**ABSTRACT:**

Soybeans are a vital global crop rich in protein and oil, making them essential for food production, livestock feed, and biofuel. Nowadays it is impossible and time consuming for humans to predict the defects in soybean seeds. Since soybeans are useful in today's global market it is essential for identifying the defects in soybean seeds to improve the production and also the quality of soybeans. Our objective is to predict the defects in soybean seeds using machine learning models. Initially the soybean seed data set was extracted from nearly 5000 different kinds of soybean seed images. The second phase is to develop a suitable machine learning model that can predict the defect of soybean seed. For ease we had categorized the soybean seed into: complete, broken, immature, spotted, and skin-damaged. The dataset includes key seed features such as size, weight, color, shape and using these key features and other supporting features we had predicted the basic defects in soybean seeds.

**KEYWORDS:**

Machine Learning, Outlier Detection, Classification,

**1.INTRODUCTION:**

Soybean seed classification, a very important tool in agriculture, has been pivotal in crop quality and market value in the soybean business. Soybeans are used in food production, animal feed, and other biofuels. Traditionally, seed companies offered a range of seed options to farmers based on multi-year yield performance trials conducted at various test farm locations. Farmers then choose seed varieties that are suitable for their farm's soil and climatic conditions, optimizing their crop yield. Quality soybean seeds, depending on characteristics such as size, shape, color, and physical damage, affect crop productivity, nutrition, and processing efficiency. While traditional manual inspection has been the norm for assessment of seed quality, it is time consuming and inclined to human error. There is thus a great need for machine learning as a transformative solution in the automation of prediction. In this work, ML algorithms was intensively used in the classification of soybean seeds, based on numerical data obtained from physical characteristics that are elicited by images from the seeds. Image-based classification techniques, although effective, heavily rely on computational resources for their processing; our approach uses structured datasets in CSV format with encapsulated numerical attributes like seed weight, dimensions, moisture content, and others. The datasets are further enhanced with features extracted from seed images, including edge detection and shape analysis, texture metrics, color distribution, and correlation measures. These features play a critical role in seed physical condition assessment and defect identification.

Key features such as edge and shape analysis capture geometric properties like seed boundaries and contour irregularities, enabling the identification of broken or misshaped seeds. Texture and correlation metrics, derived from the Gray-Level Co-occurrence Matrix (GLCM), help detect visual defects such as discoloration or contamination. Color distribution provides insights into seed maturity and potential damage due to environmental factors, while Histogram of Oriented Gradients (HOG) features highlight subtle surface irregularities and biases. Preprocessing involved applying necessary scaling, normalization, and dimensionality reduction for optimizing model performance. However, these techniques help the models to increase their efficiency as training is quicker and the performance is not hampered.

**2.LITERATURE SURVEY:**

Quadras et al. [1] suggested a three-stage process integrating discrete event simulation and machine learning to predict lead and queue times in soybean seed classification. They also created SSDINet, a deep neural network employing contour detection and basic CNN blocks to enhance seed quality identification accuracy. Sable et al. [2] introduce a computational model for the detection and measurement of defects in soybean seeds based on deep learning. They introduce a light-weight model, SSDINet, that utilizes a proprietary contour detection algorithm and combines CNN, depthwise convolution, and squeeze-and-excitation blocks to perform efficient and accurate classification. SSDINet surpasses existing state-of-the-art models and is a reliable tool for automated quality evaluation.Gadotti et al. [3] utilized machine learning to enhance soybean seed germination and vigor determination, overcoming human quality control restrictions. Random Forest and Classification by Regression yielded the highest accuracy, allowing low-cost seed lot ranking and efficient storage strategies.Ziliang Hua et al. [4] introduced an end-to-end segmentation-classification pipeline based on SNet, a light CNN network with Mixed Feature Recalibration to perform efficient soybean seed classification. SNet exceeded six state-of-the-art models and was effective for automated seed quality checking, particularly on low-resource platforms.de Medeiros, A.D [5] emphasized the drawbacks of traditional, time-consuming seed testing procedures and suggested Interactive Machine Learning (IML) as a quicker, more precise alternative. Software such as Ilastik also indicated potential in correlating seed appearance with physiological performance for enhanced quality evaluation.Yang, Y et al. [6] applied histogram-based imaging and a 1D-CNN model to detect high-oil-content soybean seeds, examining 5,510 samples of 58 varieties. The 1D-CNN performed better than traditional models such as SVM, KNN, and PLS-DA, with Multivariate Scattering Correction (MSC) producing the best preprocessing outcome.Torsoni, G.B. [7] applied Machine learning techniques such as MLR, MLP, SVM, Random Forest, and XGBoost were used with a 70/30 train-validation split and obtained high precision and accuracy. Major climate variables such as 26\_12\_ARM and 2\_10\_TDEW were identified to have significant impacts on soybean yields at certain phenological phases.

Oliveira DC de et al. [8] introduce a supervised machine learning method to classify soybean grains as standard or non-standard in order to improve accuracy and efficiency in agricultural decision-making. The paper highlights the potential of new technologies such as AI, IoT, and big data to revolutionize agriculture. It also presents a website that employs intelligent computing to make real-time predictions consistent with MAPA regulations to facilitate the use of technology in grain classification.Fletcher, R. S. et al. [9] utilized NDVIs from multispectral leaf reflectance and Random Forest models to separate soybeans from three broadleaf weeds under greenhouse conditions. The research indicated NDVIs were as good or superior to raw multispectral data for classification and emphasized the significance of spectral sensitivity in remote sensing for weed detection. Cetin, N. et al. [10] designed classification models to discriminate soybean seeds on the basis of physical characteristics such as shape, size, and weight utilizing RF, SVM, KNN, Decision Tree, and MLP algorithms. MLP and Random Forest were the best-performing, utilizing majority voting for stable classification. Statistical tests also identified significant differences between seed characteristics, and the work complements current research by enhancing binary classification of soybean seeds through machine learning. Sundaramoorthi, D et al. [11] created a decision support tool with Syngenta data and machine learning for forecasting soybean yield performance under differing weather and soil conditions. The tool supports farmers to maximize crop choice for increased yields by providing data-driven analysis.

