How much of this submission has been generated by AI?

0%

of qualifying text in this submission has been determined to be generated by AI.

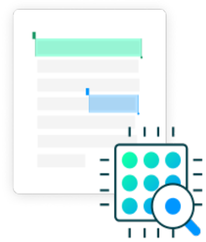
Caution: Percentage may not indicate academic misconduct. Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Frequently Asked Questions

What does the percentage mean?

The percentage shown in the AI writing detection indicator and in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was generated by AI.



Our testing has found that there is a higher incidence of false positives when the percentage is less than 20. In order to reduce the likelihood of misinterpretation, the AI indicator will display an asterisk for percentages less than 20 to call attention to the fact that the score is less reliable.

However, the final decision on whether any misconduct has occurred rests with the reviewer/instructor. They should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in greater detail according to their school's policies.

How does Turnitin's indicator address false positives?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be AI-generated will be highlighted blue on the submission text.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

What does 'qualifying text' mean?

Sometimes false positives (incorrectly flagging human-written text as AI-generated), can include lists without a lot of structural variation, text that literally repeats itself, or text that has been paraphrased without developing new ideas. If our indicator shows a higher amount of AI writing in such text, we advise you to take that into consideration when looking at the percentage indicated.

In a longer document with a mix of authentic writing and AI generated text, it can be difficult to exactly determine where the AI writing begins and original writing ends, but our model should give you a reliable guide to start conversations with the submitting student.

2024 International Conference on Advancement in Computation & Computer Technologies (InCACCT)

SoyNet: Deep Learning Approaches for Automated Soybean Seed Quality Assessment

Harshita Pratap
Dept. of Computer Science and Engg.
Graphic Era (Deemed to be University)
Dehradun, India
harshitapratap60@gmail.com

Guru Prasad M S
Dept. of Computer Science and Engg.
Graphic Era (Deemed to be University)
Dehradun, India
guru0927@gmail.com

Poorvi Agarwal
Dept. of Computer Science and Engg.
Graphic Era (Deemed to be University)
Dehradun, India
poorviagarwal22@gmail.com

Ayushi Bhardwaj
Dept. of Computer Science and Engg.
Graphic Era (Deemed to be University)
Dehradun, India
bhardwajayushi519@gmail.com

Ayushi Mehra
Dept. of Computer Science and Engg.
Graphic Era (Deemed to be University)
Dehradun, India
mehraayushi2016702@gmail.com

Shilpa Mathpati
School of computing & Information
Technology Reva University,
Bengaluru, India
shilpapm1@gmail.com

Abstract- In the realm of agriculture, soybeans have played a vital role in diets for centuries, providing a consistent source of plant protein. Beyond protein, soybeans offer essential components like fibres, vitamins (including B vitamins), minerals (such as iron, calcium, and magnesium), and antioxidants. This necessitates precise quality assessment for factors like size, shape, colour, and damage. Leveraging advanced machine learning techniques such as VGG16, AlexNet and CNN, the model aims to automate the identification and categorization of soybean quality. To implement the proposed work the standard soybean dataset is considered. It consists of
five soybean seed types complete, spotted, immature, broken, and skin-damaged. The proposed work effectively categories the quality of soybean seeds and it has the potential to improve the agricultural process as compared to traditional work. In this work different pretrained models are experimented in that VGG16 provides the highest of all showing accuracy 90%, AlexNet accuracy 78.49% and CNN accuracy 74.55%.

Keywords – Soybeans, Seed Quality Analysis, Deep Learning, CNN, VGG16, Agricultural Processes.

I. INTRODUCTION

Soybeans have played a major role in human diets from many years by providing a consistent source of protein in daily consumption. Above this, soybeans consist of essential components like fibre, vitamins (including B vitamins), minerals (such as iron, calcium, and magnesium), and antioxidants [1]. Because of rich content of nutrition users' preference on soybean is paramount. The high quality of soybeans have huge demand in the market. Several factors, including size, shape, colour, physical damage, insect damage, and fungal infection, can impact the quality of soybeans [2]. Figure 1 depicts the sample soybean seeds.

It is essential to preserve customer confidence and the quality of the soybean supply chain. Consequently, the classification of soybean seeds based on their quality becomes a pivotal task [3]. The objective of our work is to create a model capable of automatically identifying and categorizing various aspects of soybean quality. Recent strides in machine learning have facilitated this endeavour, with diverse deep learning techniques such as CNN, AlexNet, VGG16, SVM, and others offering effective means for developing classification models [4].

