Phoenix: International Multidisciplinary Research Journal
Vol 3, No. 4, Oct-Dec, 2025
ISSN:2583-6897

NAIVE BAYES CLASSIFIER TO IDENTIFY DISEASE IN FRUITS

Dikahendra Daulat Sarpate
Professor, Department of Artificial Intelligence & Data Science, ZEAL College of Engineering & Research, Pune, Email id -dikshendra@gmail.com
B. Sankaraiah
Assistant Professor, Department of Computer Science & Engineering, Malla Reddy (MR) deemed to be University, Hyderabad. Email id.: Shankar61186@gmail.com
B.Rani
Assistant Professor, Department of Computer Science & Engineering, Malla Reddy (MR) deemed to be University, Hyderabad. Email id.: rani@mrec.ac.in
Vemula Nikitha
Assistant Professor, Department of Computer Science & Engineering, Malla Reddy (MR) deemed to be University, Hyderabad. Email id.: nikitha479@gmail.com
Dr. Syed Umar
Professor, Department of Computer Science & Engineering, Malla Reddy (MR) deemed to be University, Hyderabad. Email id.: syedumar@mrec.ac.in

Abstract— Fruit infections are a serious issue that hurts the agriculture industry and the economy. In the past, tainted fruit had to be manually identified; but, as technology has advanced, image processing technologies have been created. This system operates in two stages: training and testing. The testing step determines whether the fruit is contaminated and, if so, by which illness. The training phase stores all data pertaining to both infected and non-infected fruit. This work developed a method for identifying infected and non-infected fruit by combining the K-mean clustering algorithm, the speedup robust feature (SURF) feature detector, and the Nave Bayes Classifier. A database of fruits is used for the investigations, and the results are contrasted with those of a neural network. The outcomes show how effective the Nave Bayes Classifier approach is. In recent years, fruit illnesses have been identified using clustering and fruit picture segmentation approaches. To illustrate the significance of an algorithm graphic, multiple estimations are used. Examples of ratios include the probability ratio, specificity ratio, and intensity ratio.

INDEX - K-means clustering algorithm, intensity ratio, specificity ratio, probability ratio, fruit disease, SURF (speedup robust feature), NN(Neural Network) etc.

INTRODUCTION For the horticultural industry, agricultural photos are a great source of data and information. Image processing tools are crucial for the analysis of agricultural issues. One of the most crucial tools available to farmers for identifying plant disease at every stage of growth is fruit disease detection. By providing appropriate plant management techniques, such as disease-specific chemical applications and pesticide sprays, early infection and plant health detection can help avoid fruit illnesses and boost plant yield. Because it can identify disease symptoms from gathered photos as soon as they appear on developing fruits, automatic fruit illness diagnosis is essential. Fruit diseases harm variety and reduce yield [1], which can cause cultivation to be abandoned. Fruit illnesses start out as spots on the fruit and cause significant financial loss if they are not treated right away. One of the main causes of groundwater contamination has been identified as pesticides. Since pesticides have been found to be a major cause of groundwater contamination, using them excessively to treat fruit disease may increase the risk of hazardous residual levels on soil. Consequently, a framework is presented that could identify fruit illnesses as soon as symptoms show up on the fruits, enabling the plant to receive the appropriate treatment [2,3]. Using automation techniques and decision tools that seamlessly integrate product, knowledge, and services, today's farmer can use smart farming to increase production, grading, and surplus yield.

This study discusses the problem of automatically detecting and classifying fruit illness and provides a framework for doing so. Two diseases of apples, apple rot and apple scab, and two diseases of mangos, mango anthracnose and mango fruit fly, are fought by the system. This framework uses the K-mean technique for image segmentation, which divides the image into clusters and separates the fruit's contaminated area. Next, the feature extraction approach is used to extract the feature of the damaged part. Because SURF uses the blob detection technique, it yields better results when utilized for feature extraction. Then, a Nave Bayes classifier is used for photo categorization. This classifier is more accurate and comparatively faster to implement, and it has demonstrated better results than a neural network in terms of time consumption and confusion matrix. Following the identification of fruit diseases, farmers receive treatment advice that will help them control the disease and increase fruit yield [4.5].

