https://doi.org/10.56093/JISAS.V78I1.9
Available online at http://isas.org.in/isa/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 78(1) 2024 69–74

Identification of Paddy Stages from Images using Deep Learning

Himanshushekhar Chaurasia[1, 2, 3], Alka Arora[2], Dhandapani Raju[4], Sudeep Marwaha[2], Viswanathan Chinnusamy[4], Rajni Jain[5], Mrinmoy Ray[2] and Rabi Narayan Sahoo[4]

[1]The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi
[2]ICAR- Indian Agricultural Statistics Research Institute, New Delhi
[3]ICAR-Central Institute for Research on Cotton Technology, Mumbai
[4]ICAR-Indian Agricultural Research Institute, New Delhi
[5]ICAR-National Institute of Agricultural Economics and Policy Research, New Delhi

Received 23 October 2023; Revised 27 December 2023; Accepted 03 January 2024

SUMMARY

Rice, a crucial global staple, is integral to food security. Precise identification of paddy growth stages, booting, heading, anthesis, grain filling, and grain maturity is vital for agricultural decisions. However, a gap exists in recognizing these stages using red-green-blue (RGB) images. This study uses state-of-the-art computer vision and deep learning classification (Convolutional Neural Networks) algorithms to address this gap. Among the studied algorithms, EfficientNet_B0 achieved an impressive 82.8% overall accuracy. Notably, increasing image size from 64X64 pixels to 128X128 pixels significantly enhanced accuracy. A detailed assessment of growth stages revealed varying accuracy levels, with boot leaf being the most accurately detected (95.1%) and anthesis being the most challenging (72.28%). This work significantly advances automated monitoring, empowering researchers in real-time decision-making.

Keywords: Paddy; Growth stages; Deep learning; Computer vision; Convolutional neural network.

1. INTRODUCTION

Rice, a staple food for a significant portion of the global population, is pivotal in ensuring food security and sustainability. Understanding and accurately identifying the various growth stages of the rice plant is essential for agronomists and researchers alike. It forms the basis for making informed decisions regarding planting, irrigation, nutrient management, pest control, and harvesting, significantly influencing crop yield and quality.

Beginning with the early stages of seedling emergence, where delicate shoots pierce through the soil and progress through the vegetative growth phases, the paddy gradually prepares itself for reproduction. As it reaches the booting stage, it transitions into the reproductive phase, marking a critical turning point in its development. Accurate identification at this stage is vital, as it influences the formation of panicles and is responsible for producing rice grains. Subsequent stages include heading, anthesis (flowering), grain filling, and grain maturity. Each of these stages carries its significance and requirements. For instance, proper irrigation and nutrient management are essential during grain filling to ensure that rice grains develop fully and contain ample starch, contributing to yield and grain quality. Researchers, especially physiologists, need to identify these stages to understand plants' physiology.

Additionally, determining the optimal timing for harvesting during the grain maturity stage is crucial to maximize the economic value of the crop. Understanding and identifying different stages of paddy enables researchers to develop appropriate crop management practices such as optimized nutrient application and irrigation strategies. Different stages of

Corresponding author: Alka Arora
E-mail address: alka.arora@icar.gov.in

crops present varying vulnerabilities to pests, diseases, and environmental stresses. Identifying these stages allows for timely intervention and implementation of targeted pest control measures and disease management strategies.

Identifying different stages of paddy is crucial in various aspects of agricultural management, decision making and research. Computer vision, a subset of artificial intelligence, is one such technique that transformed how we perceive images. By employing advanced image processing techniques and deep learning algorithms, we can automate and enhance the identification and monitoring of paddy stages and manage the crop monitoring of rice plants. This optimizes resource utilization and empowers agronomists and physiologists with real-time insights into their rice fields, revolutionizing the decision-making process.