H. Pratap [12] proposed a supervised machine learning method for the classification of soybean grains as standard or non-standard, providing a more efficient and robust technique. Ferraz & Pinto (2017) mentioned that technologies such as weather instruments and commodity forecasting systems improve agricultural decision-making by minimizing risks and maximizing profits. Supervised learning, already well established in agriculture, provides highly accurate and reliable results.Gulzar, Y et al. [13] used a modified Inception V3 model with transfer learning and adaptive learning rate methods to classify faulty soybean seeds with high precision and recall. This machine-based method provides a cost-effective, scalable alternative to the conventional manual seed quality determination..Wei Lin et al. [14] created a web-based application by means of deep learning for soybean seed classification, enhancing quality evaluation under non-uniform lighting. Seeds were classified as normal, damaged, abnormal, or unclassifiable with lightweight CNNs, providing a quicker and more objective method in place of conventional practices. Running on the NVIDIA Jetson TX2, the model was able to process in real-time, demonstrating suitability for non-destructive, cost-effective seed quality evaluation in resource-scarce environments.Melero, F. C et al. [15] also emphasized the significance of seed quality and its relation to genome integrity for maximum growth. Seed germination, vigor tests, and DNA analysis by electrophoresis and the comet assay revealed large differences in DNA stability. These results underscore the requirement to preserve DNA integrity for high-quality seeds and enhanced productivity of crops.

**3.METHODOLOGY:**

The dataset is preprocessed by converting the categorical target variables into a coded format using Label Encoder and the features are normalized using Standard Scaler such that all the features will have a mean of zero and a standard deviation of one. The dataset's dimensionality is then reduced to to facilitate easier visualization using principal component analysis (PCA).For eliminating noise in the data, DBSCAN is used with different epsilon(eps) values (0.2,0.3,0.4,0.5). This data clustering algorithm marks and eliminates noise using cosine distance between data points. Subsequent to noise elimination, four classification models- Random Forest, Decision Tree, Naive Bayes and XGBoost- are trained and tested on the data after cleaning with an 80/20 train-test split. The accuracy, precision, recall and f1-score of each model measured and the best classifier and value of eps are chosen based on the evaluation.The effect of noise removal is monitored by quantifying the amount of noise points eliminated at each value of eps, with visualization consisting of PCA scatter plots, accuracy vs. epsilon plots and bar plots of noise distribution by class. This approach pairs DBSCAN with noise reduction with several classification models to improve model performance. The ultimate analysis selects the best performing model and optimal value of eps, which combined result in increased classification accuracy and a cleaned dataset

Table 3.1 Distribution of Seed varieties

| **Labels** | **Count** |
| --- | --- |
| Skin damaged | 1128 |
| Intact | 1201 |
| Broken | 1089 |
| Spotted | 1058 |
| Immature | 1125 |

Table 3.1 illustrates the distribution of seed images categorized by skin condition labels. Among the five classes, 'Intact' has the highest count with 1201 samples, followed by 'Skin damaged' (1128), 'Immature' (1125), 'Broken' (1089), and 'Spotted' (1058). The dataset appears relatively balanced, supporting robust training and evaluation of classification models.

Table 3.2 Visual Classification of Soybean Seed Conditions

| Broken |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Immatured |  |  |  |  |  |
| Intact |  |  |  |  |  |
| Skin-damaged |  |  |  |  |  |
| Spotted |  |  |  |  |  |

This table shows the soybeans sorted into Broken,Immature,Intact and Skin-damaged classes,presenting variations in physical form for quality grading.Each type indicates certain defects.

Table 3.3 Feature extraction and Visualisation for Soybean Seed Image Analysis.

|  |  |  |
| --- | --- | --- |
| a.Original | b. Grayscale | c.Color histograms |
|  |  |  |
| d.LBP Histograms | e.GLCM Features | f.Gabor real part |
|  |  |  |
| g.Gabor imaginary part | h.HOG Visualization | i.Color means |

Fig (a-i) illustrates the Features extracted from a sample soybean image. The images show different feature extraction methods ranging from an image of soybean such as Color Histograms,LBP,GLCM,Gabor Features,HOG Visualization, and Color Means,utilized for in-depth texture and color analysis.

Table 3.4 Summary of Extracted Features for Soybean Seed Image Analysis

| **Feature Type** | **Description** | **Value(s)** |
| --- | --- | --- |
| Color Mean | Mean color (R, G, B) | [0.3672, 0.3938, 0.3970] |
| Color Std | Standard deviation of color (R, G, B) | [0.2571, 0.2832, 0.2657] |
| Color Skewness | Skewness of color channels (R, G, B) | [-0.0033, -0.0060, -0.0083] |
| GLCM - Contrast | Texture contrast | 24.6764 |
| GLCM - Correlation | Texture correlation | 0.9975 |
| GLCM - Energy | Texture energy | 0.1538 |
| GLCM - Homogeneity | Texture homogeneity | 0.4461 |
| HOG Shape | Shape of HOG feature vector | (72900,) |
| LBP Histogram | Histogram of LBP (first 10 values) | [0.0069, 0.0204, 0.0189, 0.1199, 0.3176, 0.1940, 0.0572, 0.0303, 0.1992, 0.0356] |

Table 3.4 represents extracted features such as Color Mean,Color Std,Color Skewness,GLCM features,HOG Shape and LBP histogram and their description with respective values.The features are responsible for retaining significant color,texture and shape details for image analysis.

