Gazi Üniversitesi
Fen Bilimleri Dergisi
PART C: TASARIM VE TEKNOLOJİ

Gazi University
Journal of Science
PART C: DESIGN AND TECHNOLOGY

GU J Sci, Part C, 13(X): XX-XX (2025)

Classification of Hazelnut Varieties Based on Shell Characteristics: An Image Processing and Machine Learning Approach

Rabia KAYMAK[1], Ferzan KATIRCIOĞLU[2*]

[1,2] *Duzce University Faculty of Engineering, Department of Mechatronics Engineering, Duzce, Turkey*

Article Info

Research article
Received: 17/03/2025
Revision: 18/05/2025
Accepted:14/06/2025

Keywords

Agricultural Technology
Image Processing Machine
Learning Hazelnut
Classification
Automation

Makale Bilgisi

Araştırma makalesi
Başvuru: 17/03/2025
Düzeltme: 18/05/2025
Kabul: 14/06/2025

Anahtar Kelimeler

Tarım Teknolojisi
Görüntü İşleme
Makine Öğrenmesi
Fındık Sınıflandırması
Otomasyon

Graphical/Tabular Abstract (Grafik Özet)

This study developed an original AI-based system for the automatic classification of hazelnut varieties based on shell characteristics. The system is suitable for integration into industrial production lines and aims to improve quality control and processing efficiency. / Bu çalışma, fındık çeşitlerinin kabuk özelliklerine göre otomatik sınıflandırılmasını sağlayan, görüntü işleme ve yapay zekâ tabanlı özgün bir sistem geliştirmiştir. Sistem, endüstriyel üretim hatlarına entegre edilebilir nitelikte olup, kalite kontrol ve işleme verimliliğini artırmaya yöneliktir.

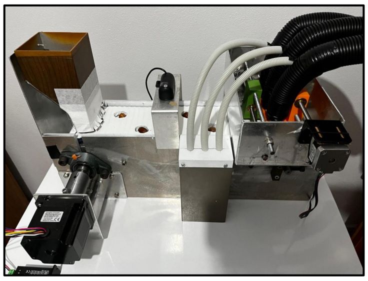


Figure A: *Detailed production image of the proposed project /* ***Şekil A:*** *Önerilen projenin detaylı üretim görseli*

Highlights (Önemli noktalar)

- *Classification of hazelnut varieties was performed based on shell characteristics. / Fındık çeşitlerinin kabuk özellikleri üzerinden sınıflandırma işlemi yapılmıştır.*
- *This study presents the first systematic design for classifying hazelnut varieties based on shell characteristics. / Bu çalışma, fındık çeşitlerinin kabuk özelliklerine göre sınıflandırılmasına yönelik ilk sistematik tasarımı sunmaktadır.*
- *Image processing techniques were used to extract color, texture, and dimensional features. / Görüntü işleme teknikleri kullanılarak renk, doku ve boyut özellikleri çıkarılmıştır.*

Aim (Amaç): *The aim of this study is to design and develop a machine learning-based intelligent system that can automatically classify hazelnut varieties based on their outer shell characteristics such as color, texture, and size using image processing techniques. / Bu çalışmanın amacı, fındık çeşitlerini dış kabuk özellikleri olan renk, doku ve boyut temelinde görüntü işleme teknikleri kullanarak otomatik olarak sınıflandırabilen makine öğrenmesi tabanlı akıllı bir sistem tasarlamak ve geliştirmektir.*

Originality (Özgünlük): *The study is the first to systematically classify hazelnut varieties based on shell characteristics using machine learning. It offers an original prototype suitable for industrial applications. / Bu çalışma, fındık çeşitlerini kabuk özelliklerine göre makine öğrenmesi ile sistematik şekilde sınıflandıran ilk çalışmadır. Endüstriyel uygulamalara uygun özgün bir prototip sunmaktadır.*

Results (Bulgular): *Among all tested algorithms, CFNN achieved the highest accuracy with 92%. Feature extraction based on shell color, size, and texture significantly improved classification performance. / Test edilen tüm algoritmalar arasında CFNN %92 doğruluk ile en yüksek başarıyı göstermiştir. Kabuk rengi, boyutu ve dokusuna dayalı öznitelik çıkarımı sınıflandırma başarısını artırmıştır.*

Conclusion (Sonuç): *The developed machine learning-based system successfully classifies hazelnuts by variety and is ready to be used in industrial processes to increase efficiency and product quality. / Geliştirilen makine öğrenmesi tabanlı sistem, fındıkları çeşidine göre başarıyla sınıflandırmakta ve verimlilik ile ürün kalitesini artırmak için endüstriyel süreçlerde kullanılmaya hazırdır.*

*Corresponding author, e-mail: ferzankatircioglu@duzce.edu.tr
DOI: 10.29109/gujsc.1659305

GU J Sci, Part C, 13(X): XX-XX (2025)

