Süleyman Demirel Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
Cilt 29, Sayı 1, 124-133, 2025

Süleyman Demirel University
Journal of Natural and Applied Sciences
Volume 29, Issue 1, 124-133, 2025

DOI: 10.19113/sdufenbed.1575098

XGBoost Tabanlı Derin Öğrenme Algoritması ile Açıklanabilir Yapay Zeka Modellerinin Kullanımı: Çürük Meyvelerin Tespiti

Nilgün ŞENGÖZ*[1], Harun KÖROĞLU[2], Beyza KIRIKTAŞ[3]

[1,2,3]Burdur Mehmet Akif Ersoy Üniversitesi, Gölhisar Uygulamalı Bilimler Yüksekokulu, Bilişim Sistemleri ve Teknolojileri Bölümü, 15400, Burdur, Türkiye

(Alınış / Received: 30.10.2024, Kabul / Accepted: 12.03.2025, Online Yayınlanma / Published Online: 25.04.2025)

Anahtar Kelimeler
Görüntü İşleme,
XGBoost,
Derin Öğrenme,
XAI Lime
SHAP

Öz: Tarladan sofraya uzanan gıdaların kayıpları her bakımdan değerlendirildiği takdirde hem ülke ekonomisine hem de insan sağlığına büyük bir tehdit olarak ortaya çıkmaktadır. İşte bu yüzden kayıpları en aza indirmek için yapay zeka tabanlı görüntü işleme teknikleri bağlamında bu çalışma taze ve çürümüş meyvelerin görüntüleri üzerinde erken teşhis sınıflandırmasına odaklanmaktadır. Görüntü işleme alanında yüksek doğruluk oranlarına ulaşmak, çoğu zaman tek bir modelin sınırlarını aşar. Bu nedenle XGBoost ve derin öğrenme modellerinin sentezlenip birlikte kullanımı daha doğru ve güvenilir sonuçlar elde etmek için yaygın bir yaklaşımdır. Bu çalışma bağlamında VGG16 transfer öğrenme algoritmasını kullanarak özellik çıkarımı yapılmıştır. Sonrasında ise çıkarılan özellikler düzleştirilerek XGBoost algoritmasıyla sınıflandırma işlemi gerçekleştirilmiştir. Bu çalışmada iki adet farklı veri seti kullanılmış olup, ilk veri seti (Dataset1) ham veriler ile algoritma üzerinde çalışma yapılmıştır. İkinci veri setinde ise (Dataset2) ver ön işleme teknikleriyle birlikte XGBoost modeli kullanılmıştır. Dataset1 F1 Skor değeri %98.756 oranına sahipken, Dataset2 %93,36 oranı elde edilmiştir. Bu modelin açıklanmasında SHAP ve LIME açıklanabilir modeller kullanılmıştır.

Detection of Rotten Fruits Using XGBoost-Based Deep Learning Algorithm with Explainable Artificial Intelligence Models

Keywords
Image Processing,
XGBoost,
Deep Learning,
XAI Lime
SHAP

Abstract: Losses of food from the field to the table are a great threat to both the national economy and human health if evaluated in all respects. Therefore, in the context of artificial intelligence based image processing techniques to minimise losses, this study focuses on early detection classification on images of fresh and rotten fruits. Achieving high accuracy rates in image processing is often beyond the limits of a single model. Therefore, synthesising and combining XGBoost and deep learning models is a common approach to achieve more accurate and reliable results. In the context of this study, feature extraction was performed using the VGG16 transfer learning algorithm. Afterwards, the extracted features were smoothed and classification was performed with the XGBoost algorithm. In this study, two different datasets were used, and the first dataset (Dataset1) was used for the algorithm with raw data. In the second dataset (Dataset2), the XGBoost model was used with data preprocessing techniques. Dataset1 has an F1 Score value of 98.756%, while Dataset2 has a rate of 93.36%. SHAP and LIME explainable models were used to explain this model

1. Introduction

Food waste has become an obstacle to preventing malnutrition, which has become a major problem worldwide. According to 2016 data, 1/3 of the agricultural products produced in the world are lost or wasted. According to the same report, the total value of post-harvest losses and food waste in developed countries has reached 680 billion dollars. This loss reaches 310 billion dollars in developing countries. In

* Corresponding author: nilgunsengoz@mehmetakif.edu.tr

other words, approximately 1 trillion dollars' worth of food is wasted every year [1]. That is why it is important to combine the great economic loss in which the food industry is involved with developing computer vision systems and to offer appropriate solutions to this problem.

Fruit rot refers to the process by which fruit loses its freshness and begins to spoil. Spoiled fruit usually shows symptoms such as discolouration, softening, bad odour and taste deterioration. This process occurs naturally through the release of ethylene gas and the action of microorganisms. For example, fruits such as apples and bananas produce ethylene gas, causing other fruits to rot faster. Therefore, fruits that produce ethylene gas should be stored separately. Rotten fruits that are overlooked by human error can cause great damage to producers, so such damage can be prevented by helping image processing and classification artificial intelligence in the sorting of fruits.

In the paper there are 2 different kinds of dataset that were used. Dataset1 has 6 classes with fresh and rotten fruits which are apple, banana, orange. Dataset1 used for the classification of fruits, worked with 13,599 images of apples, oranges and bananas. Of these images, 10,901 were used for training and 2,698 were used for testing.

