FEATURE EXTRACTION OF AGRICULTURE CROP RECOMMENDATION USING ADVANCED MACHINE LEARNING GENERATIVE ALGORITHMS

[1]Devara Nagasri, [2]Dr.T.Srinivasulu, [3]U. Venakat Rao, [4]V. Rupa

[1]Assistant Professor, Department of CSE (AI&ML), Vignan's Institute of Management and Technology for Women, Kondapur, Hyderabad.

[2]Professor, Department of Mathematics, Vignan's Institute of Management and Technology for Women, Kondapur, Hyderabad.

[3]Assistant Professor, Department of CSE (DS), Vignan's Institute of Management and Technology for Women, Kondapur, Hyderabad.

[4]Assistant Professor, Department of Information Technology, Vignan's Institute of Management and Technology for Women, Kondapur, Hyderabad.

ABSTRACT:

Advancements in machine learning algorithms have revolutionized the domain of precision agriculture, enabling data-driven decisions for crop selection and boosting overall productivity. This research aims to develop an intelligent crop recommendation system leveraging machine learning algorithms to suggest suitable crops based on historical productivity data and prevailing seasonal conditions. The proposed system utilizes a diverse set of features such as soil characteristics, climate data, historical crop performance, and geographical factors to capture the complexities of crop-environment relationships. The agricultural sector plays a critical role in providing food security and sustenance for the growing global population. However, the success of agricultural practices heavily relies on the selection of appropriate crops tailored to specific regions and seasonal conditions. In recent years, The machine learning models employed include AdaBoost, Naive Bayes, K-Nearest Neighbors , Logistic Regression, which will be trained on a comprehensive dataset of past crop yields and environmental parameters. The dataset will be collected from diverse agricultural regions across different seasons, ensuring the robustness and adaptability of the developed models. The Research outcome is expected to empower farmers with valuable insights to make informed decisions about crop selection, leading to optimized resource utilization and improved productivity. By harnessing the power of machine learning, this research aspires to contribute to sustainable agriculture practices, economic growth in rural communities, and the overall advancement of the agricultural sector.

Keywords: *Agriculture Crop Recommendation, Precision Agriculture, Machine Learning, AdaBoost, Naive Bayes, K-Nearest Neighbors , Logistic Regression.*

INTRODUCTION:

Tamil Nadu being 7th largest area in India has 6th largest population. It is the leading producer of agriculture products. Agriculture is the main occupation of Tamil Nadu people. Agriculture has a sound tone in this competitive world. Cauvery is the main source o f water. Cauvery delta regions are called as rice bow l of Tamil Nadu. Rice is the major crop grown in Tamil Nadu. Other crops like Paddy, Sugarcane, Cotton, Coconut and groundnut are grown. Bio-fertilizers are produced efficiently. Many areas Farming acts as major source of occupation. Agriculture makes a dramatic impact in the economy o f a country. Due to the change o f natural factors, Agriculture farming is degrading now-a-days. Agriculture directly depends on the environmental factors such as sunlight, humidity, soil type, rainfall, Maxim um and Minimum Temperature, climate, fertilizers, pesticides etc. Knowledge of proper harvesting o f crops is in need to bloom in Agriculture. India has seasons of

- W inter which occurs from December to March
- Summer season from April to June
- Monsoon or rainy season lasting from July to September and
- Post-monsoon or autumn season occurring from October to November.

Due to the diversity of season and rainfall, assessment of suitable crops to cultivate is necessary. Farmers face major problem s such as crop management, expected crop yield and productive yield from the crops. Farm ers or cultivators need proper assistant regarding crop cultivation as now-a-days many fresh youngsters are interested in agriculture. Impact of IT sector in assessing real world problem is moving at a faster rate. Data is increasing day by day in field of agriculture. With the advancement in Internet o f Things, there are ways to grasp huge data in field o f Agriculture. There is a need o f a system to have obvious analyzes of data of agriculture and extract or use useful information from the spreading data. To get insights from data, it has to be learnt. Crop production depends on many agricultural parameters. Proposed work is based on the production of crops in previous years, crops can be recommended to the farmers. This kind of suggestions will make farmer to know that whether that particular is yielding a good production in recent years. Production o f crops may become less due to any crop disease, water problem and many other factors. While considering about the production, farm ers may get knowledge about which crop is in high volume in the market in that year. Based on this farmer can take decision o f trend on crops in recent years. Farmers will be given recommendation by