Recently, the deep learning concept of computer vision and artificial intelligence has gained pace to solve various problems related to agriculture and allied sectors across the world. Typical uses include disease and pest detection and crop and weed identification. Haque *et al.* (2022) applied classification algorithms of deep learning to identify diseases of maize viz., Maydis Leaf Blight, Banded Leaf, Sheath Blight and Turcicum Leaf Blight with an accuracy of 95.99%. Nigam *et al.* (2023) applied VGG19, ResNet152, DenseNet169, InceptionNetV3, and MobileNetV2 classification algorithms to classify leaf rust, stem rust and stripe rust in wheat and reported accuracy ranging from 91.2% to 97.8%. Haque *et al.* (2023) proposed a 15-layer convolution neural network (CNN) to classify Gray Leaf Spot, Common Rust and Northern Corn Leaf Blight diseases of maize crops on the publicly available dataset PlantVillage. Deep learning-based model Fruit-CNN can identify the type of fruit for their quality assessment with an accuracy of 99.6% (Kumar *et al.*, 2021). Narvekar and Rao, 2020 tested classification algorithms such as VGG16, MobileNetV2 and Resnet50 on publicly available flower datasets. The broadleaf and grass weeds of soybean crop were identified with an accuracy of 97% using CNN models (dos Santos Ferreira *et al.*, 2017). Jiang *et al.* (2019) applied the AlexNet CNN model on hyperspectral images combined with a machine learning-based segmentation algorithm to detect the postharvest apple pesticide residue and reported an accuracy of 99.09%.

Some studies have also been done on paddy to identify diseases and growth stages of the paddy. Vardhini *et al.* (2020) used CNN to classify rice blast, sheath blight, false smut and rice tungro disease. Ikasari *et al.* (2016) used LANDSAT-8 remote sensing data to identify the growth stages of paddy, namely vegetative, reproductive, ripening, and harvesting. They applied CNN models with different dropout and batch normalization combinations and achieved a maximum accuracy of 71.79%. Murata *et al.* (2019) used NDVI images collected from drone at different heights to classify growth stages in paddy. Using the CNN model, they achieved a classification accuracy of 71.2% at a height of 60 meters.

Although some work has been done on the paddy to classify the growth stages using remote sensing data, there is a significant gap in identifying the stages from RGB images. In this work, we have undertaken the task of identifying the boot leaf, heading, anthesis, grain filling and grain maturity using RGB (Red-Green-Blue) images. We implemented different state-of-the-art classification algorithms of computer vision and deep learning. We also experimented with two different sizes of images.

2. MATERIALS AND METHODS

2.1 Data Collection

The image data were collected from Nanaji Deshmukh Plant Phenomics Centre (NDPPC), ICAR-Indian Agricultural Research Institute (Misra *et al.*, 2019). The images were collected from the side view as the images facilitate us not only in identifying anthesis, grain filling and grain maturity stages but also booting and heading stages, which are crucial for the physiological study of the crop. The images were collected in five categories, namely boot leaf, heading, anthesis, grain filling and grain maturity, making five classes for our experiment. The images collected were varying in size. Fig. 1 shows the sample images from each of the classes. We used boot leaf data as boot leaf emerges 1-2 days after booting and has better visibility. A total of 4000 images were collected, and the number of images was increased to 10000 to train the models by adding different augmentations.

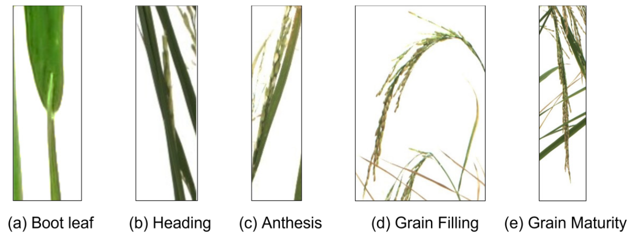

Fig. 1. Different stages used for identification

2.2 Data Preprocessing

We took two experiments of varying sizes. The first one involves resizing the image to 64X64 pixels, and the second one involves resizing the image to 128X128 pixels. After resizing, we applied data augmentation using the ‘Albumentation’ package. We applied “horizontal flip” to all the images and then “random brightness and contrast” to the original and horizontally flipped images to augment the data and increase the size of the data. Horizontal flip was applied as it generates a mirror view of the image. Random Brightness and Contrast was applied to get variation in the brightness and contrast of the dataset, making the models developed on this dataset more robust to adapt to these changes. After stacking three channels (Red, Green, and Blue) of the images, the data was then normalized by mean and variance of pixel values.