The models accuracy highly depends upon the dataset used to build such a model, as a more diversified dataset will lead to a more robust classification model. In case the dataset is not already pre-processed then various other steps like data cleaning, data visualization and data preprocessing needs to be carried out [9]. The general steps required to build such classification models are data collection and preprocessing, model selection (like CNN, AlexNet, CGG16), training, evaluation based on various performance metrics and deployment.

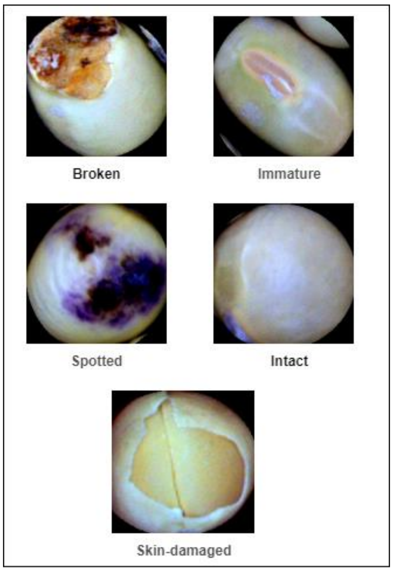


Fig. 1. Sample soybean seeds

Building such a classification model will replace labour and time-consuming job of manually determining the quality of each soybean grain and help in maintaining the public health. Further leading to minimization of food waste and leading to good quality products as well.

II. LITERATURE SURVEY

The study of this sector of classification of multiclass soyabean seeds has been done to form this literature survey. Some of the studies of are given below:

Marcelo Souza et. al., [1] The suggested strategy focuses on the classification of Soybean seed vigor to enhance productivity in soybean crops. Employing advanced deep learning methods, particularly CNN, and adopting a pipelining approach for seed image segmentation, the goal is to establish a robust vision based system. The developed system has demonstrated notable

979-8-3503-7131-4/24/$31.00 ©2024 IEEE

success with an accuracy of 80.17%±2.37, showcasing its effectiveness in seed vigor classification.

Guoyang Zhao et. al., [2] focused on The central objective is the creation of a real-time recognition system aimed at identifying defects across the entire surface of soybean seeds. Employing deep learning models, the methodology involves precise seed classification, wherein soybean seeds are categorized into six distinct classes. The proposed approach achieves an impressive overall sorting accuracy of 98.87% and operates at a sorting speed of 222 seeds per minute.

Shaolong Zhu et. al., [3] implemented is a swift and remarkably effective technique for discerning various soybean seed varieties, integrating hyperspectral images with transfer learning. Capturing hyperspectral images of both seed fronts and backs, soybean reflectance data was extracted. The most suitable Convolutional Neural Network (CNN) model for identifying soybean seed varieties was meticulously chosen.

Nikhil Kaler et. al., [4] concentrated on a resilient analysis of laser bio-speckle data for the identification of fungal-infected soybean seeds, the approach relies on deep learning. The analysis involves a combination of various techniques, including CNN, LSTM, 3D CNN, and conv(LSTM). A dataset comprising 1000 soybean seeds was curated, resulting in an impressive testing accuracy of 97.72%.

Nguyen Hong Son et. al., [5] explored were image processing algorithms and machine learning methodologies for the identification and categorization of two distinct types of rice (whole rice and broken rice). The results suggest that the automation of rice quality assessment and classification can be efficiently achieved by employing Deep Learning techniques.

Sutasinee Jitanan et. al., [6] discussed about quality grading of soybean seeds using image analysis. In this research, background subtraction was employed to mitigate the presence of shadows in captured images. Additionally, a technique was introduced to extract color features emphasizing resilience to illumination changes. The dataset comprised 1,320 soybean seeds, having 6,600 seed images exhibiting difference in brightness levels. Experimental outcomes demonstrated accuracies 98% to 100% across different seed types.

III. METHODOLOGY

In this investigative study focusing on soybean quality analysis, vibrant deep learning algorithms such as CNN, AlexNet, and VGG16 are utilized. The dataset comprises images of five distinct types of soybean seeds- complete, spotted, immature, broken, and skin-damaged totalling 5513 images [7]. Out of which spotted soybean images are 1058, skin-damaged soybean images are 1127, intact Soybean images are 1201, immature soybean images are 1125 and broken soybean images are 1002. The images, captured by an artificial camera, depict soybean seeds in physical contact, and subsequently, individual soybean images were resized to 227 × 227 pixels. Figure 2 shows flow diagram of this application.