The other dataset is less than the Dataset1 which also has 6 different classes, strawberry, grapes, pomegranate, as fresh and rotten. Dataset2 has for each class 200 images to be used for the classification. But according to Dataset1 image rates, preprocessing techniques have been applied to Dataset2 to get better accuracy rates.

The main image processing and classification models used in the classification of rotten fruits usually perform up to a certain accuracy. For example, deep learning models (such as convolutional neural network) and classical machine learning algorithms are widely used to extract and classify meaningful information from visual data. However, each model has its own limitations, so it can be difficult to get ahead of a certain performance.

In recent years, hybridization of models, i.e. combining features that two or more models are good at, has emerged to achieve better performance. Combining powerful data processing algorithms such as XGBoost with visual data processing models has yielded remarkable results. In this paper, limitations of classical and deep learning models for visual data processing, the advantages of hybrid models, and how to improve visual data preprocessing performance has been discussed. The potential of hybrid models to achieve higher accuracy and performance has been proved not only metrics but also using with Explainable Artificial Intelligence (XAI) models.

2. Material and Method

Image processing models are algorithms used to extract information from digital images, analyse images and perform specific tasks. These models are usually developed using deep learning and classical machine learning. Basically, image processing stages are pre-processing, feature extraction, classification.

Dataset2 was enlarged with data augmentation methods to improve the performance metrics while classifying and because Dataset2 is smaller in size compared to Dataset1. In addition to the classical methods used, a method based on widely preferred equations such as Gaussian blur was also used. The following methods were used for data augmentation; horizontal and vertical rotation, right and left rotation, brightness and contrast adjustment, Gaussian blur, noise addition, gamma change and color change. Horizontal and vertical rotation was applied by rotating the image 180 degrees on the x and y axes, with a maximum rotation of 15 degrees to the right and left. Brightness and contrast settings were applied between -0.2 and +0.2. Gaussian blur is randomly generated between 3x3 and 7x7 kernel sizes according to the limits of kernel sizes.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad [1]$$

Applied in two dimensions, this formula (1) produces a surface whose contours are concentric circles with a Gaussian distribution from the center point. The gamma of the image is multiplied by a value between 0.8 and 1.2. As for the color change on the image, the hue is changed between -20 and +20, while the saturation is changed between -30 and +30. The application of data augmentation methods on images by setting limits is to prevent the augmented data from resembling each other as much as possible. Before the data augmentation, the raw data set contained 200 data for each class, but after the data augmentation, this number increased to 800. The total number of data sets was increased 4 times, and 4800 data were obtained.

Feature extraction is the process of identifying meaningful and discriminative features of images. Deep learning models are very effective at this stage. In the project, feature extraction is performed using the VGG16 model after dimensioning and normalization of the images in the preprocessing of the data.

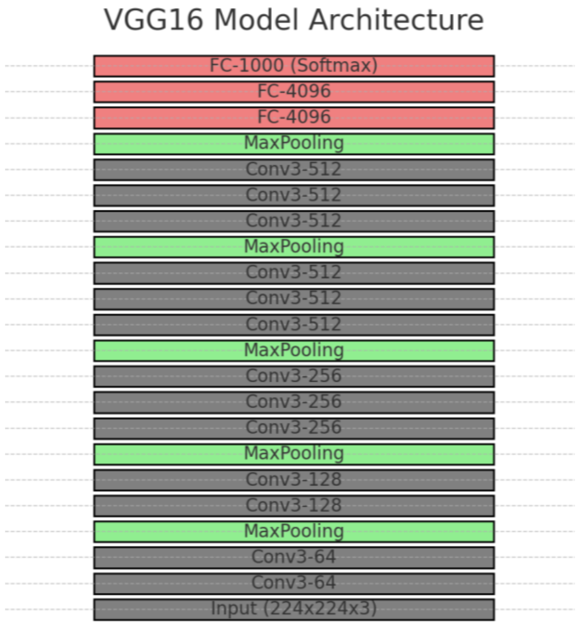

Figure 1. VGG16 Model Layers

VGG16 is a Convolutional Neural Network (CNN) model consisting of 16 layers in total. These 16 layers consist of 13 convolutional layers and 3 full connection layers. The basic architecture of the model is based on creating a deep and narrow network structure using convolution filters.

The input layer of this model takes a 224x244 image with 3 color channels (RGB). The convolution layers use 3x3 filters and after each convolution, the ReLU (Rectified Linear Unit) activation function is applied. The convolution layers are respectively 2 64-channel layers, 2 128-channel layers, 3 256-channel layers, 3 512-channel layers.

After each convolution block there is a 2x2 Max Pooling layer which is half the size. These layers reduce the number of parameters the model must learn, preventing overfitting and reducing computational cost. These blocks are followed by 3 fully connected layers. This last layer is a layer with a softmax activation function that separates the input image into 1000 classes. This layer gives the probability distribution of each class.

Deep learning models usually output features extracted from images as multidimensional tensors. These tensors must be flattened before they can be processed by traditional machine learning algorithms such as XGBoost. [2] Smoothing is the process of converting multidimensional tensors into one-dimensional vectors. This is done by summing all the values in each tensor into a single line.