2.3 CNN Models

We used AlexNet (Krizhevsky *et al.*, 2012), VGG19 (Simonyan and Zisserman, 2014), MobileNetV2 (Sandler *et al.*, 2018), ResNet152 (He *et al.*, 2016), GoogleNet (Szegedy *et al.*, 2015), EfficientNet_B0 (Tan and Le, 2019) and EfficientNet_B7 (Tan and Le, 2019) convolutional neural network models to classify and compare the model performance.

AlexNet is one of the pioneering deep convolutional neural network (CNN) architectures. It gained attention after winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. It consists of eight layers—five convolutional layers (first, second and fifth convolution layers are followed by max-pooling layers) and three fully connected layers. AlexNet introduced concepts like ReLU activation functions and dropout, significantly advancing the field of deep learning (Krizhevsky *et al.*, 2012). The architecture of AlexNet is displayed in Fig. 2. The convolutional layer is the core building block of CNNs and is responsible for feature extraction. It comprises a set of learnable filters (kernels) that slide over input data (images) to perform convolution operations. These filters extract features by performing element-wise multiplications and aggregations, detecting patterns like edges, textures, and shapes within the input data. The resulting feature maps preserve spatial relationships, enabling the network to learn hierarchical representations of visual information. ReLU is an activation function commonly used in CNNs, introducing non-linearity by replacing negative values with zero and leaving positive values unchanged. It helps the network learn complex relationships in the data by introducing non-linearities, aiding in feature learning, and mitigating the vanishing gradient problem during training. Max pooling is a downsampling technique used after convolutional layers to reduce spatial dimensions while retaining essential features. It partitions feature maps into non-overlapping regions and retains the maximum value within each region, discarding other values. By capturing the most prominent features and reducing computational complexity, max pooling helps in spatial abstraction, translation invariance, and controlling

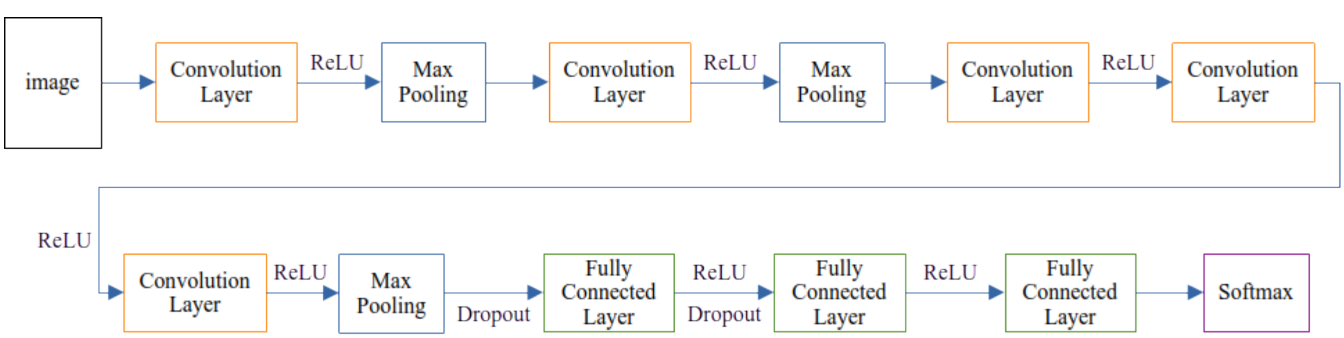

Fig. 2. Architecture of AlexNet

over fitting by providing a form of regularization. Fully connected layers, also known as dense layers, constitute the final segments of a CNN, preceding the output layer. Neurons in these layers are interconnected to all neurons in the preceding layer, enabling the network to learn complex patterns and relationships across the extracted features. Fully connected layers integrate high-level features captured by earlier layers and transform them into predictions or classifications suitable for the given task.