I. RELATED WORK

The researchers Jahanbakhshi et al. presented a method for identifying mango fruits from pictures [6]. The most important addition was the utilization of the elliptical mango shape as a detecting tool. The method's pre-processing steps included opening morphology, detecting edges, and computing distances between edges, converting to grayscale, and transforming to binary color images. To capitalize on the elliptical shape of mangos, they employ the Randomized Hough Transform to identify oval items in the input image. Utilizing a back propagation neural network, the mango fruit was categorized according to its anticipated oval forms. In this investigation, a three-layer neural network was employed. 450 neurons in the input layer, 50 neurons in a single hidden layer, and an output layer transmitted the values of the clipped oval shape image. When a mango is clearly visible, the detection rate can reach 96.26 percent; when a mango is torn, it can reach 90 percent. Uppu Lokeshet al. [7] proposed a novel technique called "extended spectral angle mapping (ESAM)" (Huang long bingor HLB) to identify citrus greening illness. A multispectral image and an HS image from an aerial HS imaging device were used in the investigation. By automatically differentiating between the stem and calyx ends, Nandipati Sai Akash et al. [5] proposed a pattern recognition technique for identifying berries that are damaged. In order to extract color and geometrical information, blue berries were first photographed under normal settings. Second, five approaches were assessed to identify the features that should be used in further cross-validation and classification algorithm assessments. The best classifiers were found to be Support Vector Machine and Linear Discriminant Analysis [8]. Habeeb, M. S et al. [9] suggested a flexible method for identifying fruit diseases. Image processing, which comprises stages like segmentation, feature extraction, and classification, is the foundation of this technique. The analysis of the findings shows that the suggested approach can greatly increase the precision of fruit disease diagnosis and detection. One technique for identifying two types of the Penicillium fungus in citrus fruits was presented by Nandipati Sai Akash al. [5]. Avoiding or at least minimizing the ensuing monetary losses is the aim. They significantly reduced the amount of characteristics by using the MRMR (minimal redundancy and maximum relevance) technique. A reduced feature set-based method for identifying and classifying fruit images into normal and impacted categories was created by Jahanbakhshi et al. [6]. The average accuracy for normal and affected kinds, using two texture features, is 89.50 percent and 93.15 percent, respectively. They can be utilized to build a machine vision system in horticulture and agriculture. Uppu Lokesh et al. have devised and validated a classification approach for apple fruit diseases based on image processing [7]. The procedure includes the four steps. In the first stage, a defect segmentation approach based on K-means clustering was used to extract a region of interest. In the second stage, innovative color, texture, and shape-based qualities were extracted from the segmented apple illnesses. To generate a more distinctive feature, the different types of features were combined in the third step. Finally, the data was trained and classified using an MSVM (Multi-class support vector machine). Three different apple diseases—apple blotch, apple rot, and apple scab—as well as regular apples were used in the experiment. The research and results showed how significant and distinctive their approach to diagnosing apple illnesses was. Based on the classification results, they came to the conclusion that normal apples could be easily distinguished from infected apples, and that those combinations of color, texture, and shape-based features performed better than the most advanced color, texture, and shape features alone, with the shape feature contributing less. A technique for recognizing fruits and vegetables from photos was created and evaluated by R. Gnanakumaran et al. [8]. The three stages of the system that are covered here are background subtraction, feature extraction, and training and classification. Background subtraction was performed using segmentation based on K-means clustering. They extracted and combined a number of innovative color and texture elements from the foreground image. When color and texture information are combined, the resulting feature is more discriminative than either feature alone. The Multi-class Support Vector Machine (MSVM) was used for both training and classification. SVM (Support Vector Machine) is a better option for training and classification, according to their analysis of the performance fused characteristics of SVM and the nearest neighbor classifier [9].