Gazi Üniversitesi
Fen Bilimleri Dergisi
PART C: TASARIM VE TEKNOLOJİ

Gazi University
Journal of Science
PART C: DESIGN AND TECHNOLOGY

http://dergipark.gov.tr/gujsc

Classification of Hazelnut Varieties Based on Shell Characteristics: An Image Processing and Machine Learning Approach

Rabia KAYMAK[1], Ferzan KATIRCIOĞLU[2*]

[1,2] *Duzce University Faculty of Engineering, Department of Mechatronics Engineering, Duzce, Turkey*

Article Info

Research article
Received: 17/03/2025
Revision: 18/05/2025
Accepted:14/06/2025

Keywords

Agricultural Technology
Image Processing
Machine Learning
Hazelnut Classification
Automation

Abstract

In the export and import field of hazelnut fruit, there are no machines specifically designed to distinguish hazelnut varieties according to shell characteristics in the production line in the sector; instead, existing machines mainly separate hazelnuts into categories such as full, empty or broken. The ability to distinguish between hazelnut varieties is essential to optimize processing workflows and increase the quality of the final product. In this context, the proposed study aims to classify hazelnuts according to their varieties as Chubby, Pointed and Almond by analyzing their physical properties such as color and texture to determine their suitability for specific applications. In the study conducted; after capturing images of hazelnut shells on the conveyor belt, feature extraction was performed on the processed images after preliminary image processing steps such as cropping, background removal and measurement standardization. The techniques used for feature extraction include Size, Color, Haralick, Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG). These extracted features were then applied to various classifiers. When the classification results are examined, according to the results of Sensitivity, Precision, F-Score and G-Mean metrics evaluated in the study, it is revealed that the Cascaded Forward Neural Network (CFNN) algorithm outperforms other classification algorithms by obtaining 92% accuracy value. As a result, the machine learning-based system developed to classify hazelnuts according to their outer shells is ready to bring a great innovation by taking efficiency to a new level and significantly increase quality in industrial processes.

Fındık Çeşitlerinin Kabuk Özelliklerine Göre Sınıflandırılması: Bir Görüntü İşleme ve Makine Öğrenmesi Yaklaşımı

Makale Bilgisi

Araştırma makalesi
Başvuru: 17/03/2025
Düzeltme: 18/05/2025
Kabul: 14/06/2025

Anahtar Kelimeler

Tarım Teknolojisi
Görüntü İşleme
Makine Öğrenmesi
Fındık Sınıflandırması
Otomasyon

Öz

Fındık meyvesinin ihracat ve ithalat alanında sektörde üretim bandında fındık çeşitlerini kabuk özelliklerine göre ayırt etmek için özel olarak tasarlanmış makineler bulunmamaktadır; bunun yerine mevcut makineler fındıkları ağırlıklı olarak dolu, boş veya kırık gibi kategorilere ayırmaktadır. Fındık çeşitleri arasında ayrım yapabilme yeteneği, işleme iş akışlarını optimize etmek ve nihai ürünün kalitesini artırmak için esastır. Bu bağlamda, önerilen çalışma, belirli uygulamalar için uygunluklarını belirlemek üzere renk ve doku gibi fiziksel özelliklerini analiz ederek fındıkları Tombul, Sivri ve Badem olmak üzere çeşitlerine göre sınıflandırmayı amaçlamaktadır. Gerçekleştirilen çalışmada taşıma bandında fındık kabuklarının görüntüleri yakalandıktan sonra, kırpma, arka plan kaldırma ve ölçüm standardizasyonu gibi ön görüntü işleme adımlarının ardından, işlenmiş görüntüler üzerinde özellik çıkarma işlemi gerçekleştirilmiştir. Özellik çıkarma için kullanılan teknikler arasında Boyut, Renk, Haralick, Yerel İkili Desenler (LBP) ve Yönlendirilmiş Gradyanların Histogramı (HOG) yer almaktadır. Çıkarılan bu özellikler daha sonra çeşitli sınıflandırıcılara uygulanmıştır. Sınıflandırma sonuçlarına bakıldığında çalışmada değerlendirilen Sensitivity, Precesion, F-Score ve G-Mean metriklerinin sonuçlarına göre Basamaklı İleri Sinir Ağı (CFNN) algoritması %92 doğruluk değerini alarak diğer sınıflandırma algoritmalarından daha iyi performans gösterdiğini ortaya koymuştur. Sonuç olarak, fındıkları dış kabuklarına göre sınıflandırmak için geliştirilen makine öğrenme tabanlı sistem, verimliliği yeni bir seviyeye taşıyarak büyük bir yenilik getirirken endüstriyel süreçlerde kaliteyi önemli ölçüde artırmaya hazırdır.