VGG19, part of the VGG (Visual Geometry Group) series, is a deep CNN architecture known for its simplicity and uniform architecture. It has 19 layers with small 3x3 convolutional filters, making the network deeper while maintaining a simple structure. VGG models are characterized by their stack of convolutional layers, often followed by max-pooling layers, and are effective at feature extraction (Simonyan and Zisserman, 2014).

MobileNetV2 is a lightweight CNN architecture optimized for mobile and edge devices. It employs depthwise separable convolutions to reduce computational complexity while maintaining good performance. MobileNetV2 introduces inverted residuals and linear bottlenecks, allowing for efficient use of parameters and faster inference on resource-constrained devices (Sandler *et al.*, 2018).

ResNet152 is part of the ResNet (Residual Network) series. It addresses the vanishing gradient problem by introducing skip connections or shortcuts that enable the flow of gradients through the network. ResNet152 has explicitly 152 layers, utilizing residual blocks that ease the training of very deep networks and achieve state-of-the-art performance in various computer vision tasks (He *et al.*, 2016).

GoogLeNet, also known as the Inception architecture, introduced the concept of inception modules employing multiple filter sizes within the convolutional layer. This architecture aims to capture features at various scales efficiently. GoogLeNet also employs global average pooling and auxiliary classifiers to aid in training deeper networks without encountering vanishing gradients (Szegedy *et al.*, 2015).

EfficientNet is a family of CNN architectures that achieve state-of-the-art performance with improved efficiency. The architecture uses a compound scaling method that scales the network's depth, width, and resolution. EfficientNetB0 represents the base model, while EfficientNetB7 is a larger, more powerful variant. These models balance accuracy and efficiency by optimizing network scaling (Tan and Le, 2019).

3. IMPLEMENTATION

We implemented all the models using PyTorch (Paszke *et al.*, 2019), a machine learning framework developed by Meta AI in the Python programming language. We used a Tesla V100 Nvidia DGX GPU server to run the experiments. While running the experiments, we used 10-fold cross-validation to avoid overfitting the data. The dataset was divided into a train set, validation set and test set in the ratio of 18:2:5, *i.e.* data was first divided into a 4:1 ratio to separate the "train+validation" set and the "test" set. The "train+validation" set was used for 10-fold cross-validation, dividing the "train+validation" set to a 9:1 ratio of the "train" set and "validation" set for each fold.

4. RESULT AND DISCUSSION

The models were developed using a cross-validation of 10 folds to validate the developed models. Further, to compare the developed models and evaluate their performance on unseen datasets, we assessed various CNN models using the test dataset and compared their accuracy. To understand how different image sizes impact model performance, the evaluation tests the models using two distinct image resolutions: 64X64 and 128X128 pixels. This analysis is essential as it helps assess how well these models adapt to different input image scales. The results of this evaluation are presented in Table 1, which offers a side-by-side comparison of the performance of each model. The overall accuracy achieved by each model is recorded for both 64X64 and 128X128 pixels image sizes.

Table 1. Performance of different CNN models

S. No.	Model	Mean accuracy of model on all the stages (%)	
		64X64 pixels image size	128X128 pixels image size
1	AlexNet	43.79	49.02
2	VGG19	62.09	65.26
3	MobileNetV2	72.08	74.43
4	ResNet152	71.24	72.59
5	GoogleNet	70.13	72.73
6	EfficientNet_B0	77.27	82.80
7	EfficientNet_B7	73.38	75.32

The study begins by evaluating a set of CNN models. This evaluation involves subjecting these models to a test dataset to determine how effectively they can classify images. Although the EfficientNet_B7 has been reported to be the best-performing model among the studied models, EfficientNet_B0 stood out as the top-performing model, achieving an impressive overall accuracy of 82.8% as it performs better for smaller-size images. On the opposite end of the spectrum, AlexNet, a model known for its historical significance in the field of deep learning, exhibits the lowest accuracy. It is dubbed the least effective model in this evaluation. An intriguing observation is made regarding the impact of image size on model accuracy. Increasing the image size from 64X64 to 128X128 pixels leads to a noticeable improvement in accuracy, especially for AlexNet. This indicates that some models benefit from higher-resolution input images. This is mainly because of increased spatial information, ultimately leading to enhanced feature extraction and reduction in information loss. Similar findings were noted by Nigam *et al.*, 2023 where they recorded that increasing image size led to increased accuracy up to a point after which the accuracy started decreasing.