II. METHODOLOGY

One of the primary objectives is to monitor fruit illnesses and come up with alternative treatments in order to maintain a healthy crop and achieve high levels of productivity. Acquisition of an image is always the initial condition for the work flow sequence of image processing. This is due to the fact that processing can only be accomplished with the assistance of a representation. Picture segmentation is accomplished through the application of the K-Means clustering technique [10,11]. Feature vectors, which include picture color, morphology, texture, and hole structure, are utilized in the process of extracting features from each image and diagnosing the morphology of the disease. For the purpose of retrieving features, the SURF method is utilized as a locator and descriptor, whereas the nave bayes classifier is utilized for the categorization of diseases. The algorithm can be broken down into the following steps (see figure 1):

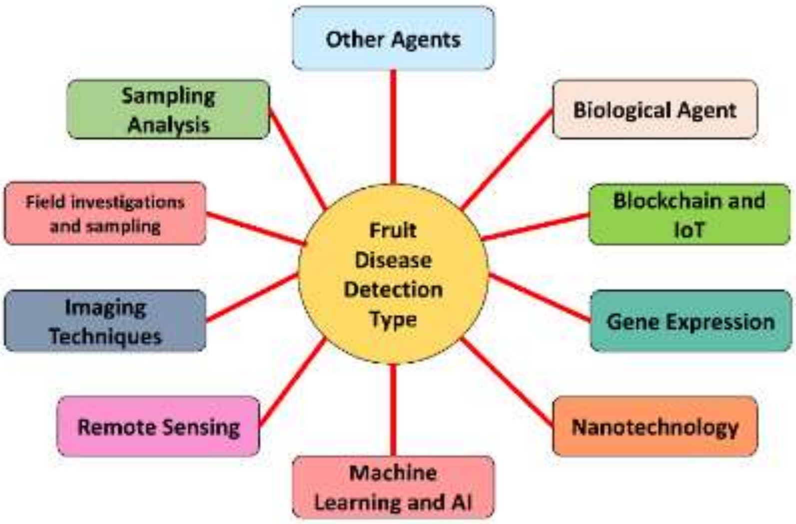


Fig 1: Steps Involved In Detection of Fruit Disease [4]

Image acquisition: Image acquisition is the action of retrieving an image from a source in image processing.
Segmentation is a technique for breaking down an image into discrete sections. Pixels with comparable properties will be found in regions, and each area will contain pixels with similar attributes.
Segmentation: Segmentation is the initial step from low-level image processing to high-level image description in terms of features, objects, and scenes [12]. It involves converting a grayscale or colour image into one or more additional images.
Feature Extraction: In this study, the K-mean segmentation technique is applied.
Color, texture, morphology, and structure of the fruits' holes are all taken into account during feature extraction. The features are extracted using the SURF (Speed up Robust Feature) technique. It's utilised as a blob detector and a local descriptor.
Classifier: The Nave Bayes (NB) Classifier is used to determine whether or not the fruit is defective.

Speeded Up Robust Features (SURF)

The SURF algorithm is a feature detector and descriptor that can be utilized for a variety of tasks, including recognition and classification. In the year 2024, Habeeb, M. S presented out the idea during the European Conference on Computer Vision, which took place in the United States [13]. SIFT, which stands for scale invariant feature transform, was the source of inspiration for SURF. When compared to SIFT, the SURF descriptor is both quicker and more reliable.

It is possible to calculate SURF, which is an integer approximation of the determinant of the hessian blob detector, using three different different approaches. In addition, the SURF descriptor is utilized in the process of locating goods, recognizing faces, reconstructing three-dimensional scenes, and tracking things.

As a Gaussian smoothing approximation, square-shaped filters are utilized in the SURF [14] algorithm. To filter the image with a square, which is significantly more efficient, integral photos are the only ones that may be used. It is possible to state it as follows:

$$S(x, y) = \sum_{i=0}^{x} \sum_{j=0}^{y} I(i, j) \qquad (1)$$

The inclusion of real images within the rectangle can be evaluated using integral images. It required four evaluations at the rectangle's four corners.