*Corresponding author, e-mail: ferzankatircioglu@duzce.edu.tr
DOI: 10.29109/gujsc.1659305

1. INTRODUCTION (GİRİŞ)

Hazelnuts are one of the most widely produced and consumed food products globally, playing a crucial role in human nutrition. A 100-gram serving of hazelnuts provides 634 calories of energy [1]. The predominant organic acid in hazelnuts is oleic acid [2]. The protein content in hazelnuts ranges from 10% to 24%, with 100 grams supplying 22% of the daily protein requirement for an average adult. The content of cellulosic compounds is approximately 1-3%. Hazelnuts are also rich in essential minerals, including iron (Fe), magnesium (Mg), copper (Cu), manganese (Mn), potassium (K), phosphorus (P), zinc (Zn), and calcium (Ca). Furthermore, hazelnut oil is the second-best source of vitamin E among vegetable oils, with 100 grams of hazelnuts providing 24% of the daily vitamin E requirement [1]. The specific value parameters for different hazelnut varieties are detailed in Table 1.

Table 1. Value parameters of some hazelnut varieties (Fındık çeşitlerinin değer parametreleri) [2].

	Kernel per. (%)	**Protein (%)**	**Oil (%)**	**K (mg/kg)**	**P (mg/kg)**	**Ca (mg/kg)**	**Mg (mg/kg)**	**Mn (mg/kg)**	**Fe (mg/kg)**	**Cu (mg/kg)**	**Zn (mg/kg)**
Tombul	54,6	16,13	62,8	6651	3453	1926	1712	51,6	30,9	33	24,7
Sivri	53,8	17,67	62,7	5875	3307	1830	1637	52,8	32,3	29,1	25,5
Badem	55	13	60	6000	2000	2323	1728	30	25	10	20

Hazelnuts are highly valued in healthy nutrition due to their rich chemical composition. They offer significant cardiovascular benefits, including protection against heart and vascular diseases by helping to prevent elevated cholesterol levels [3]. Globally, hazelnuts are the second most widely cultivated hard-shell fruit, surpassed only by almonds. Major hazelnut-producing countries, particularly those located between 36° and 41° northern latitudes where the climate is optimal, include Türkiye, Italy, Spain, and the United States [4]. Global hazelnut production has seen substantial growth, rising from approximately 250,000 tons in the 1960s to over 1 million tons in recent years. Türkiye leads the world in hazelnut production, accounting for about 62% of the global supply, followed by Italy, Georgia, and Azerbaijan. In the past five years, the global export of hazelnuts and hazelnut products, measured in shelled hazelnuts, has averaged 742,000 tons annually, with Türkiye contributing 72% of this total [5].

Hazelnuts have a broad range of industrial applications. Approximately 80% of hazelnuts are utilized in the chocolate industry, particularly in the production of biscuits, sweet pastries, ice cream, and confectionery, with hazelnut flour serving as a key ingredient in many chocolate products. Hazelnuts are also marketed as hazelnut paste. Those that are unsuitable for domestic or export markets are processed into oil. Hazelnut oil is refined for various uses, including cooking oils, cleansing and moisturizing products, grease production, and applications in the pharmaceutical and cosmetic industries. Additionally, it is used for disinfecting hands and rubber gloves, as well as in industrial processes such as surfactant production, lubrication, metal-cutting oils, and metal-cleaning operations. In the healthcare sector, raw hazelnut oil is employed for wound care and as an antiseptic in certain conditions. The residual meal left after oil extraction is used as animal feed. Furthermore, hazelnuts are consumed as a snack, either on their own or as an ingredient in other products [6].

Given the diverse industrial uses of hazelnuts, selecting the most appropriate variety for each specific application is essential to optimize both quality and efficiency. Classifying hazelnut varieties according to their characteristics not only meets these industrial needs but also reduces waste, lowers costs, and enhances the quality of the final products. Thus, the differentiation of hazelnut varieties is of critical importance for their industrial utilization.

In the hazelnut industry, traditional classifications such as shelled, empty, and broken have long been the standard. However, recent advancements in image processing techniques have introduced significant alternatives. For instance, a recent study tested the peeling values of hazelnuts after frying, using RGB values from the images and the K-Nearest Neighbors (KNN) algorithm. The results demonstrated that the KNN algorithm yielded superior outcomes [7]. In another study, a dataset of hazelnuts grown in the Düzce and Sakarya regions of Türkiye was initially categorized by size. The type and quality of the hazelnuts were then determined using image processing methods, specifically employing the geodesic diameter feature [8]. In another study, 17 different hazelnut varieties grown in Türkiye were analyzed for shape characteristics, and digital images of the hazelnuts were created. Shape-based classification was then performed using Elliptic Fourier Analysis (EFA) [9]. Another study on shelled hazelnuts involved processing images of hazelnuts after treatment with an industrial camera, categorizing them into three classes: clean, shelled, and rotten. The classification process utilized algorithms such as Support Vector Machine, Naive Bayes, Artificial Neural Networks, and Linear Discriminant Classifier [10]. In a study conducted by Solak and colleagues, hazelnuts were classified into three groups—small, medium, and large—based on the area they occupied on a plane, using K-means clustering methods [11]. Another study applied image processing and machine learning algorithms to detect damaged or defective hazelnuts after they had passed through a thresher, facilitating the identification and separation of defective hazelnuts from healthy ones, thereby improving product quality [12]. Finally, a dataset comprising eight different hazelnut varieties was classified using pre-trained DenseNet121 and InceptionV3 models [13].