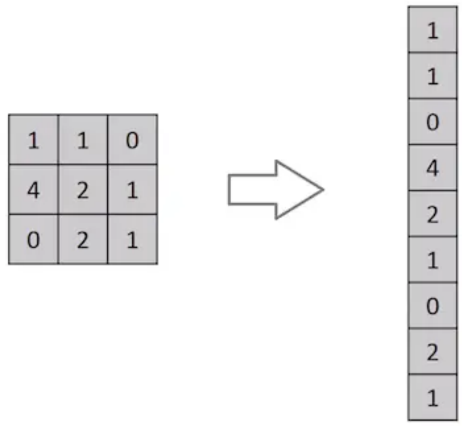

Figure 2. Example of flattening Tensors

2.1. XGboost classifier

XGBoost (Extreme Grandient Boosting) is a machine learning algorithm based on the gradient boosting framework that performs high performance classification and regression tasks on large and complex data sets. XGBoost improves prediction accuracy by using an ensemble of tree-based models and stands out with its speed, flexibility and scalability. At each iteration, the model adds new decision trees that minimize errors and strengthen weak predictors.

XGBoost starts by making initial forecasts for each observation. These forecasts are usually initialized with an average value. Then, the errors of the current forecasts are calculated. These errors represent the current performance of the model and are the points to be corrected in the next step. A new decision tree is trained to estimate these errors. Each new tree is designed to correct the errors of the previous model. The predictions of the new tree are added to the existing model, which increases the accuracy of the model. The process is repeated, giving more weight to the errors. Each iteration improves the accuracy of the model. Finally, the results of all trees are combined to make the final predictions. This iterative process gradually improves the performance of the model, increasing the accuracy of the final predictions.

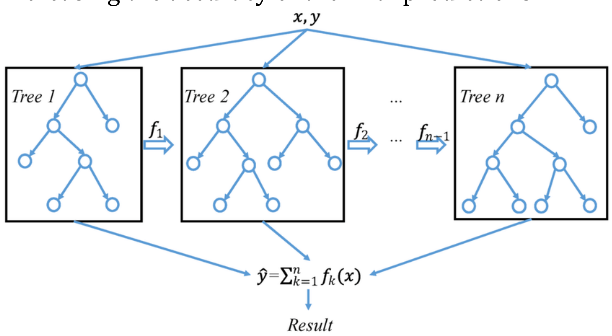

Figure 3. XGBoost Decision Trees

The smoothed features are used as input in the XGBoost model. This creates a dataset containing the smoothed features and the target variable (label). The XGBoost classifier is trained on this data set. During training, the model tries to minimize errors by creating different decision trees. When the training is completed, the model makes predictions on test and validation data and evaluates the performance.

2.2. Data set

In this paper, performance of the proposed XGBoost-based classification model was evaluated using two distinct datasets with differing characteristics. Dataset1 comprised a large collection of 13,599 raw images, which were directly utilized without any data augmentation. Dataset2, on the other hand, was initially smaller, containing 200 images per class (a total of 1,200 images across six classes). To address the limited size of Dataset2, an extensive data augmentation process was applied, resulting in 800 images per class, increasing the total number of images to 4,800.

The Dataset1 is named 'Fruits fresh and rotten for classification' [3] obtained on Kaggle website and Dataset2 is named "Fresh and Rotten Fruits Dataset for Machine-Based Evaluation of Fruit Quality" [4] also obtained on Mendeley data website was used in the study. In Dataset1 10.901 of these photographs were used as training (approximately 80%) and 2698 of these photographs were used as test (approximately 20%), in total 13.599 photographs were used.

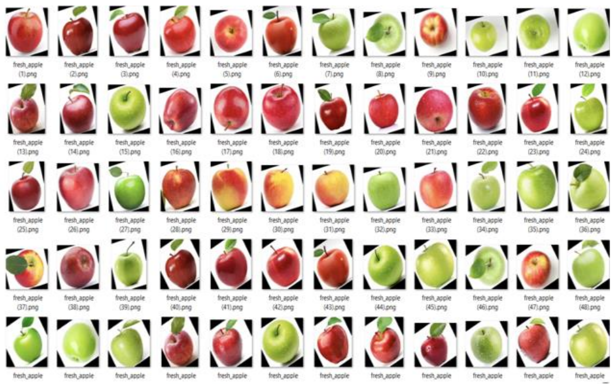

Figure 4. Fresh Apple Data Set Example

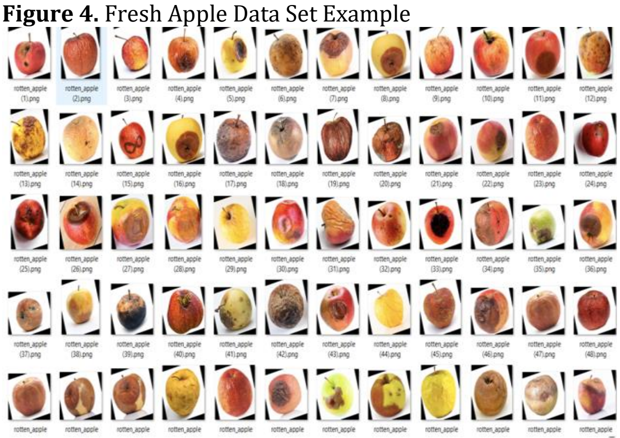

Figure 5. Rotten Apple Data Set Example

In Dataset2 preprocessing techniques have been applied to get better accuracy rate when we compare Dataset1 images size. Above preprocessing techniques are clarified and explained so in the beginning Dataset2 for each class 200 images after preprocessing techniques, each class have been up to 800 images. So, at the beginning Dataset2 raw images are 1200, then 4800 images are obtained. Dataset2 were traced %80-%20 rule.