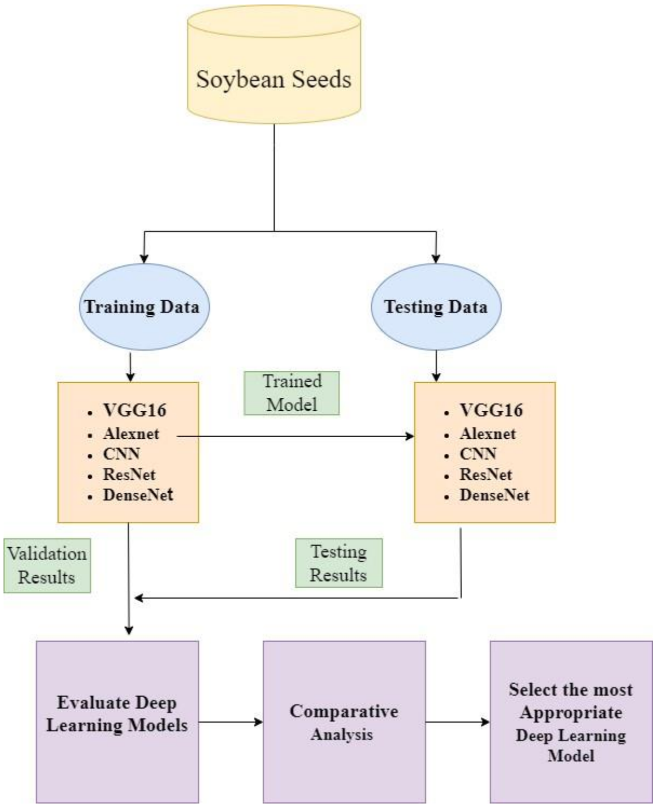


Fig. 2. Flow diagram for the application

Step 1: First and the foremost step is collection of dataset which is of soybean seeds containing 5513 images comprising of total 5 categories (intact, immature, skin-damaged, broken and spotted).

Step 2: In the second step the dataset is divided into training and testing part in the ratio of 80:20.

Step 3: After that various deep learning algorithms are applied on the dataset such as VGG16, AlexNet and CNN.

Step 4: Based on the accuracy and classification report, out of three models the best one is considered.

A. Deep Learning

Deep learning, a branch of artificial intelligence (AI), employs a methodology that instructs computers to analyze data in a manner inspired by the human brain. These models have the capability to discern intricate patterns within various types of data, including images, text, and sounds, facilitating the generation of precise insights and predictions [8].

Employing deep learning techniques allows for the automation of tasks traditionally reliant on human intelligence, such as image description or transcription of audio files into text. The underlying structure of deep learning algorithms consists of neural networks, which emulate the interconnected nature of the human brain. Analogous to the millions of interconnected neurons in the human brain that collaborate to learn and process information, deep learning neural networks, or artificial neural networks, comprise numerous layers of artificial neurons working in tandem within a computer system.

B. Deep learning Techniques

The following deep learning techniques are implemented in this work

- VGG16
- AlexNet
- CNN
- DenseNet
- ResNet

1) VGG16

VGG16, denoting the VGG model or VGGNet, represents a convolutional neural network (CNN) model characterized by its 16-layer architecture. This model encompasses convolutional layers utilizing compact 3×3 filters, alongside max-pooling layers and fully connected layers. It is easy and simple to understand unlike some other architecture which are difficult to understand. VGG16 can identify low-level features like edges and textures, and using these it builds up model to recognize more complex patterns in images. [9].

2) AlexNet

AlexNet is a deep learning framework, which is a modified/updated version of the convolutional neural network, consisting of total 8 layers which includes 5 convolutional layers, max-pooling layers and 3 fully connected layers.

3) CNN

Convolutional Neural Networks is a deep learning model mainly used for image classification and recognition. It automatically learns the hierarchical features from input image with the help of its convolutional layers, pooling layers and fully connected layers. It is the most used neural network for computer vision[10].