In addition to hazelnuts, image processing and machine learning techniques have been applied to the classification of various fruits. One noteworthy study focused on supporting agricultural automation and reducing labor costs by classifying strawberries based on their shapes using machine learning algorithms. The study identified eight distinct strawberry shape types, with features such as width-to-length ratios, Ellipse Similarity Index (ESI), Elliptical Fourier Descriptors (EFD), and Chain Code Subtraction (CSS) being extracted. The classification was successfully conducted using Random Forest algorithms [14].

In another study, seven different varieties of dates, commonly used in food, medicinal, and cosmetic products, were classified. This classification was based on features such as color, size, diameter, and shape, with a total of 34 features extracted. Logistic Regression and Artificial Neural Network models were used for the classification, and the performance was further improved by employing a stacking method that combined both models [15].

For the classification of five grape varieties, a dataset was created from images of grape clusters and individual grapes. A 15-layer Convolutional Neural Network (CNN) was employed, achieving a classification accuracy of 96.10% [16].

Additionally, in a study highlighting the importance of detecting harvest products in robotic systems, cabbage and broccoli plants were classified using Python. The dataset included 176 images of cabbages and 83 images of broccoli, captured from different angles. These images were first converted to grayscale, followed by histogram equalization to enhance the classification process [17].

Literature reviews show that hazelnuts are generally classified according to quality criteria such as broken, perforated or rotten today and no distinction is made based on variety. However, since the different oil, protein and mineral contents of hazelnut varieties directly affect processing efficiency and final product quality, it is of great importance to distinguish according to variety. Failure to make this distinction in current industrial processes creates a significant deficiency in terms of both quality and economic efficiency. This study aims to fill this gap and designs and develops an artificial intelligence-supported machine that allows hazelnuts to be classified accurately and quickly according to their varieties based on their shell characteristics. Thus, variety-based distinctions can be made in post-harvest processes and suitable hazelnuts can be evaluated in the most accurate applications. The proposed method aims to increase product quality in export and domestic markets, optimize processing processes and use resources more effectively. This study is original as it is the first systematic design and application in the literature on the classification of hazelnut varieties according to their shell characteristics; and brings an important innovation in the fields of quality control and processing efficiency in the food and agricultural industry.

2. MATERIALS AND METHODS (MATERYAL VE METOD)

2.1. Machine Design and Manufacturing (Makine Tasarımı ve İmalatı)

The prototype model designed for the project consists of three main components. The first section contains the hazelnut hopper. Inside the hopper, it is divided into three compartments, allowing the hazelnuts to be held within the hopper without piling up in the conveyor slots. A hopper curtain is added at the hopper exit to ensure that any hazelnuts that do not settle into the slots as the conveyor advances are swept back into the hopper. The conveyor belt, as shown in Figure 1, consists of slots with a diameter of 30 mm and a depth of 20 mm, aligned at a 45-degree angle in the same direction. This design ensures that the hazelnuts will move within the designated slots on the belt. The second section of the prototype model incorporates a webcam system, which captures images of the hazelnuts as they move along the conveyor belt, enabling their classification and identification by variety (Figure 1). Utilizing the data gathered from the webcam, the system determines the appropriate hopper for each hazelnut and relays the corresponding positional information to the vacuum tubes. The design and placement of the sliding plates, along with the positioning of the vacuum hoses, are depicted in Figures 2.a and 2.b.

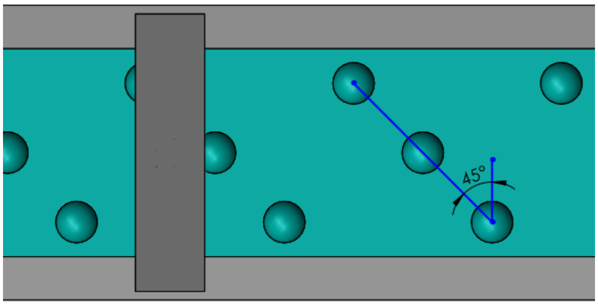
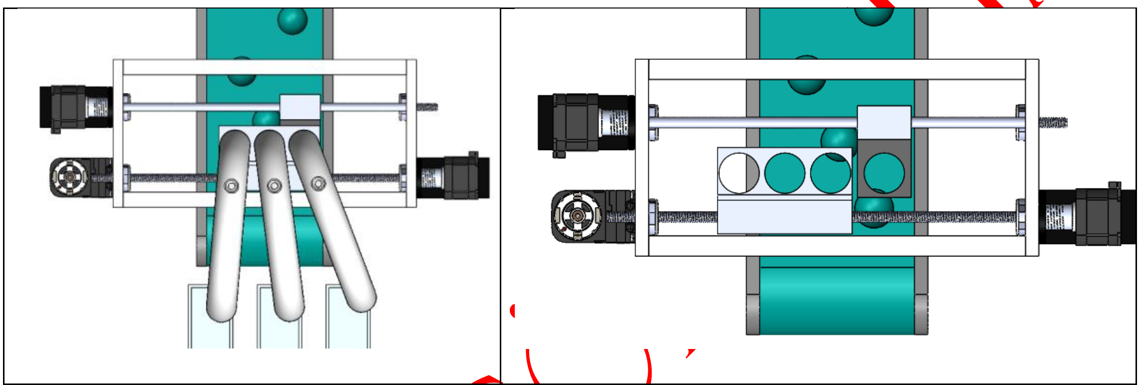