considering the season of crop production. Tamilnadu Agriculture Dataset of about 1,20,000 records were taken. It contains fields like crop year, crop name, District, Season, A reacultivated and production. Recommendations were given to user based on the production of crops, season when the crops cultivated. Recommender systems have lent its hands to users to choose items they like. Recommendation system is the approach to provide the suggestions to the users of their interest This can be practiced for agricultural use too. Based upon the factors o f agriculture, farm ers are given with ideas for their cultivation process. New techniques to increase crop cultivation can also be recommended. Pesticides, fertilizers can also be recommended. Hybrid Recommender system built by AgajiIorshase to recommend agricultural products solves issues like serendipity, ratio diffusion and ramp-up. M any crop prediction yield models have been developed. Clustering approaches such as k-means, km eans++ are used to perform grouping o f data as clusters to predict crop yield is used. Tripathy et al., provided a system to have management of pesticides for crop cultivation using data mining process. Essential parameter for agriculture analysis is nature o f soil. Diverse varieties o f soil are available in this India. Crops are cultivated depending on the type of soil in the land. The role o f soil in im proving crop cultivation is discussed. Data mining techniques are applied to analyze the soil parameter. Naive Bayes techniques are applied which produces m ore reliable results in analyzing red and Black soil. Impact of parameters o f agriculture in crop management is studied to improve productivity [5]. Neural networks, soft computing, big data and fuzzy logic methods are being used to examine the agricultural factors. Pritam Bose developed a SNN model to have a spatiotemporal analysis with crop estimation. Prediction of crops was done according to farmer's experience in the past years. Although farmer's knowledge sustains, agricultural factors has been changed to astonishing level. There comes a need to indulge engineering effect in crop prediction. Data mining plays a novel role in agriculture research. This field uses historical data to predict; such techniques are neural networks, K-nearest Neighbor. K-means algorithm does not use historical data but predicts based on-computing centers o f the samples and forming clusters. Computational cost of algorithm acts as a major issue. Use of Artificial Neural Network is a boon to agriculture field which computes accurately even with more input.

LITERATURE SURVEY:

In methodology of remote sensing was demonstrated and validated. The author also represented rice yield estimation. For this authors developed a system that considered remote sensing data

based on SAR and MODIS data as input [1]. This developed system, based on crop growth model, generates dimensional explicit inputs for rice. Further, the study considered the Red River Delta in Vietnam for rice yield estimation. The study considered eight rice producing provinces of area under consideration. In case of the MODIS (MOD13Q1 and MYD13Q1 products), a combined time series of the TERRA and AQUA 16-day Composite Vegetation Indices (VIs) distributed by the NASA were used. The time-series images that were selected were at 250-m resolution. MODIS mentioned products include indices like NDVI and EVI spectral indices. On the other hand, C-band VV and VH polarization (SAR data distributed by ESA) with a regular repeat cycle of 12-days were used. The spatial resolution of 20-m were used. Sentinel-1 SLC time-series is converted into terrain-geocoded values in the pre-processing of the SAR sentinel-1 data [2]. The steps in the pre-processing include image calibration, speckle filtering (time-series), radiometric terrain correction and normalization, filtering of atmospheric attenuation. Here, the authors proposed an algorithm for the Rice Crop detection. The proposed algorithm was based on the analysis of time-series of NDFI and EVI spectral indices. The authors in utilized time-series dual polarized (VV/VH) C-band SAR imagery from the Sentinel-1A and 1B satellites. The SAR imagery captured from a small region of central North Dakota was analyzed for the Crop analysis [3]. All the Sentinel-1 images (Wide Swath Mode, ascending orbit, Level 1 GRD format) captured over a period of April 2016 to November 2016 were first pre-processed on SNAP. Authors suggested few steps in the pre-processing, that includes orbit corrections, followed by multi-looking the images to 100m pixels, and applying the terrain correction [4]. In the research, the authors represented a classification algorithm, in which individual pixel was compared with a model of average crop backscattered response. Taking the least difference from the model, each and every single pixel was classified as the specific crop [5]. Further, the authors analyzed the classification accuracy based on few parameters. These parameters included the iterations in model building, influence of polarization, and number of training fields [6]. The proposed approach achieved overall accuracies of above 90%, using both VV and VH polarizations independently or in combination. Authors performed two different types of analysis a). Based on twenty training field trails, b). with full CDL layer (excluding training pixels) [7]. In the former one, overall accuracy of 85%- 96% was achieved and in the latter one, it was only 65%-74%. Thus the former case, irrespective of the polarization selected or count of iterations used, produced more classification accuracy than that of the latter one [8]. The result of the classification algorithm proposed by the author was compared with the most common, full time–