Fig 3.1 Pair Plot of Soybean Classification Based on Feature Relationships

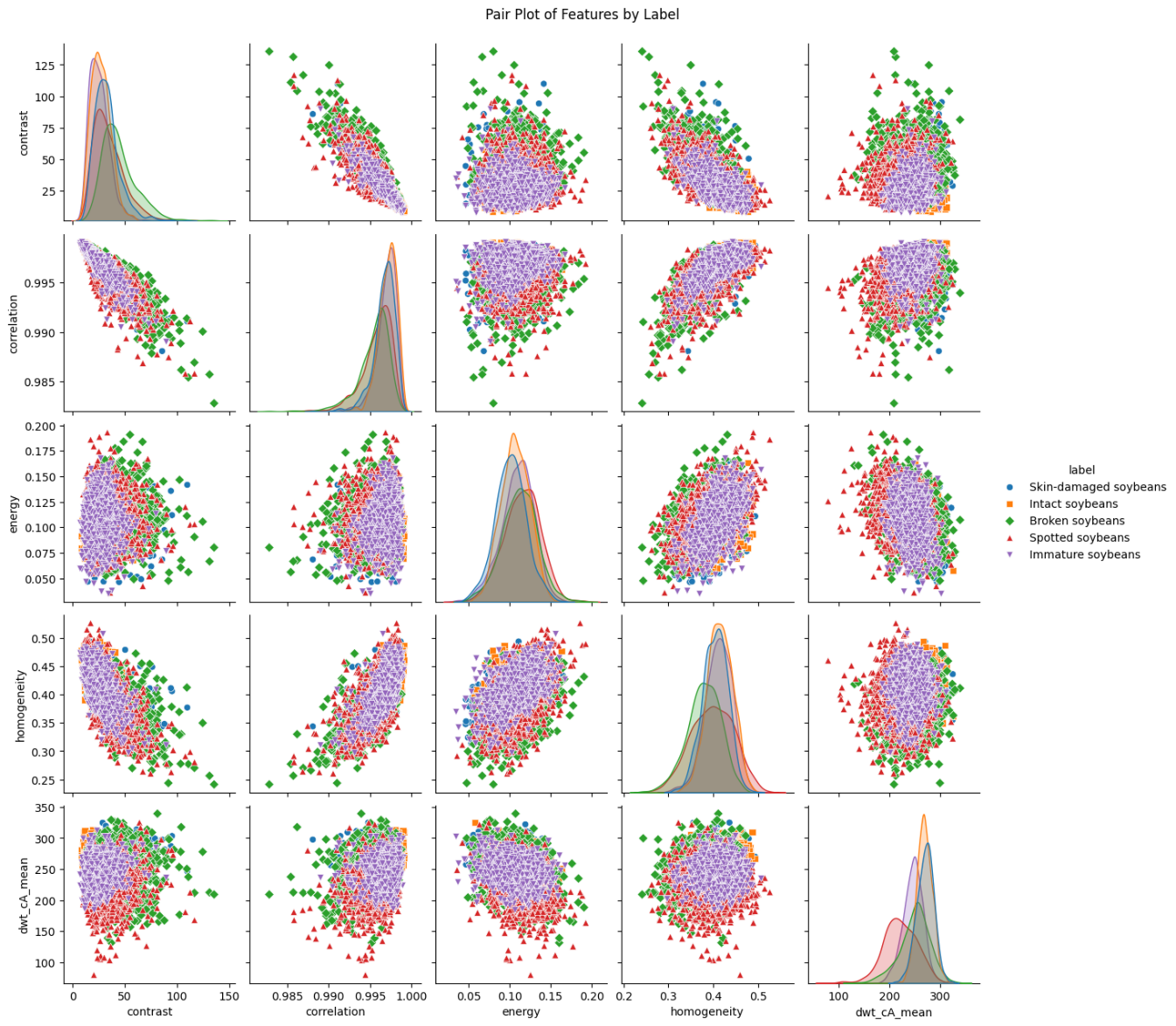
****

Fig 3.1 a pair plot showing feature relationship in soybean classification.The diagonal plots display kernel density estimates while scatter plots reveal correlations.

Fig 3.2 Feature Correlation Matrix

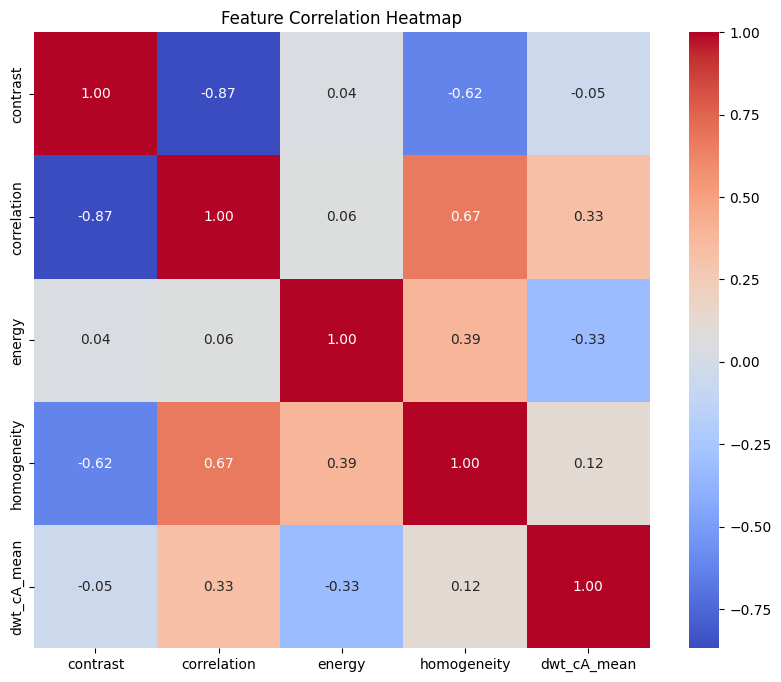
****

Fig 3.2 heatmap of correlations between features for a soybean dataset.Positive correlations are indicated in red and negative correlations are indicated in blue. The most correlated feature is homogeneity.