Figure 1. Arrangement of conveyor belt slots and camera placement (Konveyör bant yuvalarının düzenlenmesi ve kamera yerleşimi)



a)	b)

Figure 2. (a) Design of the vacuum system; **(b)** Sliding plate mechanism ((a) Vakum sisteminin tasarımı; (b) kayar plaka mekanizması).

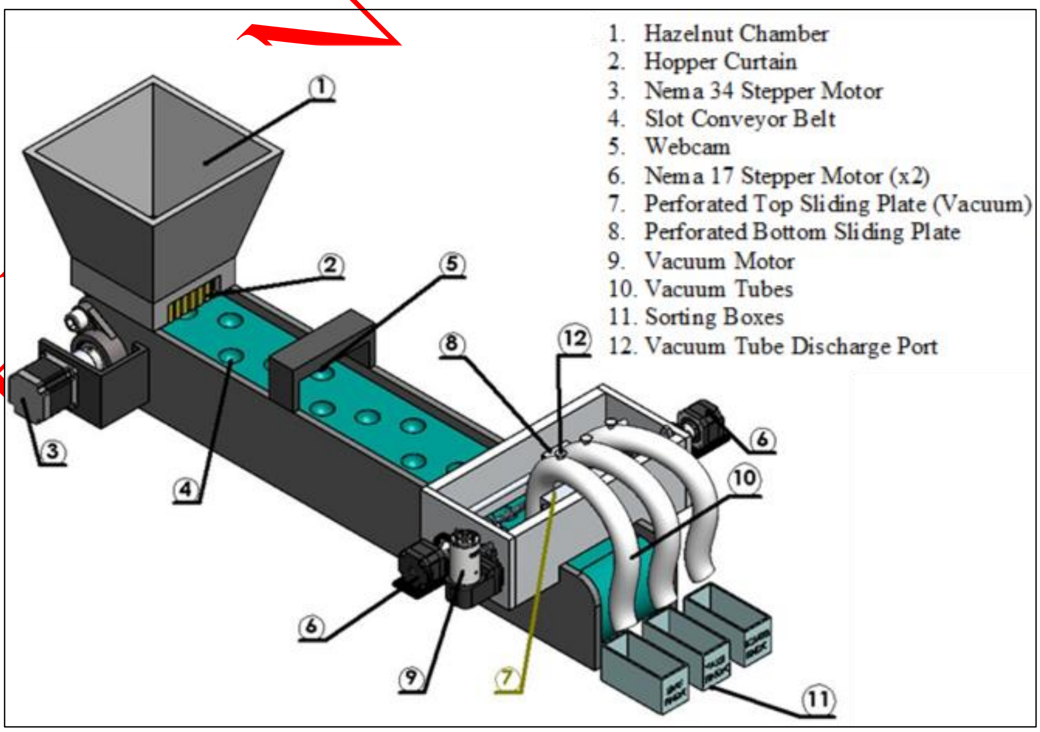
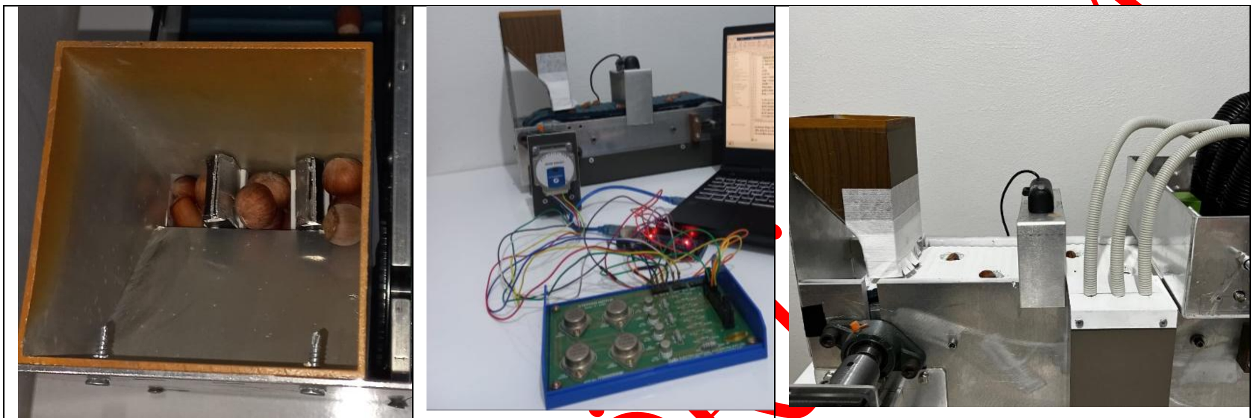


Figure 3. Prototype design scheme and used machine elements (Prototip tasarım şeması ve kullanılan makine elemanları).