series of VH-polarized images, using a RF classifier. This classifier was implemented using Random Tree tool in ArcGIS Desktop 10.5. RF algorithm is a version of the Breiman‟s (2001) RF algorithm known for its robustness and accuracy [9]. The SAR sensor (Sentinel-1) allows an accurate chronological follow-up of agricultural crop growth. Implementing various deep learning techniques, author highlights the capability of Sentinel-1 radar imagery for mapping agricultural land cover. Multi-temporal Sentinel-1 data was improved by applying temporal filtering to minimize noise, though preserving the fine structures present in the images. As per the analysis and the results proposed by the author, two deep recurrent neural networks (RNN)-based classifiers gives the better results than that of the machine learning approaches (random forest, support vector machines and K-nearest neighbours). The authors preferred RNN over classical machine learning approaches as per the results retrieved in his research [10]. The sentinel-1 satellite data has already been considered by many researchers for research in the area of agricultural with remarkable results. The sentinel-1 data of agricultural fields was processed to assess parameters such as type of crop , Green Area Index (GAI), plant height and vegetation water content, illustrated the analyses and processing of the Sentinel-1 images for the estimation of Rice crop acreage. Author suggested pre-processing of the Sentinel-1 images before performing estimation and correlation in the area of agriculture. To process the SAR images author preferred SNAP, before utilizing it for a particular purpose. Author presented the pre-processing pipeline procedure illustrated in the SNAP graph as sequence of operations Read, Calibration, Speckle Filter, Terrain Correction, and Write. Further, author performed k-means clustering for the sampling of the field. Authors analyzed and processed the 8 OLI multi-temporal data of year 2013. Author downloaded the images of Northern Italy, Lombardy region, to be considered for his research. Overlapped region of the two WRS-2 paths (193 and 194), cloud cover less than 20%, was selected for the analysis purpose. Multi-temporal spectral indices (NDFI, RGRI, and EVI) used as input for the supervised algorithm.

METHODOLOGY:

Step 1: Raw data and Weather Statistics This is the first step in the crop recommendation process where we collect agriculture data. Agriculture data collected from the region which contains agriculture parameters, crop details, farmers details and yield details. Agriculture parameters include rainfall, temperature, soil features such as PH, nitrogen, potassium, iron etc…

Step 2: Extract and Segment Data (Data Preprocessing) Here agriculture data analyzed and only relevant data extracted. The data required for processing extracted and segmented according to the different regions. Required data extraction is done because entire agriculture data not required for processing and if we input all data, it requires too much of time for processing, so data processing is done.

Step 3: Train Data Once required data extracted and segmented, we need to train the data, train means converting the data into the required format such as numerical values or binary or string etc.. conversion depends on the algorithm type. Step 4: Supervised Learning ML concerns with construction and study of system that can learn from data. Naive Bayes Algorithm "Naive Bayes Algorithm" is used for crop recommendation because of the following reasons;

1. efficient classifier
2. Works fine for less number of parameters as well as more number of parameters.
3. Works fine for small data-set as well as big data-set.
4. more accurate results

Step 5: Crop Recommendation (Priority wise) Here suitable crops recommended for the farmers which may yield high profits. Naive Bayes algorithm generates outputs (crop recommendations) based on the priority wise.