3. Findings

The accuracy metrics of the model results are calculated on the training and test data sets. First, the data set is split into training and test sets; the training set is used for learning the model, while the test data set is used to evaluate the performance of the model. After the model is trained on the training data sets, it makes predictions on the test data sets. These predictions are compared with the actual labels and the accuracy rate is calculated. The accuracy rate is determined as the ratio of correct predictions to total predictions. The difference in performance between the training and test sets is a critical indicator to assess whether the model is overfitting or underfitting.

3.1 Confusion matrix

In accuracy metric calculations, the confusion matrix is a table used to evaluate the performance of classification models and shows in detail the correct and incorrect classifications of the model. It consists of four cells: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

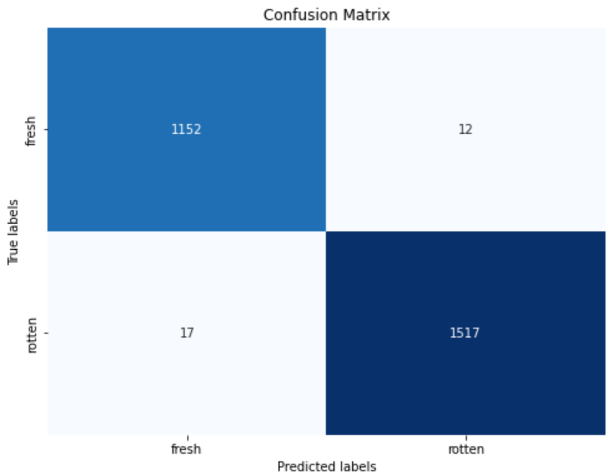

Figure 6. Confusion Matrix for Dataset1

While classifying the Dataset1, the model correctly predicted 1152 fruits labelled as fresh, and 1517 fruits labelled as rotten. It incorrectly predicted 12 fruits labelled as rotten as fresh and 17 fruits labelled as fresh as rotten.

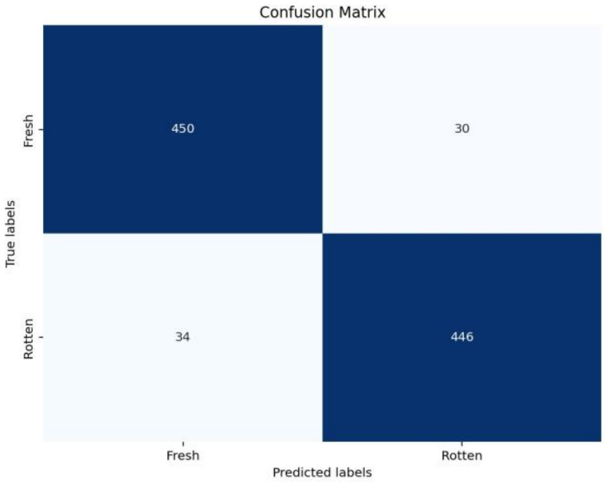

Figure 7. Confusion Matrix for Dataset2

At the Dataset2 confusion matrix figure, the model prediction for the correct classes is 450 and 446, fresh and rotten respectively. Misclassified rates are for the fresh and rotten classes are 30 and 34 respectively.

True Positive model correctly classifies as positive and True Negative model correctly classifies as negative. False Positive is what the model incorrectly classifies as positive, False Negative is what the model incorrectly classifies as negative.

Accuracy, which is formula 2, measures how many of the model's classifications are correct and is calculated as follows:

$$\frac{TP + TN}{TP + TN + FP + FN} \quad [2]$$

Precision is a metric that measures how many of a model's positive predictions are correct. In other words, it shows the ratio of the model's 'true positive' predictions to its total positive predictions. Precision is a critical metric, especially when the number of false positives is important. Precision, which is formula 3 is calculated as follows:

$$\frac{TP}{TP + FP} \quad [3]$$

Recall is a performance indicator that measures how many of the samples that are actually positive the model correctly identifies as positive. Recall is particularly critical in situations where false negatives are costly (e.g. disease diagnosis or safety warning systems). This metric, expressed mathematically by Formula 4, is calculated as follows:

$$\frac{TP}{TP + FN} \quad [4]$$

The F1 score is the harmonic mean of the precision and recall metrics and is a measure of the balance between these two indicators. It is an ideal tool for evaluating model performance, especially in scenarios that require a balance between precision and recall (e.g., in datasets with high class imbalance). The F1 score, expressed by formula 5, is calculated as follows:

$$2\,x\,\frac{Precision * Recall}{Precision + Recall} \quad [5]$$

Matthews Correlation Coefficient (MCC) is a metric that evaluates the performance of classification models at the level of all classes (positive and negative). By measuring the correlation between actual and predicted classes, this metric provides an objective performance analysis even in imbalanced datasets. MCC, denoted by Formula 6, is calculated using the following formula:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad [6]$$

The MCC value is a metric that takes values between -1 and +1. A value close to +1 indicates that the model produces an almost perfect classification, while values around 0 indicate random predictions. A value close to -1 indicates that the model produces inaccurate results in the opposite direction of the actual classification.

The MCC value calculated in this study was found to be 0.978109. The fact that this value is extremely close to +1 proves that the model exhibits high consistency and accuracy in fresh and rotten fruit classification.