With EfficientNet_B0 emerging as the top-performing model, the evaluation delves deeper into its capabilities. It specifically focuses on its performance at the 128X128 pixels image size, examining its ability to classify images at individual growth stages. Various metrics such as precision, recall, accuracy, and the F1 score are employed to assess its performance in detail. These metrics provide a more comprehensive understanding of how well EfficientNet_B0 performs at different growth stages. Fig. 3 visually represents the performance metrics at each growth stage. This visual aid offers a clear and intuitive way to understand the model's strengths and weaknesses in classifying images at different stages of growth.

This deeper analysis reveals that the "boot leaf" stage is where EfficientNet_B0 shines the brightest, achieving a high accuracy rate of 95.1% and good precision, recall and F1 score. In contrast, the "anthesis" stage presents a challenge, with an accuracy of 72.28%. Although the "heading" stage displayed better accuracy, it also displayed poor precision, recall and F1 score. This is mainly because the "heading" and "anthesis" stages have some similarities, which can be easily viewed in Fig. 1. A similar situation can be seen in the "grain filling" and "grain maturity" stages, where the "grain maturity" stage displayed good accuracy but also displayed comparatively poor precision, recall and F1 score. This highlights potential improvement in the model's ability to detect these specific stages accurately.

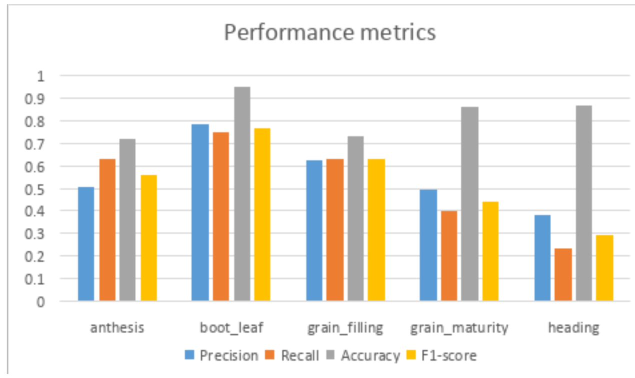

Fig. 3. Comparison of performance metrics for each class of EfficientNet_B0

CONCLUSION

Computer vision, a subset of artificial intelligence, has emerged as a transformative tool in agriculture. We can automate and improve the identification and monitoring of rice plant growth stages by employing advanced image processing techniques and deep learning algorithms. This innovation optimizes resource utilization and empowers stakeholders with real-time insights into their rice fields, revolutionizing decision-making processes.

The results demonstrated that the EfficientNetB0 model outperformed others with an overall accuracy of 82.8%. Increasing the image size from 64X64 to 128X128 pixels significantly improved model accuracy, particularly for AlexNet. Furthermore, a detailed analysis of the EfficientNetB0 model revealed varying accuracy levels for different growth stages, with boot leaf being the most accurately detected (95.1%) and anthesis being the most challenging (72.28%).

This study highlights the potential of computer vision and deep learning in accurately identifying rice growth stages from RGB images. The findings provide valuable insights for further advancements in precision agriculture, ultimately contributing to optimizing rice farming practices and global food security. Further research and refinement of these models can unlock even greater potential in leveraging technology for sustainable agricultural development.

REFERENCES

dos Santos Ferreira, A., Freitas, D.M., da Silva, G.G., Pistori, H. and Folhes, M.T., 2017. Weed Detection in Soybean Crops using ConvNets. *Computers and Electronics in Agriculture* **143**: 314-324. https://doi.org/10.1016/j.compag.2017.10.027

Haque, M.A., Marwaha, S., Deb, C.K., Nigam, S. and Arora, A., 2023. Recognition of Diseases of Maize Crop using Deep Learning Models. *Neural Computing and Applications* **35** (10): 7407-7421. https://doi.org/10.1007/s00521-022-08003-9