Step 6: Location and Year Based The crop recommendation is done based on the region wise as well as year wise.

Step 7: Results Recommending suitable and high profit crops for the farmers and recommendation is done based on the priority wise. Here high probability crops are extracted and sorted and top 3 crops recommended for the farmers.

Step 8: Visual Representation Crops recommended for the farmers on GUI. When users gets login to the application system recommends suitable and high profit crops for the farmers on a GUI.

Machine Learning Algorithms

As fertilizer is composed of chemicals and harmful for the plants and society, this procedure has to be changed to find out the suitable crop for the soil. The nutrient requirements of soils using in agriculture for crop production are examined to determine the need of a data mining system to detect the crops suite for the soil by analyzing the features of the soil. This helps in giving information to the farmer about suitable crops that regulate the nutrition levels of the soil. Farmer can know which crop is suitable for that area and also he will get 100% yield on

that year. System used by administrator, staffs and farmers. System is a browser based application which predicts crops based on the soil test results.

System makes use of data mining technique for crop prediction. System makes use of previous dataset for the crop prediction System makes use of navie bayesian algorithm for crop prediction System generates accurate results based on the size of the dataset. Administrator of the system creates the staffs and sets the unique Id and password for each staff.

ADABOOST: AdaBoost short for Adaptive Boosting is an ensemble learning used in machine learning for classification and regression problems. The main idea behind AdaBoost is to iteratively train the weak classifier on the training dataset with each successive classifier giving more weightage to the data points that are misclassified. The final AdaBoost model is decided by combining all the weak classifier that has been used for training with the weightage given to the models according to their accuracies. The weak model which has the highest accuracy is given the highest weightage while the model which has the lowest accuracy is given a lower weightage. A single weak model may not be enough for our complex problems such cases we aggregate various weak models to make a powerful and more accurate model for our problem this process of aggregating several small problems to create a strong model is what we do in boosting. Boosting is an ensemble modeling technique that attempts to build a strong classifier from the number of weak classifiers. It is done by building a model by using weak models in series. Firstly, a model is built from the training data. Then the second model is built which tries to correct the errors present in the first model. This procedure is continued and models are added until either the complete training data set is predicted correctly or the maximum number of models are added.

NAIVE BAYES: Naïve Bayes Classifier Algorithm Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles. Why is it called Naïve Bayes? The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, Which can be described as: Naïve: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other

features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other. Bayes: It is called Bayes because it depends on the principle of Bayes' Theorem. Many crop prediction yield models have been developed. Clustering approaches such as k-means, kmeans++ are used to perform grouping o f data as clusters to predict crop yield is used. Tripathy et al., provided a system to have management of pesticides for crop cultivation using data mining process. Essential parameter for agriculture analysis is nature o f soil. Diverse varieties o f soil are available in this India. Crops are cultivated depending on the type of soil in the land. The role of soil in improving crop cultivation is discussed. Data mining techniques are applied to analyze the soil parameter.

K-NEAREST NEIGHBORS : K-nearest Neighbor. K-means algorithm does not use historical data but predicts based on-computing centers o f the samples and forming clusters. Computational cost of algorithm acts as a major issue. Use of Artificial Neural Network is a boon to agriculture field which computes accurately even with more input. An architecture developed in uses input; selects needed features; classification and association rule mining is applied and visualized. Bangladesh has its high production as rice. Statistical Methodologies has been used to predict its crop production. Shakil Ahamed applied clustering and classification techniques on 15 districts of Bangladesh to recommend for yield and planting of crops. Factors implementing crop yield were considered. They are a. Environmental factors-rainfall, humidity, Minimum and maximum temperature b. Biotic factors-soil pH and salinity c. Area factors-irrigated or cultivated.