Fig 3.3 Bar chart of feature importance

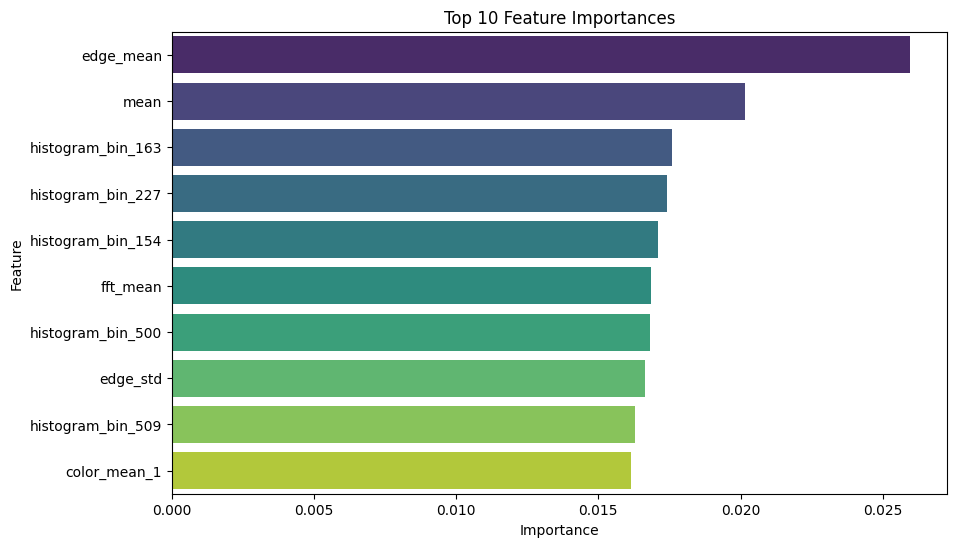
****

Fig 3.3 represents the top 10 most influential features in a model. The importance of “Edge\_mean” is the maximum, followed by “mean” and some of the histogram features.

XGB Classifier is an ensemble learning model that improved the accuracy of the

prediction. These base models are initially trained individually on the same dataset. Each model has different accuracy to prediction. Advantages of using this classifier is, the accuracy is improved and also flexibility. We also used up sampling which is a technique used to increase the number of instances in a dataset. The invocation of up sampling technique and utilization of XGB Classifier produced a better accuracy compared to other models and classifiers.











**Input data**















Fig 3.4

**4.RESULTS AND DISCUSSIONS:**

Table 4.1 Performance Metrics of Classification Models

| MODEL | ACCURACY | PRECISION | RECALL | F1 SCORE |
| --- | --- | --- | --- | --- |
| XGBoost | 97.72 | 98 | 98 | 98 |
| Random Forest | 97.17 | 97 | 97 | 97 |
| Decision Tree | 94.70 | 95 | 95 | 95 |
| Naive Bayes | 54.04 | 65 | 54 | 52 |

Table 4.1 illustrates the four performance metrics—Accuracy, Precision, Recall, and F1 Score—of four different classification models.XGBoost has the best scores on all four metrics, followed by Random Forest, with Naive Bayes doing the worst among the four models.

Fig 4.1 Performance Metrics

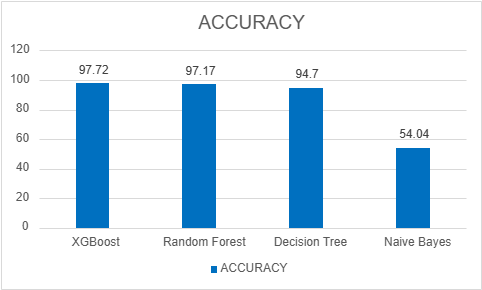


Fig 4.1 illustrates the accuracy of various machine learning models.XgBoost has the highest accuracy followed by Random Forest, Decision Tree etc.The results suggest that ensemble approaches such as XGBoost and Random Forest work better than standard models

.

Fig 4.2 Performance Measures for Different Classes

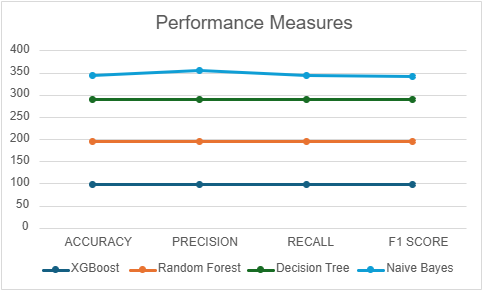


Fig 4.2 illustrates the performance of ML models based on Accuracy Precision, Recall, and F1 Score.It shows how each model fares on these most important evaluation metrics.