SURF can also be used as a hessian matrix-based blob detector. The determinant of a hessian matrix can be used to calculate the local change around points. If the determinant is maximal, these spots are chosen. The Hessian determinant is used by SURF to select the scale.

Given a point p=(x, y) in an image I, the Hessian matrix H(p, σ) at point p and scale σ, is defined as follows [15]:

$$H(p, \sigma) = \begin{pmatrix} L_{xx}(p, \sigma) & L_{xy}(p, \sigma) \\ L_{xy}(p, \sigma) & L_{yy}(p, \sigma) \end{pmatrix} \qquad (2)$$

Where$L_{xx}(p, \sigma)$ etc. are the second-order derivatives of the grayscale image.

Analysis of Blobs

Blob detection algorithms [14] are used to recognise sections in digital photographs that have different qualities from the surrounding regions, such as brightness or hue. Blob is a term used to describe a region of an image where some attributes are nearly constant. The stages of the Blob Analysis are as follows:

Extraction: It detects a region belonging to a single object or objects using an image thresholding technique.

Refinement: It employs transformation techniques for region refining.

Analysis: It is the final step in the refinement process. Divide the region into several blobs for investigation if it contains many objects.

Pattern Matching

The act of testing a given sequence of tokens for the presence of the constituents of some pattern is known as pattern matching [16]. In this study, the Nave Bayes idea is used to classify disease via pattern matching.

Naïve Bayes (NB) Classifier [16]

The Bayes theorem serves as the foundation for the classification approach known as the nave bays classifier. When dealing with extremely large data sets, the Naive Bayes model is quite useful. Even the most complicated classification algorithms are outperformed by it in the majority of instances. These models are utilized rather frequently in applications that involve machine learning. One of the fundamental tenets of the Naive Bayes classifier is that the presence of a particular feature inside a category does not influence the presence of any other feature within that category. Because of this quality, every component is able to make an equal contribution to the final result while still maintaining its independence from the other components. These models are adaptable and can be utilized in a wide range of contexts. In the process of making decisions, NB classifiers are considered to be computationally efficient. Numerous classes can be predicted with the help of these classifiers.

III. RESULTS AND DISCUSSIONS

Matlab R2020b was used to implement the work that was given in this research for the purpose of identifying diseases that affect fruit. Case studies are used to investigate illnesses such as apple rot, apple scab, mango anthracnose, and mango fruit fly. Apple and mango fruit are used as the subjects of the research. JPEG format is used to capture and save images of apple and mango fruit samples—both those that are not contaminated and those that are sick.

First, these photos are reduced in size to 480 * 640 pixels, then they are translated from srgb (standard rgb format) to lab color format, and finally, the k-mean segmentation method is utilized to split them into clusters. The following four features are taken into consideration after the segmentation process: color, texture, morphology, and hole structure. These features are retrieved with the help of SURF, which stands for Speedup Robust Feature Extractor. The command save ("new name.mat, "feature variable") was utilized in order to extract features from each and every photograph that was included in the dataset and then store them in the form of a matrix. The query picture feature will then be matched with the features matrix by the naive Bayes classifier, which will then produce an output based on it, based on these features. Figures 2 through 6 present the findings of the study.

A Nave Bayes (NB) classifier is used to assess whether or not the image is flawed, and if it is, it detects whether or not the fault was produced by a disease, such as apple rot or apple scab. Under the circumstances of mango anthracnose, the classifier will ascertain if the query image was recognized by mango fruit fly or mango anthracnose. In addition, this system offers disease remedies, such as determining which pesticides should be applied and instructions on how to maximize yield output. In the case of apples, the nave bayes classifier has a 57.63 percent accuracy rate, while the NN has a 42.2 percent accuracy rate. On the other hand, when it comes to mangoes, the nave bayes classifier has a 75.7 percent accuracy rate, while the NN has a 46.5 percent right rate.