LOGISTIC REGRESSION: Impact of parameters o f agriculture in crop management is studied to improve productivity . Neural networks, soft computing, big data and fuzzy logic methods are being used to examine the agricultural factors. Pritam Bose developed a SNN model to have a spatiotemporal analysis w ith crop estimation. An automatic system to gather the information about soil nature, weather conditions was developed with clustering techniques to extract the knowledge and use it by farmers in crop cultivation. Prediction of crops was done according to farmer’s experience in the past years. Although farmer’s knowledge sustains, agricultural factors has been changed to astonishing level. There comes a need to indulge engineering effect in crop prediction. Data mining plays a novel role in agriculture research. This field uses historical data to predict; such techniques are neural networks,

RESULT ANALYSIS:

Table1: Trained Datasets and Tested Results

S.No	Model Type	Accuracy
1	AdaBoost	98.5
2	K- Neighbor Classifier	94.6
3	Naïve Bayes	93.4
4	Logistic Regression	92.8

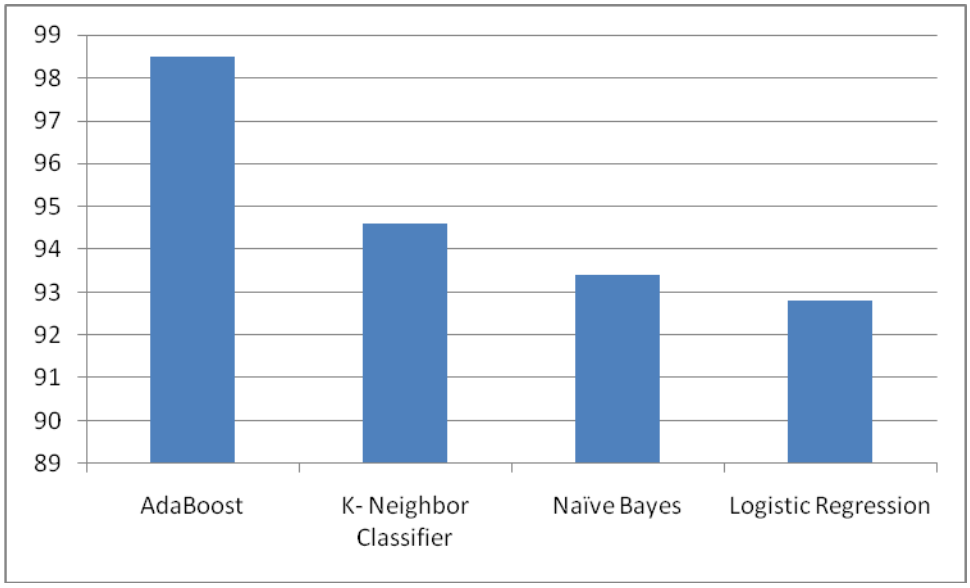

Figure2: Accuracy of ML Algorithms

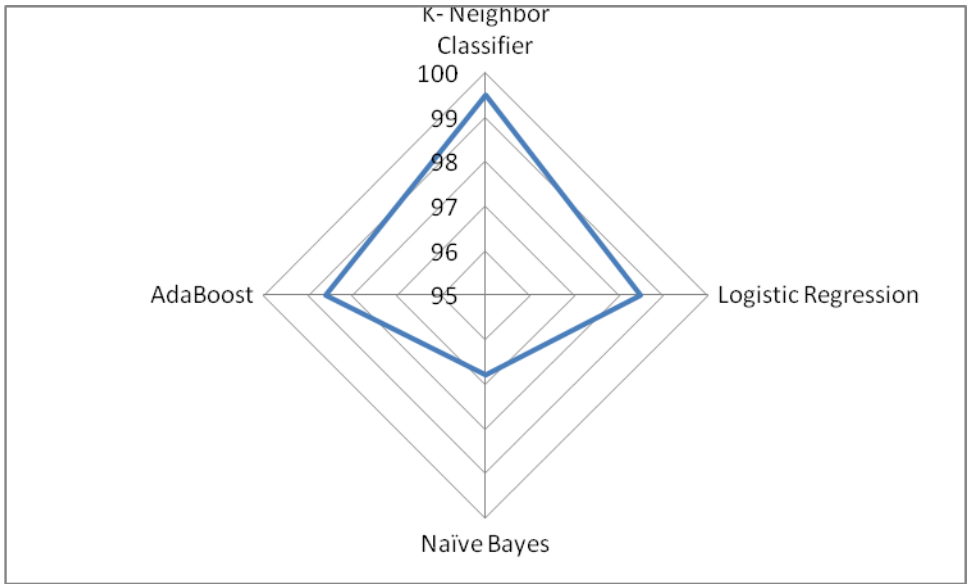

Figure2: Crop Recommendations

CONCLUSION

Farmers need assistance with recent technology to grow their crops. Proper prediction of crops can be informed to agriculturists in time basis. Many Machine Learning techniques have been used to analyze the agriculture parameters. Some o f the techniques in different aspects o f agriculture are studied by a literature study. Blooming Neural networks, Soft computing techniques plays significant part in providing recommendations. Considering the parameter like production and season, more personalized and relevant recommendations can be given to farmers which makes them to yield good volume of production. The pre-processing of the imagery data from satellite and Offline images can be performed using AdaBoost, Naive Bayes, K-Nearest Neighbors , Logistic Regression methods. Further, we need to study pre-processed images of the same polarization with different dates preferably at a gap of say two or three months. As different dates data to be studied for Variation analysis over time considerably over different areas needs to be studied.

REFERENCES

1. Li, L.; Wang, B.; Feng, P.; Liu, D.L.; He, Q.; Zhang, Y.; Wang, Y.; Li, S.; Lu, X.; Yue, C.; et al. Developing machine learning models with multi-source environmental data to predict wheat yield in China. Comput. Electron. Agric. 2022, 194, 106790.