Table 4.2 Model Accuracy at Varying EPS Values

| EPS Value-> | EPS=0.2 | EPS=0.3 | EPS=0.4 | EPS=0.5 |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Accuracy | Accuracy | Accuracy |
| XGBoost | 99.25 | 92.67 | 87.55 | 97.72 |
| Random Forest | 97.74 | 89.18 | 84.17 | 97.17 |
| Decision Tree | 93.98 | 81.85 | 74.13 | 94.70 |
| Naive Bayes | 86.47 | 72.08 | 61.03 | 54.04 |

Table 4.2 illustrates the performance of four models of classification—XGBoost, Random Forest, Decision Tree, and Naive Bayes—on varying EPS (perturbation) values between 0.2 and 0.5. It indicates that a systematic loss in performance exists across all the models with every rise in EPS value, XGBoost reporting the best performance on all scales.

Fig 4.3 Impact of EPS value on Model Accuracy

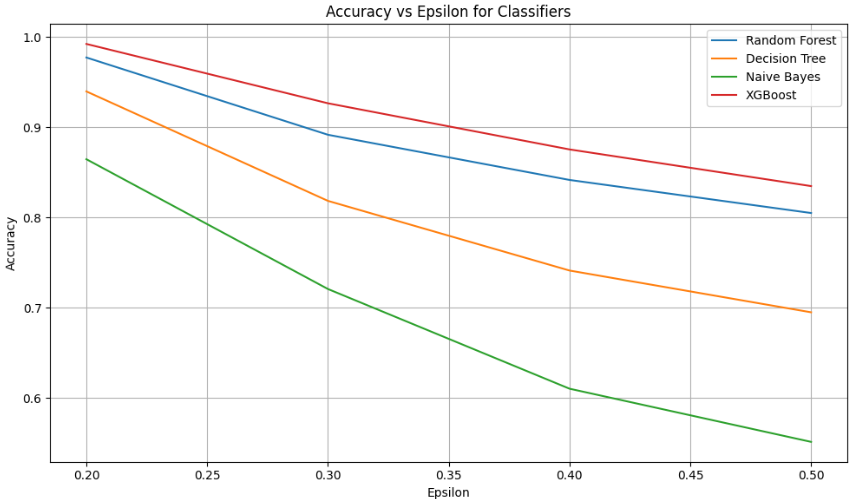


Fig 4.3 illustrates how the accuracy of four machine learning models,namely XGBoost,Random Forest, Decision Tree, and Naive Bayes, changes with increasing EPS values.Among the models, XGBoost always produces the highest accuracy for all EPS values,whereas Naive Bayes exhibits the sharpest decline.

Fig 4.4 Confusion matrix Fig 4.5 Confusion matrix

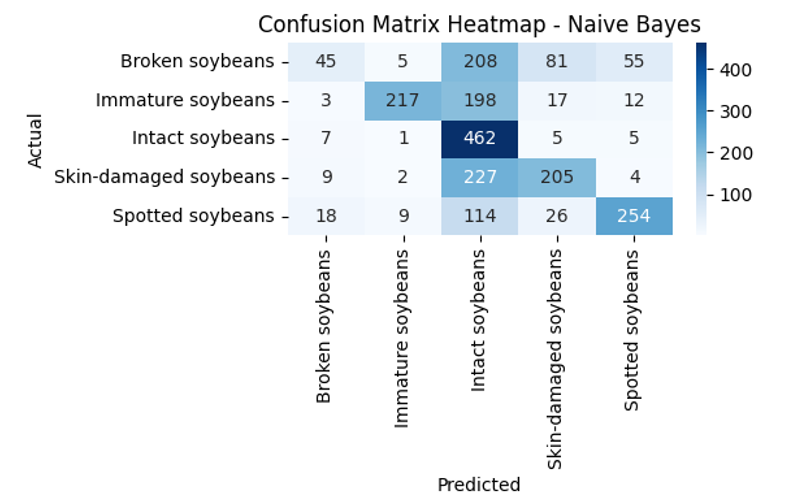
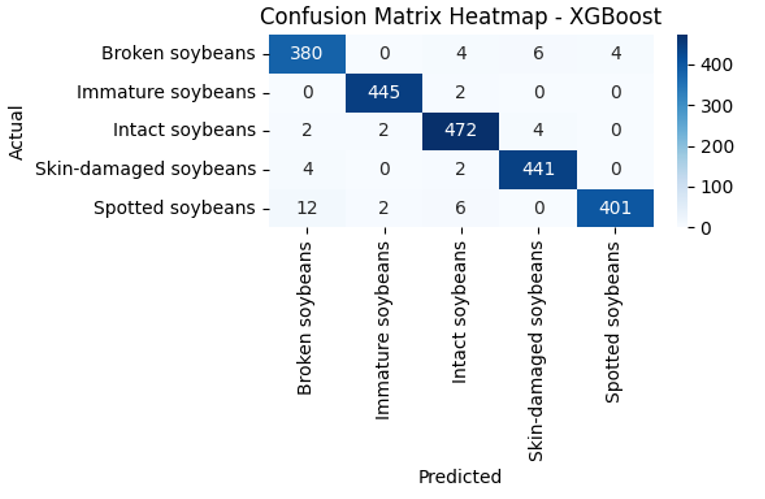


Fig 4.4 and Fig 4.5 shows the confusion matrix for the XGBoost and Naive Bayes Models across multiple classes.It highlights the Correct and incorrect predictions made for each class.