The confusion matrix of the NB classifier provides an indication of the proportion of photographs belonging to a particular category that are correctly identified out of the total number of images. Both the first column and the first row of the confusion matrix provide information regarding the amount of photographs that have been correctly identified out of the total number of apple rot images. Figure 2 presents a sample of fifteen photos taken from various databases. One of the many capabilities that NB possesses is the ability to detect exactly. Figures 5 and 6 illustrate the operation of the feed forward neural network, which is the foundational technology, as well as the performance of the NB-based technique. The blue solid line in the graph illustrates how an NB-based technique accurately detects a variety of diseases that can affect apples and mangoes, whereas the red dotted line illustrates how a NN-based technique detects diseases.

In contrast to the Feed Forward Neural Network, the NB classifier displays a higher degree of precision and may be implemented with a greater degree of speed. Comparatively, the Nave Bayes Classifier takes 0.0113 seconds to identify a disease, whereas the Neural Network takes 0.3959 seconds to do it. NN's ability to correctly categorize a condition is demonstrated via the NN Confusion Matrix. When it comes to classification, the green cells represent the correct classification, while the red cells represent the erroneous classification.

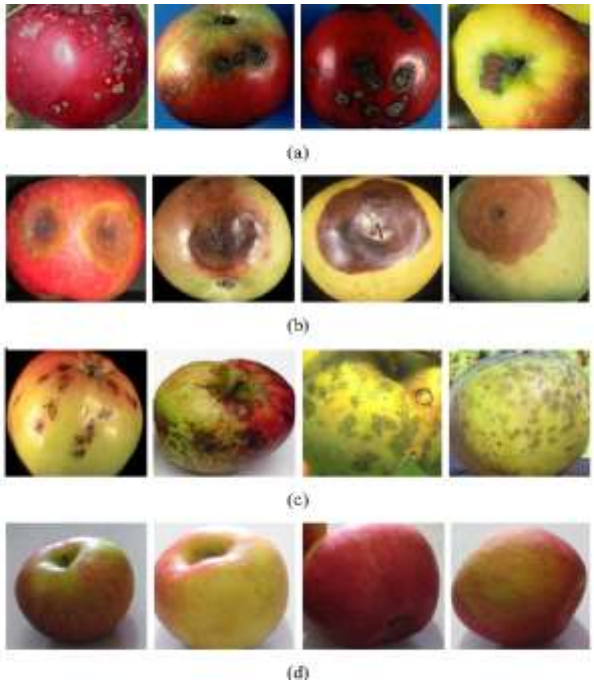


Fig 2: Sample images from data set of type (a) Apple scab, (b) Apple rot, (c) Apple blotch, and (d) Normal apple

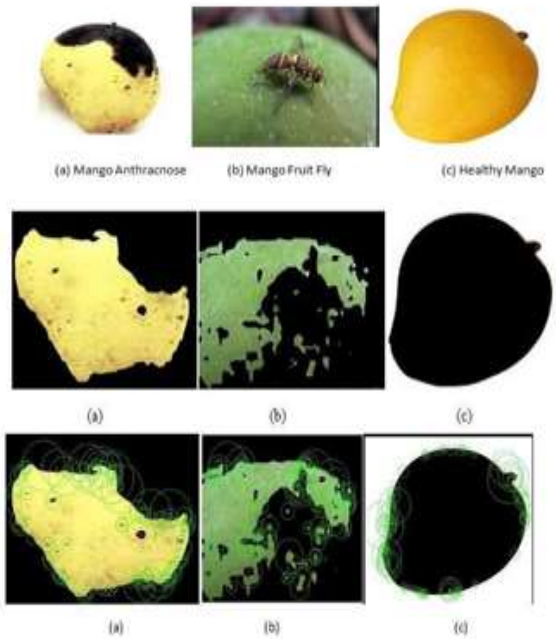


Fig 3: a) Mango Fruit Samples, b) Segmented Portions of Mango Sample, c) Extracted Features of Segmented of Mango using SURF