Table 1. Accuracy Metrics

Performance Metrics	Dataset1	Datasest2
Accuracy	0.98813	0.93333
Precision	0.98969	0.9375
Recall	0.98545	0.929752
F1 Score	0.98756	0.933609
MCC	0.978109	0.866696

The performance metrics presented in Table 1 show that the model performs significantly better on Dataset1 than Dataset2 in all evaluation criteria. The 98.81% accuracy rate obtained on Dataset1 outperforms Dataset2's 93.33%, indicating that the overall classification performance of the model is more consistent. However, since accuracy metric alone can be misleading in case of class imbalance, metrics such as precision, recall, F1 score and Matthews Correlation Coefficient (MCC) provide a more comprehensive analysis. The precision value in Dataset1 (0.98969) is significantly higher than Dataset2 (0.9375), proving that the model is less likely to produce false positives in the first dataset. Similarly, the sensitivity value in Dataset1 (0.98545) was higher than Dataset2 (0.929752), indicating that the model is more capable of recognising true positive examples. This metric is especially important in applications where false negatives can have critical consequences. The F1 score, which is the harmonic mean of sensitivity and precision, also confirms a more balanced performance in Dataset1 (0.98756) compared to Dataset2 (0.933609). In addition, the MCC value calculated for Dataset1 (0.978109) is well above Dataset2 (0.866696), indicating that the model maintains a strong correlation between actual classes and predictions. This high value of MCC, which is of particular importance for performance evaluation in unbalanced datasets, supports that the model achieved a more stable classification balance in Dataset1.

The increasing accuracy of machine learning models makes it difficult to understand their decision processes. High accuracy often requires the use of 'black box' models with complex architectures, which negatively affects model explainability. However, transparency of model decisions, especially in fields such as medicine or quality control, is critical for reliability and acceptability. This paradox constitutes one of the main obstacles to the adoption of AI in practical applications.

3.3 Explainable artificial intelligence (XAI) models

In this study, the integration of SHAP and LIME methods offers significant advantages for an in-depth analysis of the interpretability of the decision mechanisms of the XGBoost model in fresh and rotten fruit classification. While SHAP reveals the impact of dataset-wide features on the model from a holistic perspective, LIME complements this analysis by detailing localised decision dynamics in specific instances. This two-tier approach illuminates both the overall behaviour of the model and the complex relationships in individual predictions, taking interpretability to a comprehensive level. In particular, whereas SHAP emphasises the key features that determine classification performance, LIME describes rare or ambiguous feature interactions on an example-by-example basis. The combined use of both methodologies provides a robust strategy that supports the transparency and reliability of the model, while compensating for shortcomings that may arise when they are applied alone.

The combination of LIME's sample-oriented analyses and SHAP's descriptions of general data patterns provides a multidimensional interpretation of the decision processes of the XGBoost-based classifier. This integrated framework not only increases the transparency of the model but also reveals critical dependencies and interactions between features. Thus, it allows systematic assessment of both global trends (e.g. humidity or colour variations) and anomalies or uncertainties in individual cases. As a result, this work contributes to interpretability standards for the reliable use of AI models in areas such as agriculture and quality control.

3.3.1 LIME model

If we evaluate it in the context of Interpretable Artificial Intelligence (XAI) and its methodological contributions, the XAI methodology provides transparency by structuring the internal operational mechanics and decision-making dynamics of complex models in a way that is suitable for human perception. This interpretability is critical for increasing model reliability, detecting algorithmic biases and ensuring ethical compliance in high-risk areas such as medicine, finance and law. The added value of XAI is not only reflected in statistical accuracy metrics, but also in the auditability of systems and user adoption.

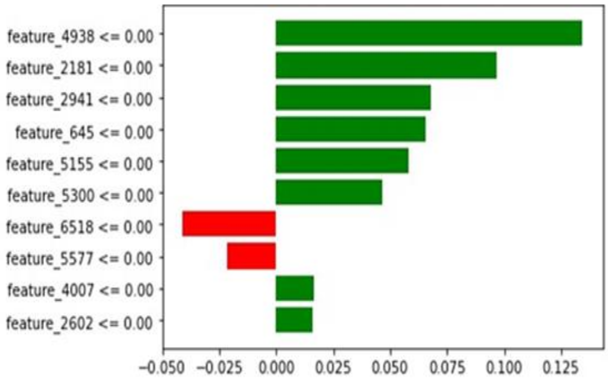

Figure 8. LIME for Dataset1

Localizable Model Analysis with LIME is one of the important parts of the study and in this context, the visual presented in Figure 8 represents an XAI output produced by the 'Local Interpretable Model-Independent Explanations' (LIME) technique. This method is used to decipher the classification decisions of opaque algorithmic systems, especially those considered as black boxes, at the level of individual data points. The main function of LIME is to map the local decision boundaries of the model with a nonparametric approach and visualize the feature impact distributions.

In the presented graph, in the analysis made for a sample belonging to the 'Rotten' class, it is seen that the green-toned bars represent the features that support the classification decision, while the red ones represent the inhibitory factors. For example, while the parameters with indexes feature_4938, feature_2181 and feature_2941 significantly affect the classification threshold positively (weight values +0.22, +0.18, +0.15, respectively), feature_6518 (-0.21) and feature_5577 (-0.17) made negative contributions.