4) ResNet

ResNet stands for Residual Network which is a deep learning model designed specifically to excel in computer vision tasks. ResNet functions within the Convolutional Neural Network (CNN) framework, designed with a distinctive capability to manage an extensive range of convolutional layers, spanning from hundreds to potentially thousands in quantity. As researchers sought to enhance the depth of these networks, they faced the hurdle of the "vanishing gradient" problem. In addressing this concern, ResNet introduces a creative solution called "skip connections." In the context of ResNet, identity mappings(connections that allow the input to flow directly to the output) are introduced within the convolutional layers of the network. Instead of being altered by these layers, the input simply passes through these mappings unchanged. As the input skips certain convolutional layers, the activations (or outputs) from the preceding layers are reused. By bypassing certain convolutional layers and reusing activations, the training process is accelerated , mitigating the vanishing gradient problem often encountered in deep neural networks and enabling the training of significantly deeper models[11].

5) DenseNet

DenseNet emerged as a groundbreaking response to the constraints inherent in conventional CNN architectures. It represents a specific category of convolutional neural network that incorporates dense connections between its layers, achieved through the integration of Dense Blocks. Within these blocks, all layers with corresponding feature-map sizes are directly interconnected.

C. Optimizer

VGG-16 model achieves 84% accuracy using the Adam optimizer which is one of the most widely used optimization algorithms in CNNs and in deep learning models. The Adam optimizer adjusts the learning rate for each parameter on the basis of historical gradients by combining features in RMSprop with Momentum. Adaptively scaling learning rates by first-order and second-order variance of gradients increases the optimization of the model.

To increase the accuracy beyond 84%, we implemented Batch Normalization which is a technique that improves training stability and accelerates convergence. It normalizes the layer activations during each mini-batch of training by subtracting the batch mean and dividing it by the batch standard deviation, which reduces internal covariate shift and ensures a consistent distribution of the network activations after normalization. By scaling and shifting the activations using learnable parameters such as gamma and beta, the network learns optimal representations for every layer.

Another advantage of BN is that it allows for higher learning rates during the training. This proves advantageous as the model learns faster and explores the solution space more efficiently. The combination of Adam optimization and BN increases the accuracy of the model.

IV. RESULTS AND DISCUSSION

The proposed work is experimented on standard soybean dataset [7]. There are totally 5513 soybean seed images in the dataset. Table I shows the description of dataset. Figure 3 depicts the soybean seed category with its count.

TABLE I. DATASET DESCRIPTION

Soybean seed category	Number of samples
Spotted soybean	1058
Skin-damaged soybeans	1127
Intact soybeans	1201
Immature soybeans	1125
Broken soybeans	1002

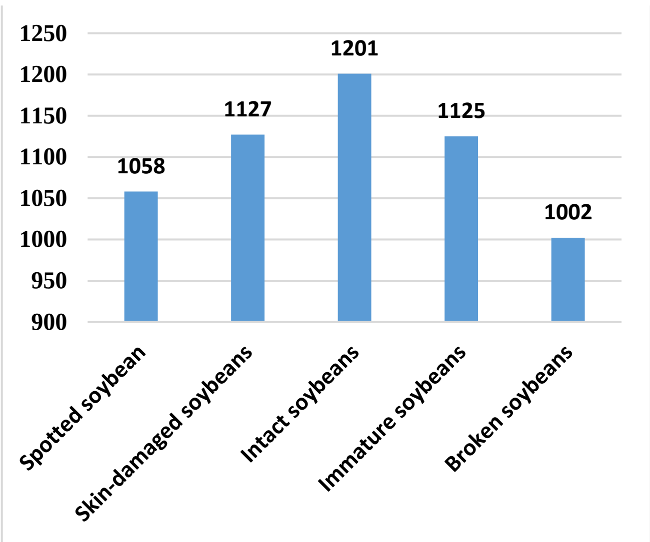


Fig. 3. Soybean seed category

In this section the performance of our system is analysed. For evaluation, three performance evaluation metrics i.e. Training Accuracy, Testing Accuracy, and Support are used. The metrics can be defined as follows:

$$Accuracy = (TN + TP) / (TN+TP+FN+FP) \text{ -----(1)}$$

Support represents the count of actual instances of a class within the dataset, serving as a crucial metric for assessing class distribution balance.