The final section of the prototype model embodies the project's most original and innovative feature. In this stage, a vacuum system is utilized to carry out the classification process. Custom code developed in the Matlab App Designer interface determines the appropriate hopper for each identified hazelnut variety. The stepper motors then precisely position the vacuum system to direct the hazelnuts into their designated hoppers. In this configuration, the vacuum tubes are attached to sliding plates, with their movement orchestrated by the stepper motors. The overall system design and the specific details of the vacuum system are depicted in Figure 3.

Following the design phase, the manufacturing process of the proposed machine is depicted in Figure 4 and Figure 5. In Figure 4.a, the controlled flow of hazelnuts onto the conveyor belt is demonstrated, facilitated by the internal partitions within the hopper. Figure 4.b showcases the speed and movement operations of the conveyor system in the prototype model, driven by the stepper motor, while Figure 4.c and Figure 5 offers a detailed view of the hazelnuts neatly positioned in the slots. The names and technical specifications of the equipment used in the machine's manufacturing are provided in Table 2.



a) b) c)

Figure 4. Manufacturing stages of the proposed project **(a)** Hazelnut receiving chamber; **(b)** Electronic control system; **(c)** Camera and vacuum system (Önerilen projenin üretim aşamaları (a) Fındık alım odası; (b) Elektronik kontrol sistemi; (c) Kamera ve vakum sistemi).

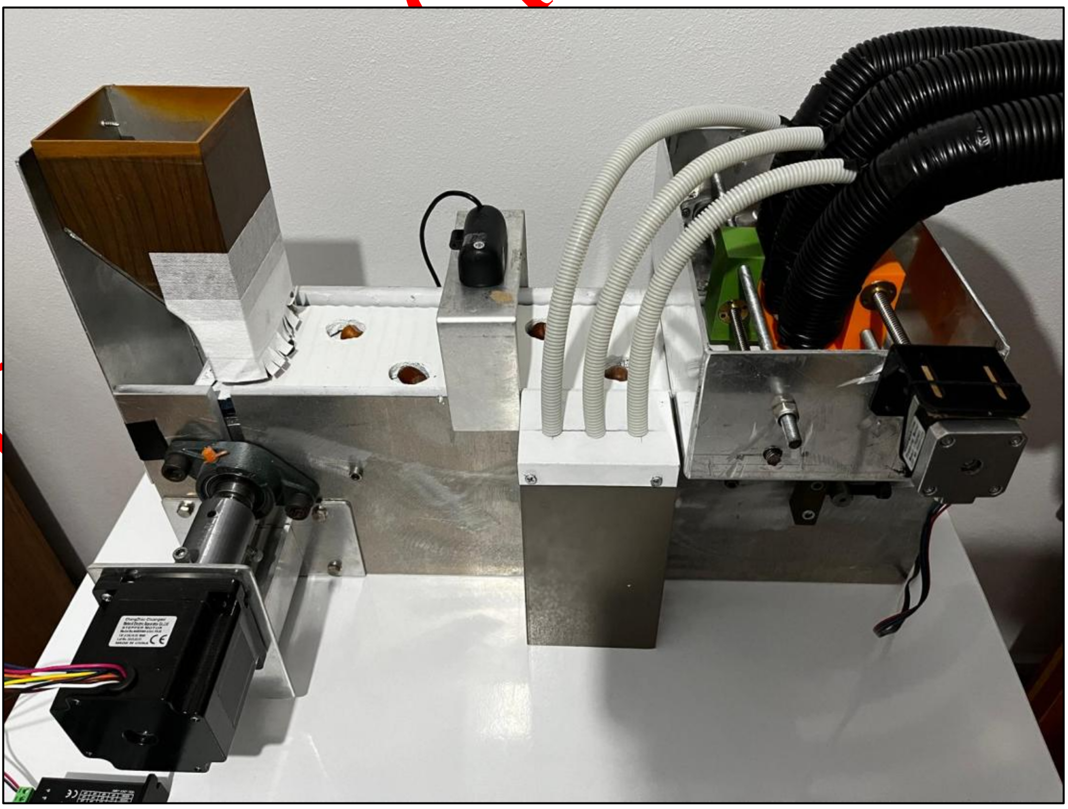


Figure 5. Detailed production images of the proposed project (Önerilen projenin detaylı üretim görüntüleri).