SHAP-LIME Synergy and Feature Impact Mapping is also one of the Explainable Artificial Intelligence models and since it is important in the originality part of the study, the decision architecture of the XGBoost model used was examined with a multi-layered perspective using the dual analytical frameworks of SHAP (Shapley Values) and LIME. The local explanation in Figure 9 reveals the decisive positive effect of feature_4741 (+0.31) and feature_3671

(+0.26) in a prediction belonging to the 'Fresh' class, while the parameters such as feature_3350 (-0.29) and feature_5393 (-0.24) play an inhibitory role in the separation process between classes.

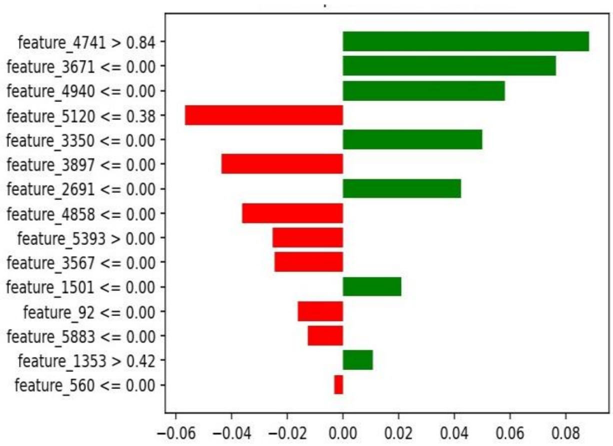

Figure 9. LIME for Dataset2

This hybrid analysis approach captures the nuances of feature interactions at the local level that can be overlooked at the global scale, revealing the causal relationships of model behaviours. Especially in data-limited conditions, these granular insights at the feature level are indispensable for the validation of algorithmic decision-making processes.

Using these graphs, we can see which features contributed to or hindered the model's classification as 'rotten' and by how much. This is extremely important to understand why the model made such a decision and is used to assess the reliability of the model.
While assessing the accuracy and reliability of the model's decisions, it also allows us to understand the reasons for erroneous predictions. By increasing the transparency of the model, LIME helps users better understand the model's decisions and more accurately interpret the results of those decisions.

Rather than understanding the overall structure of the LIME model, the local model is built to understand how a particular data point is classified by the model. Firstly, it creates data points around a given data point, i.e. by slightly modifying this point. The predictions of the model are taken on these new data points. These prediction results are associated with a simpler and more understandable model that locally reflects the decision process around this data point. As a result, it shows which features are influential in this prediction and how much. These local descriptions of the features are used to understand how the model arrived at a particular prediction.

3.3.2 SHAP model

The SHAP (SHapley Additive exPlanation) algorithm is a powerful technique for improving model explainability. By calculating the contribution of each feature to the model output, SHAP explains how the model works and what decisions it influences. However, this may not be sufficient in complex and hybridized models.

This graph is used to assess which features the model gives more 'importance' to and the effects of these features on the model's predictions.

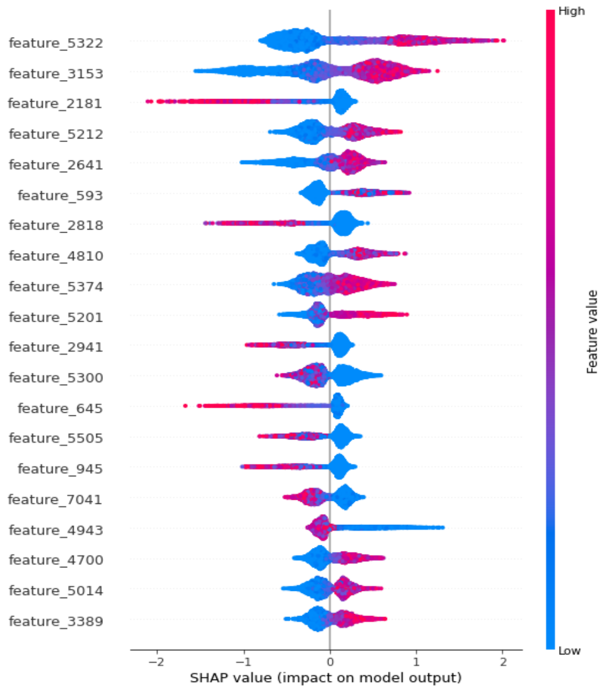

Figure 10. SHAP for Dataset1

Image Above (Figure 10); Y-Axis (Features); the features used in the model are listed on the Y-axis. Each row represents one feature.
X-Axis (SHAP Values); the X-axis shows the SHAP values. The more positive the SHAP value, the more this feature increases the prediction of the model. If it is negative, it lowers the prediction.
Color Scale: The colors indicate the magnitude of the feature values. Pink/red tones represent high values, and blue tones represent low values. For example, in the feature 'feature_5322', high SHAP values usually correspond to high (pink) feature values.

Feature 5322 has a significant influence on the model's decision-making and feature 3389 has a lesser influence on the model's decision-making, but it is difficult to say where these features influence the data.

When analysing the chart structure, for each feature, the distribution of SHAP values is shown in a 'violin plot'. This shows the distribution and density of SHAP values. The thicker parts of the graph indicate that SHAP value is more frequent.

