-holding capacity and fertility of sandy soil* Science of the Total Environment, 566, 387-397, 2016.

[8] LITTMANN, R. J. *U.S. Patent Application No. 15/092,079*, 2016.

[9] SALADINI, F., BETTI, G., FERRAGINA, E., BOURAOUI, F., CUPERTINO, S., CANITANO, G., & BIDOGLIO, G *Linking the water-energy-food nexus and sustainable development indicators for the Mediterranean region* Ecological Indicators, 91, 689-697, 2018

[10] HAK, T., JANOUŠKOVÁ, S., & MOLDAN, B. *Sustainable Development Goals: A need for relevant indicators* Ecological Indicators, 60, 565-573, 2016.

[11] DUNCAN, J. A. B., & PASCUCCI, S. *Circular solutions for linear problems: Principles for sustainable food futures.* Solutions, 7(4), 58-65, 2016

[12] MIGLIORINI, P., & WEZEL, A. *Converging and diverging principles and practices of organic agriculture regulations and agroecology. A review* Agronomy for sustainable development, 37(6), 63, 2017.

[13] STRUIK, P. C., & KUYPER, T. W. *Sustainable intensification in agriculture: the richer shade of green. A review* Agronomy for Sustainable Development, 37(5), 39, 2017.

[14] PRETTY, J. *Intensification for redesigned and sustainable agricultural systems.* Science, 362(6417), eaav0294, 2018.

[15] GUMMA, M. K., THENKABAIL, P. S., TELUGUNTLA, P. G., OLIPHANT, A., XIONG, J., GIRI, C., & WHITBREAD, A. M. *Agricultural cropland extent and areas of South Asia derived using Landsat satellite 30-m time-series big-data using*

random forest machine learning algorithms on the Google Earth Engine cloud GIScience & Remote Sensing, 1-21, 2019.

[16] DELGADO, J., SHORT, N. M., ROBERTS, D. P., & VANDENBERG, B. *Big Data Analysis for Sustainable Agriculture* Frontiers in Sustainable Food Systems, 3, 54, 2019.

[17] BOUIS, H. E., & SALTZMAN, A. *Improving nutrition through biofortification: a review of evidence from HarvestPlus, 2003 through 2016* Global food security, 12, 49-58, 2017

[18] TOGLIATTI, K., ARCHONTOULIS, S. V., DIETZEL, R., PUNTEL, L., & VANLOOCKE, A. *How does inclusion of weather forecasting impact in-season crop model predictions?* Field Crops Research, 214, 261-272, 2017.

[19] VAN EEUWIJK, F. A., BUSTOS-KORTS, D., MILLET, E. J., BOER, M. P., KRUIJER, W., THOMPSON, A., & MULLER, O *Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. Plant science*, 282, 23-39. 2019.

[20] THOMPSON, L. M. *Weather and technology in the production of corn and soybeans*, 1963.

[21] SUBUDHI, A., DASH, M., & SABUT, S. *Automated segmentation and classification of brain stroke using expectation-maximization and random forest classifier* Biocybernetics and Biomedical Engineering, 40(1), 277-289. 2020.

[22] BULLOCK, D. W. *The Influence of State-Level Production Outcomes Upon US National Corn and Soybean Production: A Novel Application of Correlated Component Regression* No. 1187-2019-1851, 2017.

[23] YOST, M. A., SUDDUTH, K. A., WALTHALL, C. L., & KITCHEN, N. R. *Public–private collaboration toward research, education and innovation opportunities in precision agriculture. Precision Agriculture, 20(1), 4-18.*

[24] MA, Y *, (2019). Seed coating with beneficial microorganisms for precision agriculture. Biotechnology advances, 37(7), 107423.*

[25] LOWENBERG-DEBOER, J., & ERICKSON, B. *(2019). Setting the record straight on precision agriculture adoption. Agronomy Journal*, 111(4), 1552-1569.

[26] BASHA, S. M., RAJPUT, D. S., & VANDHAN, V. (2018). Impact of gradient ascent and boosting algorithm in classification. *International Journal of Intelligent Engineering and Systems (IJIES)*, 11(1), 41-49.

[27] BASHA, S. M., & RAJPUT, D. S. (2018, November). Evaluating the Importance of each Feature in Classification task. *In 2018 8th International Conference on Communication Systems and Network Technologies* (CSNT) (pp. 151-155). IEEE.

[28] BALIARSINGH, S. K., VIPSITA, S., GANDOMI, A. H., PANDA, A., BAKSHI, S., & RAMASUBBAREDDY, S. (2020). Analysis of high-dimensional genomic data using MapReduce based probabilistic neural network. *Computer methods and programs in biomedicine*, 195, 105625.

[29] ATTILI, V. R., ANNALURI, S. R., GALI, S. R., & SOMULA, R. (2020). Behaviour and Emotions of Working Professionals Towards Online Learning Systems: Sentiment Analysis. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, 12(2), 26-43.

[30] RAVINDRANATH, V., RAMASAMY, S., SOMULA, R., SAHOO, K. S., & GANDOMI, A. H. (2020, July). Swarm Intelligence Based Feature Selection for Intrusion and Detection System in Cloud Infrastructure. *In 2020 IEEE Congress on Evolutionary Computation (CEC)* (pp. 1-6). IEEE.