2.2. Dataset Creation (Veri Kümesi Oluşturma)

In this study, we selected the three most commonly cultivated hazelnut varieties in Türkiye. The dataset was constructed by choosing hazelnut varieties that display distinct differences in shape, size, texture, and shell color, ensuring that feature extraction would yield high accuracy. Accordingly, a broad classification of hazelnut varieties grown in Türkiye was conducted, resulting in the selection of three types: Tombul, Badem and Sivri. Tombul hazelnuts are a prominent variety in Türkiye. They feature a bright brown shell with a fuzzy, dirty white texture extending halfway from the tip. The base is broad, flat, and slightly elevated in the center. The average shell thickness is 1.1 mm, making it easy to crack. The average length is 17.58 mm, and the width is 17.04 mm. Thanks to their plump kernels, Tombul hazelnuts offer high yield efficiency. Their oil content is approximately 69-72%. Figure 6 (a) illustrates their general appearance [18].

Table 2. Equipment specifications (Ekipman özellikleri).

Equipment and Sensors	Technical Specifications
Nema 17 Stepper Motor	2 Phase, 200 steps, 1.8 degree step angle, 28 N.cm Holding Torque, 42x42 flange size.
Nema 34 Stepper Motor	2 Phase, 1.8 degree step angle, 4.5 Nm motor torque, 86x86 mm motor flange size.
DC 12V Vacuum Motor	Air vacuum motor,12V, 10.5 L/dk, 80 Kpa vacuum power.
Screw shaft-nut	Stainless Steel, 400 mm Shaft length, 8 mm Diameter, 8 mm Pitch
ESA IF-Step Stepper Motor Driver	4.5 A current, 20-50V DC, over-current, over-low voltage, 1.8 degree step angle, Pulse and direction signals
CWD-556 Stepper Motor Driver	Max 5.6 A, 20-50 V DC, over-current, over-low voltage, over-temperature protection,
Arduino Mega 2560	54 Digital I/O pins, 16 Analog pins, 256 KB Flash memory, 5V Operating Voltage, 7-12 V, Pulse and direction signals.
FRM CJT 05	Stepper motor fixing bearing
Webcam	1080p Quality (1920x1080 pixels), CMOS, Anti-glare sensor, 16 Mpixel image capture (4608x3456), USB 2.0 port connection

a) b) c)

Figure 6. Hazelnut Varieties **(a)** Tombul Hazelnut **(b)** Sivri Hazelnut **(c)** Badem Hazelnut (Fındık Çeşitleri (a) Tombul Fındık (b) Sivri Fındığı (c) Badem Fındığı).

Sivri hazelnuts are a variety that thrives in nearly all hazelnut-producing regions of Türkiye. As illustrated in Figure 6. (b), they are characterized by a flattened, elongated, and pointed structure. The shell is bright and light brown, with a fuzzy, dirty white tip. The average length is 20.71 mm, and the width is 14.88 mm. The yield rate is approximately 49-50%, with an oil content of 65-68%. Due to the greater damage they sustain during cracking, Sivri hazelnuts are typically marketed in-shell appearance [19]. Badem hazelnuts are distinguished by their flattened and elongated shape. The shell is bright brown with an average thickness of 1.3 mm. The nuts measure 24.55 mm in length and 14.93 mm in width. The kernel fits the shape of the shell but is encased in a rather thick membrane. The yield rate is approximately 48-49%, and the oil content ranges from 57-62%. Despite their large and striking appearance, Badem hazelnuts are considered lower in quality. Their appearance is depicted in Figure 6. (c) appearance [20].

A high-resolution webcam was primarily used to create the dataset. This webcam offers 1080p image quality (1920x1080 pixels), features a CMOS sensor, an anti-reflective sensor, 16-megapixel image capture (4608x3456), automatic focus and white balance, and a USB 2.0 port connection for direct data transfer without the need for external software. The camera was mounted on a perforated plate, positioned 190 mm from the hopper exit to capture images at a 90-degree vertical angle to the conveyor belt. The camera's distance from the belt surface was set to 60 mm, enabling it to simultaneously capture three hazelnut slots within its field of view.

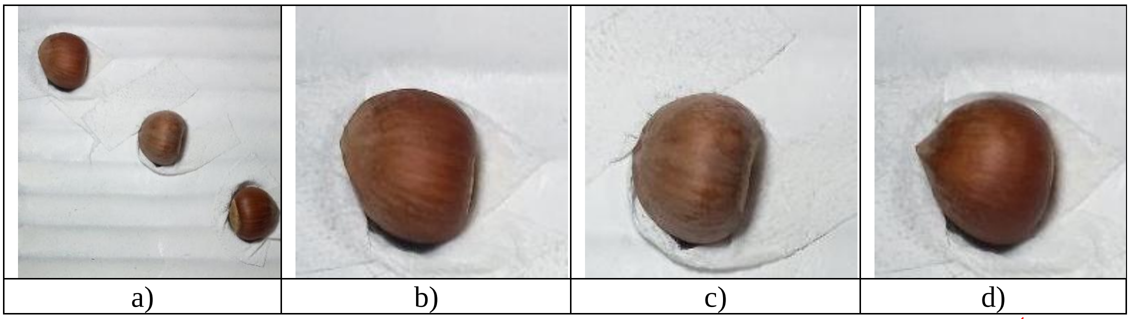


Figure 7. (a) Original **(b)** Cropped images captured from the conveyor belt1 **(c)** Cropped images captured from the conveyor belt2 (d) Cropped images captured from the conveyor belt3 ((a) Orijinal (b) Konveyör banttan çekilen kırpılmış görüntüler1 (c) Konveyör banttan çekilen kırpılmış görüntüler2 (d) Konveyör banttan çekilen kırpılmış görüntüler3).