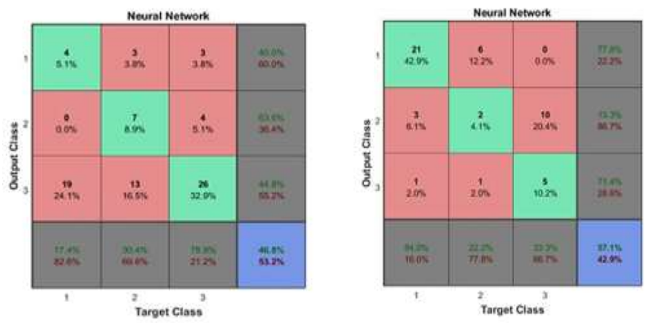
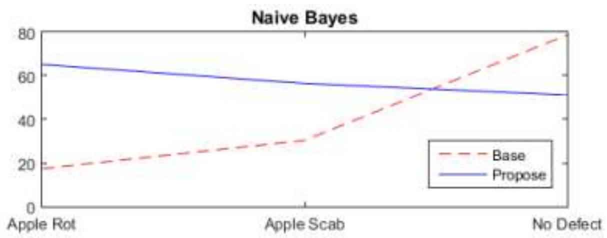
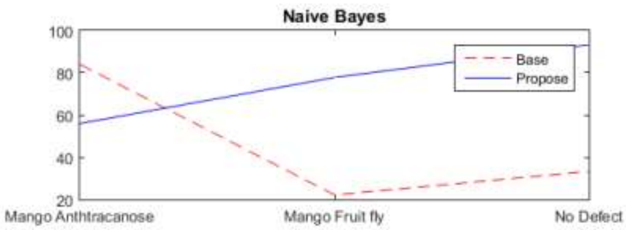


Fig 4: Confusion Matrix for Apple Disease Detection and Mango Disease Detection respectively by using Neural Network



	Apple Rot	Apple Scab	No defect
1	15	0	8
2	2	13	8
3	11	5	17

Fig 5: Performance of Naïve Bayes for Apple disease detection



	Mango Anthracanose	Mango Fruit	No defect
1	14	11	0
2	1	7	1
3	0	1	14

Fig 6: Performance of Naïve Bayes for Mango disease detection

The figures 5 and figure 6 depict the confusion matrix for diseases in Apple and Mango fruits using Naïve Bayes. (The upper graph with X is redundant).

IV. CONCLUSION

The purpose of this work was to investigate a variety of approaches to dividing the sick area of the plant. Additionally, this study investigated the development of feature extraction and classification algorithms for the purpose of extracting diseased leaf characteristics and classifying plant diseases of various kinds. The objective of this research is to enhance the performance of automatic illness diagnostics for fruit. According to the findings, the method that was proposed is not only useful, but it also has the potential to contribute to a relatively accurate study of fruit illnesses in particular. Regarding the implementation, the Nave Bayes classifier is not only more accurate but also significantly faster than other classifiers. It is recommended that the appropriate treatment be administered once these fruit diseases have been identified. In contrast, the work that will be done in the future can involve the development of better data sets in order to enhance the training component. This is due to the fact that the dataset that is currently being used might not be adequate for training the classifier. As a result, another way of imaging the fruit that is suitable for both training and testing is necessary.

.References

[1] Asad Khattak, Muhammad Usama Asghar, Ulfat Batool, Muhammad Zubair Asghar, Hayat Ullah, Mabrook Al-Rakhami, (Member, Ieee), And Abdu Gumae, Automatic Detection of CitrusFruit and Leaves Diseases Using Deep Neural Network Model, IEEE access, VOLUME 9, 2021.

[2] Umar, Syed, Bommina Naveen Sai, Nagineni Sai Lasya,Doppalapudi Asutosh, and LohithaRani. "Machine Learning based Sentiment Analysis of Product Reviews Using DeepEmbedding." Journal of Optoelectronics Laser 41, no. 6(2022): 108-113.