Haque, M.A., Marwaha, S., Deb, C.K., Nigam, S., Arora, A., Hooda, K.S., Soujanya, P.L., Aggarwal, S.K., Lall, B., Kumar, M. and Islam, S., 2022. Deep Learning-based Approach for Identification of Diseases of Maize Crop. *Scientific reports***12** (1): 6334. https://doi.org/10.1038/s41598-022-10140-z

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep Residual Learning for Image Recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition* 770-778. https://doi.org/10.48550/arXiv.1512.03385

Ikasari, I.H., Ayumi, V., Fanany, M.I. and Mulyono, S., 2016. Multiple Regularizations Deep Learning for Paddy Growth Stages Classification from LANDSAT-8. *In 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* 512-517. IEEE. https://doi.org/10.1109/ICACSIS.2016.7872790

Jiang, B., He, J., Yang, S., Fu, H., Li, T., Song, H. and He, D., 2019. Fusion of Machine Vision Technology and AlexNet-CNNs Deep Learning Network for the Detection of Postharvest Apple Pesticide Residues. *Artificial Intelligence in Agriculture* **1**: 1-8. https://doi.org/10.1016/j.aiia.2019.02.001

Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in neural information processing systems* **25**.

Kumar, A., Joshi, R.C., Dutta, M.K., Jonak, M. and Burget, R., 2021. Fruit-CNN: An Efficient Deep learning-based Fruit Classification and Quality Assessment for Precision Agriculture. *In 2021 13th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)* 60-65. IEEE. https://doi.org/10.1109/ICUMT54235.2021.9631643

Misra, T., Arora, A., Marwaha, S., Roy, M., Raju, D., Kumar, S., Goel, S., Sahoo, R.N. and Chinnusamy, V., 2019. Artificial neural network for estimating leaf fresh weight of rice plant through visual-NIR imaging. *Indian Journal of Agricultural Sciences* **89** (10): 1698-1702. https://doi.org/10.56093/ijas.v89i10.94631

Murata, K., Ito, A., Takahashi, Y. and Hatano, H., 2019. A Study on Growth Stage Classification of Paddy Rice by CNN using NDVI Images. *In 2019 Cybersecurity and Cyberforensics Conference (CCC)* 85-90. IEEE. https://doi.org/10.1109/CCC.2019.000-4

Narvekar, C. and Rao, M., 2020. Flower Classification using CNN and Transfer Learning in CNN- Agriculture Perspective. *In 2020 3rd international conference on intelligent sustainable systems (ICISS)* 660-664. IEEE. https://doi.org/10.1109/ICISS49785.2020.9316030

Nigam, S., Jain, R., Marwaha, S., Arora, A., Haque, M.A., Dheeraj, A. and Singh, V.K., 2023. Deep Transfer Learning Model for Disease Identification in Wheat Crop. *Ecological Informatics* **75**: 102068. https://doi.org/10.1016/j.ecoinf.2023.102068

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L. and Desmaison, A., 2019. Pytorch: An imperative style, high- performance deep learning library. Advances in neural information processing systems, 32.

https://doi.org/10.48550/arXiv.1912.01703

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted Residuals and Linear Bottlenecks. *In Proceedings of the IEEE conference on computer vision and pattern recognition* 4510-4520. https://doi.org/10.48550/arXiv.1801.04381

Simonyan, K. and Zisserman, A., 2014. Very Deep Convolutional Networks for Large-scale Image Recognition. *arXivpreprint arXiv:1409.1556*. https://doi.org/10.48550/arXiv.1409.1556

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. *In Proceedings of the IEEE conference on computer vision and pattern recognition* 1-9. https://doi.org/10.48550/arXiv.1409.4842

Tan, M. and Le, Q., 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *In International conference on machine learning* 6105-6114. PMLR. https://doi.org/10.48550/arXiv.1905.11946

Vardhini, P.H., Asritha, S. and Devi, Y.S., 2020. Efficient Disease Detection of Paddy Crop using CNN. *In 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)* 116-119. IEEE. https://doi.org/10.1109/ICSTCEE49637.2020.9276775