Fig 4.6 Confusion matrix Fig 4.7 Confusion matrix

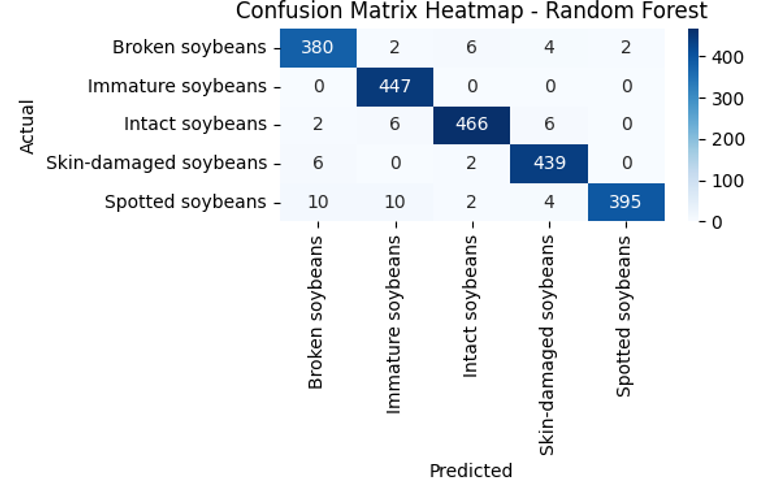
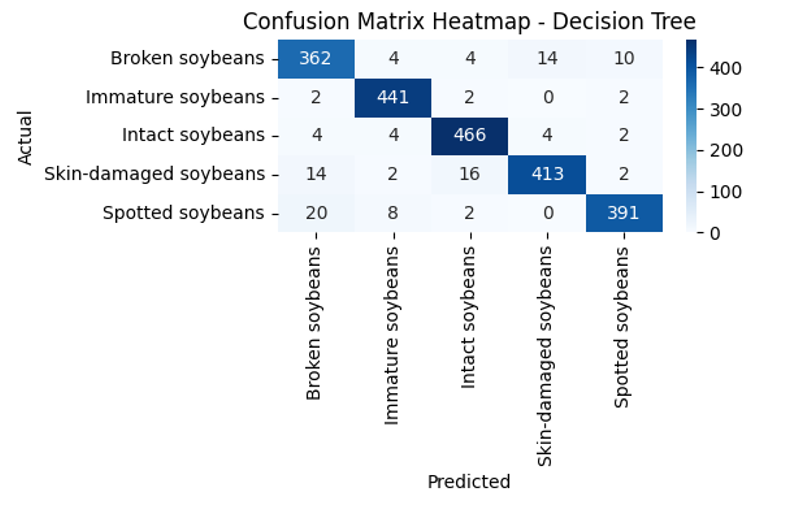


Fig 4.6 and Fig 4.7 shows the confusion matrix for the Decision Tree and Random Forest Models across multiple classes.It highlights the Correct and incorrect predictions made for each class.

**ROC CURVE**

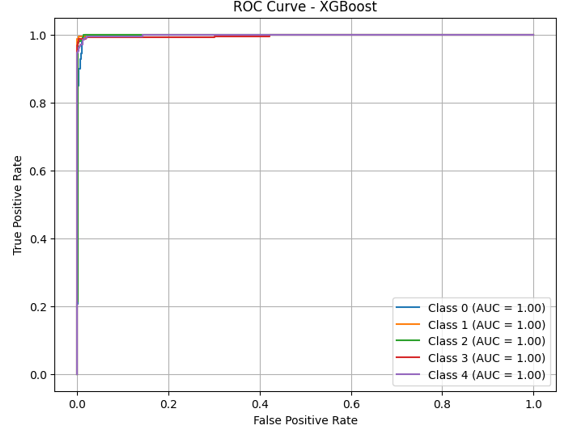
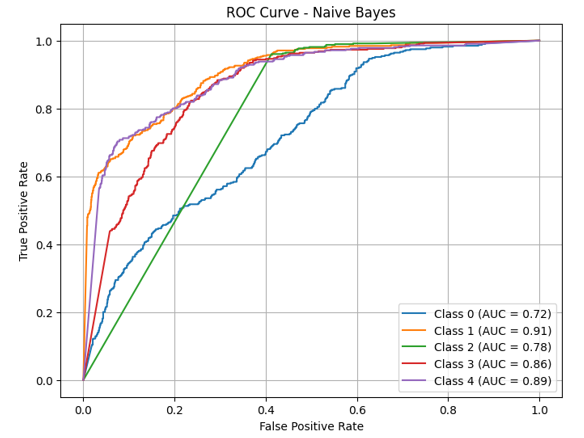
Fig 4.8 ROC Curve Fig 4.9 ROC Curve 

Fig 4.8 and 4.9 shows the ROC curves to compare the performance of Naive Bayes and XGBoost classifiers on five classes. Naive Bayes performs moderately, with lower AUC for Class 0 (0.72) and higher for others (up to 0.91). Conversely, XGBoost has ideal AUC (1.00) for all classes, demonstrating great prediction quality.