[3] Habeeb, M. S., & Babu, T. R. (2022). Network intrusion detection system: a survey on artificial intelligence‐based techniques. Expert Systems, 39(9), e13066.

[4] Han, L.; Haleem, M.S.; Taylor, M. A Novel Computer Vision-based Approach to Automatic Detection and Severity Assessment of Crop Diseases. 2015, on 20 December 2021).

[5] Nandipati Sai Akash, Naveen Sai Bommina, Uppu Lokesh, Hussain Syed, Syed Umar, "Optimized Block Chain-Enabled Security Mechanism for IoT Using Ant Colony Optimization", International Journal on Recent and Innovation Trends in Computing and Communication, (2023), 11(10), 1226–1233.

[6] Jahanbakhshi, A.; Momeny, M.; Mahmoudi, M.; Zhang, Y.D. Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks. Sci. Hortic. 2020, 263, 109133.

[7] Uppu Lokesh, Naveen Sai Bommina, Nandipati Sai Akash, Dr. Hussain Syed, Dr. Syed Umar, "Designing Energy-Efficient and Secure IoT Architectures Using Evolutionary Optimization Algorithms", International Journal of Applied Engineering & Technology, Vol. 4 No.2, September, 2022.

[8] R. Gnanakumaran, Divya Rohatgi, A K Sampath, Nidhi Nagar, D. Amuthaguka, Raj Kumar Gupta, "Robust Extreme Learning Machine based Sentiment Analysis and Classification", 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), (2023), DOI: 10.1109/ICSSIT55814.2023.10061017.

[9] Habeeb, M. S., & Babu, T. R. (2024). MS-CFFS: Multistage Coarse and Fine Feature Selection for Advanced Anomaly Detection in IoT Security Networks. International Journal of Electrical and Electronics Research, 12(3), 780-790.

[10] Uppu Lokesh , Naveen Sai Bommina , Nandipati Sai Akash , Dr. Hussain Syed , Dr. Syed Umar. (2021). Deep Reinforcement Learning with Genetic Algorithm Tuning for Intrusion Detection in IoT Systems. International Journal of Communication Networks and Information Security (IJCNIS), 13(3), 582–595.

[11] RS Supriya Khaitan, Divya Rohatgi, Sana Nalband, Tejali Mhatre, Shweta Patil, "Enhancing Essay Grading Efficiency and Consistency through Two-Layer LSTM Models and Attention Mechanisms", Journal of Information Systems Engineering and Management 10 (2), 191-202.

[12] Naveen Sai Bommina, Uppu Lokesh, Nandipati Sai Akash, Dr. Hussain Syed, Dr. Syed Umar, "Optimizing AI-Driven Security Protocols in IoT Networks Using Metaheuristic Algorithms", International Journal of Intelligent Systems and Applications in Engineering, IJISAE, 2024, 12(23s), 3339–3347.

[13] Habeeb, M. S., & Babu, T. R. (2024). Coarse and fine feature selection for network intrusion detection systems (IDS) in IoT networks. Transactions on Emerging Telecommunications Technologies, 35(4), e4961.

[14] K Sankar, Divya Rohatgi, S Balakrishna Reddy, "COX Regressive Winsorized Correlated Convolutional Deep Belief Boltzmann Network for Covid-19 Prediction with Big Data", Grenze International Journal of Engineering & Technology (GIJET), Grenze ID: 01.GIJET.9.1.547, © Grenze Scientific Society, 2023.

[15] Nandipati Sai Akash, Uppu Lokesh, Naveen Sai Bommina, Hussain Syed, Syed Umar, "Swarm Intelligence-Based Hyperparameter Optimization for AI-Powered IoT Threat Detection", International Journal of Intelligent Systems and Applications in Engineering, (2024), 12(17s), 941.

[16] Abbas, A.; Jain, S.; Gour, M.; Vankudothu, S. Tomato plant disease detection using transfer learning with C-GAN synthetic images. Comput. Electron. Agric. 2021, 187, 106279.