Where,

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

Table II gives an analysis of the dataset and shows that, with an accuracy of 90%, the VGG16 excels the other 2 classification methods. Three different performance factors were taken into consideration when evaluating the algorithms. To put it clearly, Table I provides an overview of the experimental results associated with the review of these techniques.

TABLE II. ALGORITHM EVALUATION METRIC RESULTS

Classifier	Accuracy	Precision	Recall	F1-Score
VGG 16	90%	75%	76%	72%
Alex Net	78.49%	45%	43%	46%
CNN	74.55%	40%	42%	44%

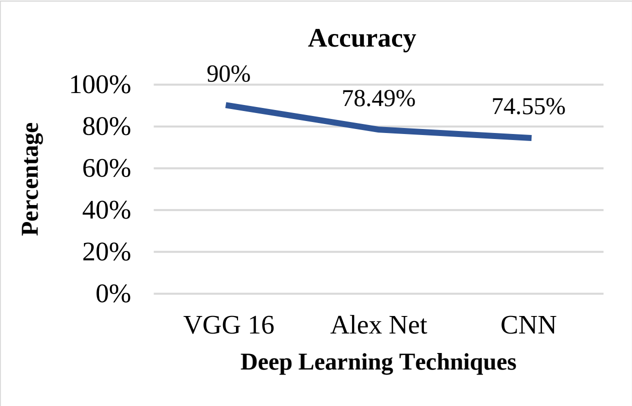


Fig. 4. Graph depicting Accuracy of different classification Algorithms

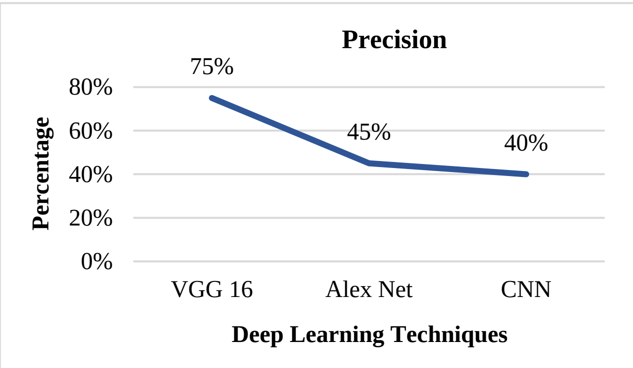


Fig. 5. Graph depicting Precision of different classification Algorithms

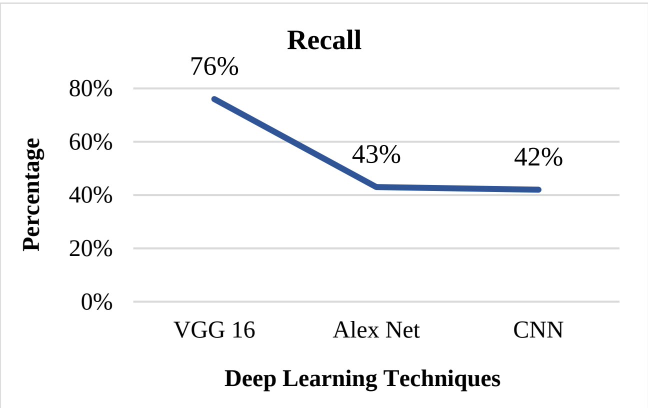


Fig. 6. Graph depicting Recall of different classification Algorithms

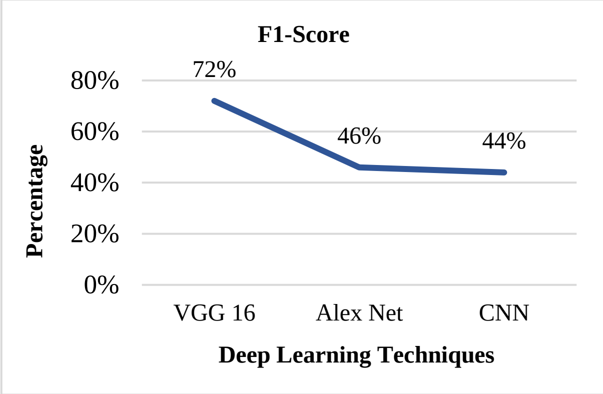


Fig. 7. Graph depicting F1 Score of different classification Algorithms

Collectively, Fig. 4,5,6 and 7 provides a complete graphic representation of the classifiers' performance in terms of accuracy, precision, recall and f1 score. It can be inferred that VGG 16 gives the highest accuracy of all the deployed models. VGG 16 has been pretrained and used some fine tuning for deployment process. Pre-training involves training the network on a large dataset (e.g., ImageNet) for image classification tasks.