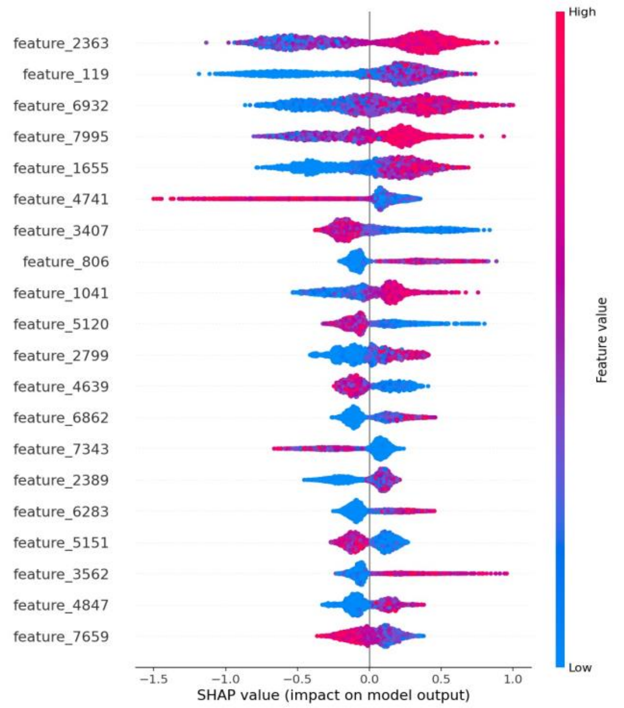

Figure 11. SHAP for Dataset2

In contrast, the SHAP summary plot (Figure 11) provides a comprehensive global perspective by ranking features based on their overall impact on the model's predictions across the dataset. The analysis reveals that feature_4741, feature_2363, and feature_6932 are the most influential predictors, as evidenced by their broad range of SHAP values. The color gradient within the plot (blue representing low feature values and red representing high feature values) further illustrates how feature magnitudes affect the classification outcomes. For instance, higher values of feature_4741 are strongly correlated with predictions for the 'Fresh' class, while lower values detract from it. Additionally, features like feature_6932 exhibit significant variability in their SHAP value distributions, indicating diverse impacts across individual instances.

4. Discussion and Conclusion

In this study, it is aimed to develop a model that classifies fruits as fresh or rotten using artificial intelligence. Thus, rotting of fruits will be prevented and separated more easily, both economic losses will be reduced, and human health will be protected. The model to be developed is an image processing model and it is aimed to outperform the existing approaches. Being different from other studies using the same dataset, it is thought to make a unique contribution to the literature. Instead of traditional image processing methods, it is aimed to create a hybrid model by using the XGBoost algorithm, which is known to be effective in different fields, and thus increase performance.

Visual datasets were prepared within a methodological framework prior to model development. Images were rescaled to standard dimensions and subjected to semantic labeling for classification efficiency. In the preprocessing phase, multidimensional features of images were converted to a single vector form through dimensional reduction and made suitable for XGBoost architecture. The obtained dataset was trained with this gradient boosting-based model and the classification task was performed.

In the context of model evaluation and comparative analysis, a comparison was made with the accuracy metric of previous studies conducted on the same dataset to measure the effectiveness of the model. This comparative evaluation reveals the performance superiority of the methodology over alternative approaches.

When the results of Dataset 1 are examined in detail, it is seen that; the positive effect of using large-scale raw data on model performance was quantified with 98.97% precision, 98.55% recall, 98.76% F1-score and 0.9781 Matthews Correlation Coefficient (MCC). These results prove the contribution of comprehensive feature representation provided by heterogeneous data structure to algorithmic learning.

Focusing on Dataset 2 and augmented data analysis; stochastic data augmentation techniques were applied to increase class diversity in Dataset2 with limited initial data. With augmented data, the model achieved 93.75% accuracy, 92.97% sensitivity, 93.36% F1-score and 0.8667 MCC value. Despite the difference in initial data volume, the generalization ability of the model showed that it remained effective even in limited data scenarios.

Since the interpretation of Confusion Matrix is one of the most important criteria in performance evaluations, the confusion matrix of Dataset2 in Figure 7 reveals that 450 out of 480 test samples were correctly classified as 'Fresh' and 446 as 'Rotten', with 30 and 34 incorrect predictions, respectively. These findings confirm that the methodology exhibits stable performance even in augmented datasets.

Table 2. Accuracy Ratios of other models in the literature using for the Dataset1 also evaluated metric for Dataset2

Model(s)	Optimizer	Accuracy Ratio	Reference
Seguential 8	Adam	98.21%	[5]
VGG-16	Adam	(0.6) 90.41%	[6]
VGG-19		(0.6) 98.36%	
InceptionV3	-	96.33%	[7]
InceptionV3	-	97.34%	[8]
Xception		97.16%	
VGG-16		96.47%	
MobilNet		95.47%	
NASNetMobile		75.29%	
MobilNetV2	SGDM	88.62%	[9]
ResNet50		73.26%	

VGG-16		96.10%	
InceptionV3		97.10%	
MobileNetV2	Adam	87.10%	[10]
MobilNetV2	Adam	87%	[11]
InceptionV3	Adam	85%	[11]
VGG-16	Adam	89.42%	[12]
VGG-19		76.18%	
MobilNet		68.72%	
Xception		78.68%	
CNN ReLu	Adam	99.04%	[13]
CNN Sigmoid	Adam	94.18%	
XGBoost Model for Augmented Dataset2 (evaluated model)	-	**93.333%**	
XGBoost Model for Dataset1 (proposed model)	**-**	**98.813%**	

In terms of the analysis of the relationship between data scale and algorithmic flexibility, the analysis attributes the superior metrics of Dataset1 to the large data volume, while Dataset2 proves that methodological adjustments can balance the data constraints. The fact that XGBoost produces consistent results in both scenarios highlights the flexible nature of the algorithm.