Fig 4.10 ROC Curve Fig 4.11 ROC Curve

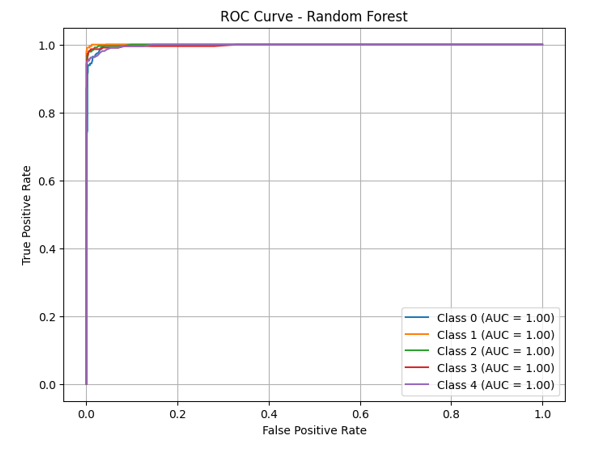
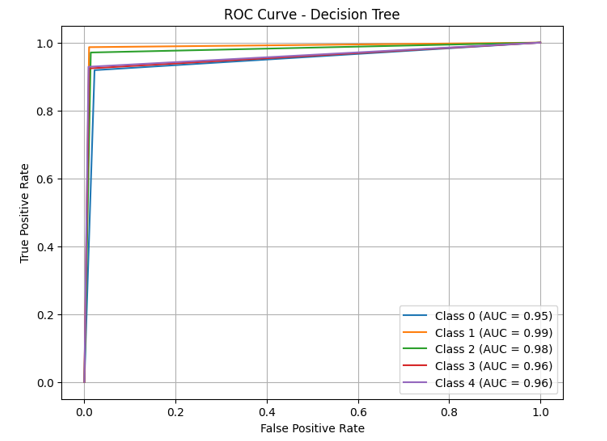


Fig 4.10 and 4.11 shows the ROC curves for Decision Tree and Random Forest classifiers across the same five classes. Decision Tree is good for AUC = 0.95 to 0.99, showing that the classification is strong but a little less accurate. Random Forest, similar to XGBoost, has perfect AUC (1.00) for all classes, indicating excellent model performance.

Fig 13 Misclassification Comparison of ML Models at Epsilon=0.2

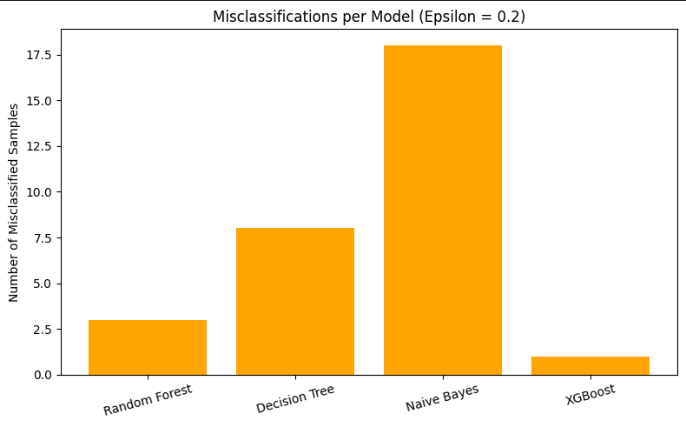


Fig 13 demonstrates the misclassified samples by various ML models at Epsilon=0.2. Naive Bayes has the greatest number of misclassifications, while XGBoost has the least.

**REFERENCES:**

* **[1]** Quadras, D. L. de O., Cavalcante, I., Kück, M., Mendes, L. G., & Frazzon, E. M. (2023). Machine Learning Applied to Logistics Decision Making: Improvements to the Soybean Seed Classification Process. Applied Sciences, 13(19), 10904. <https://doi.org/10.3390/app131910904>
* **[2]** Sable, A., Singh, P., Kaur, A., Driss, M., & Boulila, W. (2024). Quantifying Soybean Defects: A Computational Approach to Seed Classification Using Deep LearningTechniques.Agronomy,14(1098). [https://doi.org/10.3390/agronomy14061098&#8203;:contentReference[oaicite:0]{index=0](https://doi.org/10.3390/agronomy14061098&#8203;:contentReference%5Boaicite:0%5D%7Bindex=0)}.
* **[3]** Gadotti, G. I., Ascoli, C. A., Bernardy, R., Monteiro, R. de C. M., & Pinheiro, R. de M. (2022). Machine Learning for Soybean Seeds Lots Classification. Engenharia Agrícola, 42(Special Issue), e20210101. <https://doi.org/10.1590/1809-4430-Eng.Agric.v42nepe20210101/2022>
* **[4]** Ziliang Huang, Rujing Wang, Ying Cao, Shijian Zheng, Yue Teng, Fenmei Wang, Liusan Wang,Jianming Du,Deep learning based soybean seed classification,Computers and Electronics in Agriculture,Volume202,2022,107393,ISSN0168-1699,<https://doi.org/10.1016/j.compag.2022.107393>.
* **[5]** de Medeiros, A.D., Capobiango, N.P., da Silva, J.M. et al. Interactive machine learning for soybean seed and seedling quality classification. Sci Rep 10, 11267 (2020). <https://doi.org/10.1038/s41598-020-6827>
* **[6]** Yang, Y., Liao, J., Li, H., Tan, K., & Zhang, X. (2023). Identification of high-oil content soybean using hyperspectral reflectance and one-dimensional convolutional neural network. Spectroscopy Letters, 56(1), 28-41. <https://doi.org/10.1080/00387010.2022>
* **[7]** Torsoni, G.B., Aparecido, L.E.O., dos Santos, G.M., Chiquito, A.G., Moraes, J.R.S.C., & Rolim, G.S. (2023). Soybean yield prediction by machine learning and climate. Theoretical and Applied Climatology, 151, 1709–1725. [https://doi.org/10.1007/s00704-022-04341-9&#8203;:contentReference[oaicite:0]{index=0](https://doi.org/10.1007/s00704-022-04341-9&#8203;:contentReference%5Boaicite:0%5D%7Bindex=0)}.
* **[8]**Oliveira DC de, Barbosa UC, Bergland ACRO, Resende O, Oliveira DEC de. G-SOJA - WEBSITE WITH PREDICTION ON SOYBEAN CLASSIFICATION USING MACHINE LEARNING. Eng Agríc [Internet]. 2022;42(spe):e20210140. Available from: <https://doi.org/10.1590/1809-4430-Eng.Agric.v42nepe20210140/2022>
* **[9]** Fletcher, R. S. (2016). Using vegetation indices as input into random forest for soybean and weed classification. American Journal of Plant Sciences, 7(15), 2186-2198. [https://doi.org/10.4236/ajps.2016.715193&#8203;:contentReference[oaicite:0]{index=0](https://doi.org/10.4236/ajps.2016.715193&#8203;:contentReference%5Boaicite:0%5D%7Bindex=0)}.
* **[10]** Cetin, N. Machine Learning for Varietal Binary Classification of Soybean (Glycine max (L.) Merrill) Seeds Based on Shape and Size Attributes. Food Anal. Methods 15, 2260–2273 (2022). <https://doi.org/10.1007/s12161-022-02286-3>
* **[11]** Sundaramoorthi, D., Dong, L. Machine learning and optimization based decision-support tool for seed variety selection. Ann Oper Res (2022). <https://doi.org/10.1007/s10479-022-04995-8>
* **[12]** H. Pratap, M. S. Guru Prasad, P. Agarwal, A. Bhardwaj, A. Mehra and S. Mathpati, "SoyNet: Deep Learning Approaches for Automated Soybean Seed Quality Assessment," 2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India, 2024, pp. 613-617, doi: 10.1109/InCACCT61598.2024.10551180.
* **[13]** Gulzar, Y., 2024. Enhancing soybean classification with modified inception model: A transfer learning approach. Emirates Journal of Food and Agriculture, 36, pp.1-9.
* **[14]** Wei Lin, Lei Shu, Weibo Zhong, Wei Lu, Daoyi Ma,Yizhen Meng,Online classification of soybean seeds based on deep learning, Engineering Applications of Artificial Intelligence,Volume 123, Part C,2023,106434,ISSN 0952-1976,https://doi.org/10.1016/j.engappai.2023.106434.