V. CONCLUSION

The proposed model for analysing soybean seed quality utilizes deep learning algorithms capable of automatically identifying and categorizing various aspects of soybean quality. In this research, we implemented three distinct models: VGG-16, AlexNet, and CNN.VGG-16 exhibited the highest accuracy of 90%. This result suggests that the intricate and deep layers of VGG-16 are adept at capturing intricate patterns and features in soybean seeds, leading to more precise and dependable quality assessments.

The incorporation of deep learning into seed quality analysis presents a promising avenue for improving efficiency and accuracy in agricultural processes. While the findings emphasize the success of VGG-16 in this particular context, it is crucial to recognize the continuous evolution of deep learning models and the potential for further enhancements. Future research could explore integrating additional data sources, refining model parameters, and employing more advanced architectures to continually push the boundaries of seed quality analysis.

REFERENCE

[1]. M. M. S. De Souza, W. C. Horikoshi, P. T. M. Saito, and P. H. Bugatti, “Soybean seed vigor classification through an effective image learning-based approach,” Multimedia Tools and Applications, vol. 83, no. 5, pp. 13113–13136, Jul. 2023, doi: 10.1007/s11042-023-15804-0.

[2]. G. Zhao et al., “Real-time recognition system of soybean seed full-surface defects based on deep learning,” Computers and Electronics in Agriculture, vol. 187, p. 106230, Aug. 2021, doi: 10.1016/j.compag.2021.106230.

[3]. S. Zhu et al., “A Rapid and Highly Efficient Method for the Identification of Soybean Seed Varieties: Hyperspectral Images Combined with Transfer Learning,” Molecules, vol. 25, no. 1, p. 152, Dec. 2019, doi: 10.3390/molecules25010152.

[4]. N. Kaler, V. Bhatia, and A. K. Mishra, “Deep Learning-Based robust analysis of Laser Bio-Speckle data for detection of Fungal-Infected soybean seeds,” IEEE Access, vol. 11, pp. 89331–89348, Jan. 2023, doi: 10.1109/access.2023.3305273.

[5]. N. Hong Son and N. Thai-Nghe, "Deep Learning for Rice Quality Classification," 2019 International Conference on Advanced Computing and Applications (ACOMP), Nha Trang, Vietnam, 2019, pp. 92-96, doi: 10.1109/ACOMP.2019.00021.

[6]. S. Jitanan and P. Chimlek, “Quality grading of soybean seeds using image analysis,” International Journal of Power Electronics and Drive Systems, vol. 9, no. 5, p. 3495, Oct. 2019, doi: 10.11591/ijece.v9i5.pp3495-3503.

[7]. W. Lin et al., “Soybean seeds,” Mendeley Data, May 2023, doi: 10.17632/v6vzvfszj6.6.

[8]. A. S. Rao, N.-T. Nguyen, M. Palaniswami, and T. Ngo, “Vision-based automated crack detection using convolutional neural networks for condition assessment of infrastructure,” Structural Health Monitoring,

vol. 20, no. 4, pp. 2124–2142, Nov. 2020, doi: 10.1177/1475921720965445.
[9]. S. Mittal, A. K. Mishra, V. Tripathi, P. Singh and P. Pandey, "A Comparative Analysis of Supervised Machine Learning Models for Smart Intrusion Detection in IoT Network," 2023 3rd Asian Conference on Innovation in Technology (ASIANCON), Ravet IN, India, 2023, pp. 1-6, doi: 10.1109/ASIANCON58793.2023.10270377.
[10]. A. Shrivastava, P. Singh, H. Gupta, R. Kumar, V. Singh and O. Krishna, "Fuzzy Based Sentiment Analysis of E-Com and S-Com Data Using Transformer Neural Network," 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 2023, pp. 730-737, doi: 10.1109/ICPCSN58827.2023.00126.
[11]. R. Barthwal, D. Kathuria, S. Joshi, R. S. S. Kaler, and N. Singh, “New trends in the development and application of artificial intelligence in food processing,” Innovative Food Science and Emerging Technologies, p. 103600, Feb. 2024, doi: 10.1016/j.ifset.2024.103600.