The literature comparison and deep learning algorithm comparison are already the most important part of the whole study, and the originality of the article stands out in this part, and it is observed in Table 2 that the proposed XGBoost model on Dataset1 outperforms deep learning architectures such as InceptionV3 (97.34%), Xception (97.16%) [8] and VGG-16 (96.10%) [9] with an accuracy rate of 98.813%. However, the higher performance of models such as CNN ReLU (99.04%) [13] and VGG-19 (98.36%) [6] reveals the critical role of hyperparameter optimization in deep learning.

Considering the proposed optimization strategies, the findings indicate that despite the competitive performance of XGBoost, it can be improved with hybrid models. In particular, the integration of pre-trained layers based on transfer learning can expand the model capacity without increasing the computational cost. In addition, group-based learning strategies or optimizing existing models with fine-tuning techniques can be effective in reducing the performance gap between traditional machine learning and deep neural networks.

Such studies encourage the adoption of more innovative approaches in machine learning beyond the development of existing methods. Hybrid models offer an important opportunity to overcome the limits of classical methods. Combining deep learning methods with powerful algorithms such as XGBoost can increase accuracy and generalisation capacity.

In addition, traditional approaches such as transfer learning can be adapted to different datasets and application areas to provide innovative solutions. The combination of different models makes it possible to develop more flexible and powerful solutions.

In conclusion, the proposed hybrid model proves to be an effective solution by providing higher accuracy metrics compared to other studies performed on the same dataset. This success is achieved by combining the strengths of both deep learning models and traditional machine learning algorithms. Since similar advantages can be achieved on different datasets and application domains, the use of hybrid models is considered an important strategy in the field of machine learning and data science.

Declaration of Ethical Code

In this study, we undertake that all the rules required to be followed within the scope of the 'Directive on Scientific Research and Publication Ethics of Higher Education Institutions have been complied with, and that none of the actions specified under the heading 'Actions Contrary to Scientific Research and Publication Ethics' of the said directive have been carried out.

References

[1] Food Losses and Waste (FAO) Inventory and management at each stage in the food chain, 2016. https://openknowledge.fao.org/server/api/core/bitstreams/36cb45bc-392c-41fb-97f1-90ca1f16ee7f/content (Access date: 25.12.2024)

[2] Bhattiprolu, S. 2021. XGBoost for image classification using VGG16. https://github.com/bnsreenu/python_for_microscopists/blob/master/195_xgboost_for_image_classification_using_VGG16.py (Access date: 26.03.2024)

[3] Kaggle, 2018, Fruits Fresh and Rotten for Classification, https://www.kaggle.com/datasets/sriramr/fruits-fresh-and-rotten-for-classification (Access date: 26.02.2024)

[4] Mendeley Data, 2022, Fresh and Rotten Fruits Dataset for Machine-Based Evaluation of Fruit Quality, https://data.mendeley.com/datasets/bdd69gyhv8/1 (Access date: 25.12.2024)

[5] Mehedi, H., Moinul, H. 2021. Fresh and Rotten Fruit Classification Using Deep Learning. Daffodil International University. Computer Science and Engineering.

[6] Ayşin, B. 2022. Determining the Freshness of Fruits with Deep Learning Methods. Tekirdağ Namık Kemal Üniversitesi. Bilgisayar Mühendisliği Anabilim Dalı. Yüksek Lisans Tezi. Tekirdağ.

[7] Hithesh Kumar, C.M., Vikash, VArun, Dyamanagouda, P. 2021. Detection of Quality of Fruits Usıng AI. International Research Journal of Modernization in Engineering Technology and Science, 3(7).

[8] Sohel, M., Tayeeba, T., Mirajul., I., Mumenunnesa, K., Riazue, R., Syed, A.H. 2021. An Advanced Method of Identification Fresh and Rotten Fruits using Different Convolutional Neural Networks. International Conference on Computing Communication and Networking Technologies (ICCCNT) 12th

[9] Amin, U.; Shahzad, M.I.; Shahzad, A.; Shahzad, M.; Khan, U.; Mahmood, Z. Automatic Fruits Freshness Classification Using CNN and Transfer Learning. Appl. Sci. 2023, 13, 8087.

[10] Tapia-Mendez, E.; Cruz-Albarran, I.A.; Tovar-Arriaga, S.; Morales-Hernandez, L.A. Deep Learning-Based Method for Classification and Ripeness Assessment of Fruits and Vegetables. Appl. Sci. 2023, 13, 12504

[11] Deyner Julian, N. O., Silvia Alejandra, M. L. (2022). Automatic Identification of banana quality with Deep Neural Network Classification. ISSN online 2344-9217

[12] Sai Sudhan, S.P., Venkata Rami, R.C., Yakobu, D., Suneetha B. 2020. Fresh and Rotten Fruits Classification Using CNN and Transfer Learning. Revue d'Intelligence Artificielle 34(5), pp. 617-622

[13] Florence, S., Nur Shabira, B. 2023. Hyperparameter Tuning of Convolutional Neural Network for Fresh and Rotten Fruit Recognition. 2023 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)