2. Moysiadis, V.; Tsakos, K.; Sarigiannidis, P.; Petrakis, E.G.M.; Boursianis, A.D.; Goudos, S.K. A Cloud Computing web-based application for Smart Farming based on microservices architecture. In Proceedings of the 11th International Conference on Modern Circuits and Systems Technologies (MOCAST), Bremen, Germany, 8–10 June 2022; pp. 1–5.

3. Vashisht, S.; Kumar, P.; Trivedi, M.C. Improvised Extreme Learning Machine for Crop Yield Prediction. In Proceedings of the 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 27–29 April 2022; pp. 754–757.

4. Rashid, M.; Bari, B.S.; Yusup, Y.; Kamaruddin, M.A.; Khan, N. A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction. IEEE Access 2021, 9, 63406–63439.

5. Memon, R.; Memon, M.; Malioto, N.; Raza, M.O. Identification of growth stages of crops using mobile phone images and machine learning. In Proceedings of the International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), Quetta, Pakistan, 26–27 October 2021; pp. 1–6.

6. Malathy, S.; Vanitha, C.N.; Kotteswari, S.; Mohankkanth, E. Rainfall Prediction for Enhancing Crop-Yield based on Machine Learning Techniques. In Proceedings of the International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 9–11 May 2022; pp. 437–442.

7. Kumar, R.; Shukla, N.; Princee. Plant Disease Detection and Crop Recommendation Using CNN and Machine Learning. In Proceedings of the International Mobile and Embedded Technology Conference (MECON), Noida, India, 10–11 March 2022; pp. 168–172.

8. Sahu, P.; Singh, A.P.; Chug, A.; Singh, D. A Systematic Literature Review of Machine Learning Techniques Deployed in Agriculture: A Case Study of Banana Crop. IEEE Access 2022, 10, 87333–87360.

9. Swamy, S. Ranga, et al. "Dimensionality reduction using machine learning and big data technologies." Int. J. Innov. Technol. Explor. Eng.(IJITEE) 9.2 (2019): 1740-1745.

10. SWAMY, SIRISATI RANGA, and SRIDHAR MANDAPATI. "A RULE SELECTED FUZZY ENERGY & SECURITY AWARE SCHEDULING IN CLOUD." Journal of Theoretical & Applied Information Technology 96.10 (2018).

11. Sirisati, R. S. "Machine learning based diagnosis of diabetic retinopathy using digital fundus images with CLAHE along FPGA Methodology." Int. J. Adv. Sci. Technol.(IJAST-2005-4238) 29.3 (2020): 9497-9508.

12. SwamySirisati, Ranga, Mekala Srinivasa Rao, and Srinivasulu Thonukunuri. "Analysis of Hybrid Fusion-Neural Filter Approach to detect Brain Tumor." 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC). IEEE, 2020.