(<https://www.sciencedirect.com/science/article/pii/S0952197623006188>)

* **[15]** Melero, F. C., Masetto, T. E., Rocha, L. G., Silva, D. M., Crispim, B. A., & Barufatti, A. (2024). DNA degradation is involved with low physiological potential of soybean. *Sci. Agric.*, 81, e20230250. [**https://doi.org/10.1590/1678-992X-2023-0250**](https://doi.org/10.1590/1678-992X-2023-0250)**.**
* **[16]** Park, Y.-H., & Jun, T.-H. (2022). Classification of Soybean [Glycine max (L.) Merr.] Seed Based on Deep Learning Using the YOLOv5 Model. Plant Breed. Biotech., 10(1), 75-80. [https://doi.org/10.9787/PBB.2022.10.1.75&#8203;:contentReference[oaicite:0]{index=0](https://doi.org/10.9787/PBB.2022.10.1.75&#8203;:contentReference%5Boaicite:0%5D%7Bindex=0)}.
* **[17]** Yan, H., Tian, D., Zhang, Q., Wen, J., Wang, Z.-Y., & Chai, M. (2024). GmHs1-1 and GmqHS1 Simultaneously Contribute to the Domestication of Soybean Hard-Seededness. Plants, 13(15), 2061. <https://doi.org/10.3390/plants13152061>
* **[18**] Wang, M., Ding, X., Zeng, Y., Xie, G., Yu, J., Jin, M., Liu, L., Li, P., Zhao, N., Dong, Q., Liu, B., & Xu, C. (2024). Identification of multiple genetic loci and candidate genes determining seed size and weight in soybean. Agronomy, 14(9), 1957. <https://doi.org/10.3390/agronomy14091957>
* **[19]** Chen, X., He, W., Ye, Z., Gai, J., Lu, W., & Xing, G. (2024). Soybean seed pest damage detection method based on spatial frequency domain imaging combined with RL-SVM. Plant Methods, 20, 130. [https://doi.org/10.1186/s13007-024-01257-5&#8203;:contentReference[oaicite:0](https://doi.org/10.1186/s13007-024-01257-5&#8203;:contentReference%5Boaicite:0)]{

index=0}.

* **[20]** Liang, S., Duan, Z., He, X., Yang, X., Yuan, Y., Liang, Q., Pan, Y., Zhou, G., Zhang, M., Liu, S., & Tian, Z. (2024). Natural variation in GmSW17 controls seed size in soybean. *Nature Communications, 15*(1), 7417. [**https://doi.org/10.1038/s41467-024-51798-5**](https://doi.org/10.1038/s41467-024-51798-5)
* **[21]** Zhang, L., Cao, Z., Gao, Y., Huang, W., Si, Z., Guo, Y., Wang, H., & Wang, X. (2024). Soybean Yield Simulation and Sustainability Assessment Based on the DSSAT-CROPGRO-Soybean Model. Plants, 13(17), 2525. <https://doi.org/10.3390/plants13172525>
* **[22]** Li X, He Z, Liu F, Chen R. Fast Identification of Soybean Seed Varieties Using Laser-Induced Breakdown Spectroscopy Combined With Convolutional Neural Network. Front Plant Sci. 2021 Oct 6;12:714557. doi: 10.3389/fpls.2021.714557. PMID: 34691095; PMCID: PMC8527016.
* **[23]** Wu, C., Acuña, A., Florez-Palacios, L., Harrison, D., Rogers, D., Mozzoni, L., Mian, R., & Canella Vieira, C. (2024). Across-environment seed protein stability and genetic architecture of seed components in soybean. Scientific Reports, [https://doi.org/10.1038/s41598-024-67035-4&#8203;:contentReference[oaicite:0]{index=0](https://doi.org/10.1038/s41598-024-67035-4&#8203;:contentReference%5Boaicite:0%5D%7Bindex=0)}.