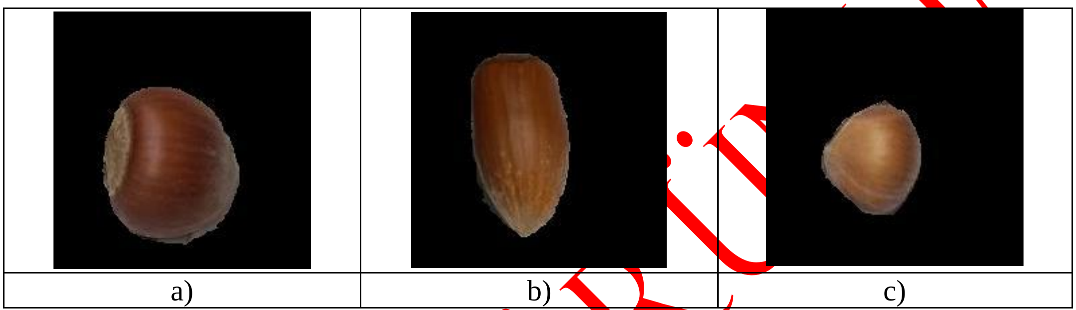


Figure 8. (a) Preliminary image processing steps for Tombul; **(b)** Preliminary image processing steps for Badem; **(c)** Preliminary image processing steps for Sivri ((a) Tombul için ön görüntü işleme adımları; (b) Badem için ön görüntü işleme adımları; (c) Sivri için ön görüntü işleme adımları).

2.3. Preliminary Image Processing (Ön Görüntü İşleme)

In the initial phase, images of the three slots on the conveyor belt (Figure 7.a) were captured and subsequently cropped to a size of 190x190 pixels using Matlab (Figures 7.b, c, d).

To facilitate further processing of the hazelnut images, background removal was performed in Matlab. To maintain the original dimensions of the hazelnuts, the canvas size was restricted to 190x190 pixels during the background removal process (Figure 8).

2.4. Feature Extraction Methods (Özellik Çıkarma Metotları)

Following object extraction and background removal, feature extraction was performed on the images. The selected feature extraction methods focused on highlighting differences in size, color, and texture, thereby ensuring high classification accuracy. This process was applied to a total of 300 images, resulting in the generation of 100 feature columns.

Dimension Properties: To analyze the hazelnuts in the images, the first step involved converting the images to grayscale. Subsequently, thresholding was applied to create binary images. In these binary images, dimensional properties such as area, perimeter, eccentricity, roundness, convex area, and filled area were calculated for each hazelnut. This approach facilitated a comprehensive analysis of the hazelnuts' shape characteristics, enabling the extraction of key features.

Color Properties: Color properties were extracted to capture the hazelnuts' color characteristics. The initial step involved analyzing each image's RGB channels, with the average values calculated to obtain the RGB features. In the next step, the images were converted from the RGB color space to the HSV (Hue, Saturation, Value) color space, and the average pixel values for each channel were calculated individually. Finally, the images were transformed from the RGB space to the Lab color space, where the brightness, the green-to-red color axis, and the blue-to-yellow color axis values were obtained, providing both brightness and color values as essential features.

Haralick Method: The Haralick method utilizes the Grey Level Co-occurrence Matrix (GLCM) to generate matrices for each directional orientation within an image. After extracting the GLCM features for each direction, the average values are computed to facilitate segmentation. The Haralick method defines 14 distinct features for each image

[21]. In this study, 10 feature values were extracted from hazelnut images using the Haralick method, including contrast, correlation, difference, inverse difference moment, total average, total variance, total entropy, and maximum correlation coefficient.

LBP Algorithm (Local Binary Patterns): The Local Binary Patterns (LBP) algorithm is a robust method for extracting texture features from images. It analyzes images at the pixel level by first converting them to grayscale. The intensity values of each pixel are then labeled within a range of 0-255. In the resulting binary image, a central reference region is selected; pixels with intensity values lower than the reference are labeled as 0, and those with higher values are labeled as 1, forming a binary pattern [22]. The LBP algorithm is particularly effective when combined with the HOG algorithm, which is why it was selected for this study.

HOG Algorithm (Histogram of Oriented Gradients): Initially introduced in 2005 for human detection in images [23], the HOG algorithm calculates gradient orientations that represent local edge structures within an image. The process begins by dividing the image into small cells. The gradient directions of the pixel values within each cell are calculated and grouped into specific intervals to create histograms. These histograms are then aggregated into larger blocks and normalized, enhancing their robustness to variations and changes in the image.

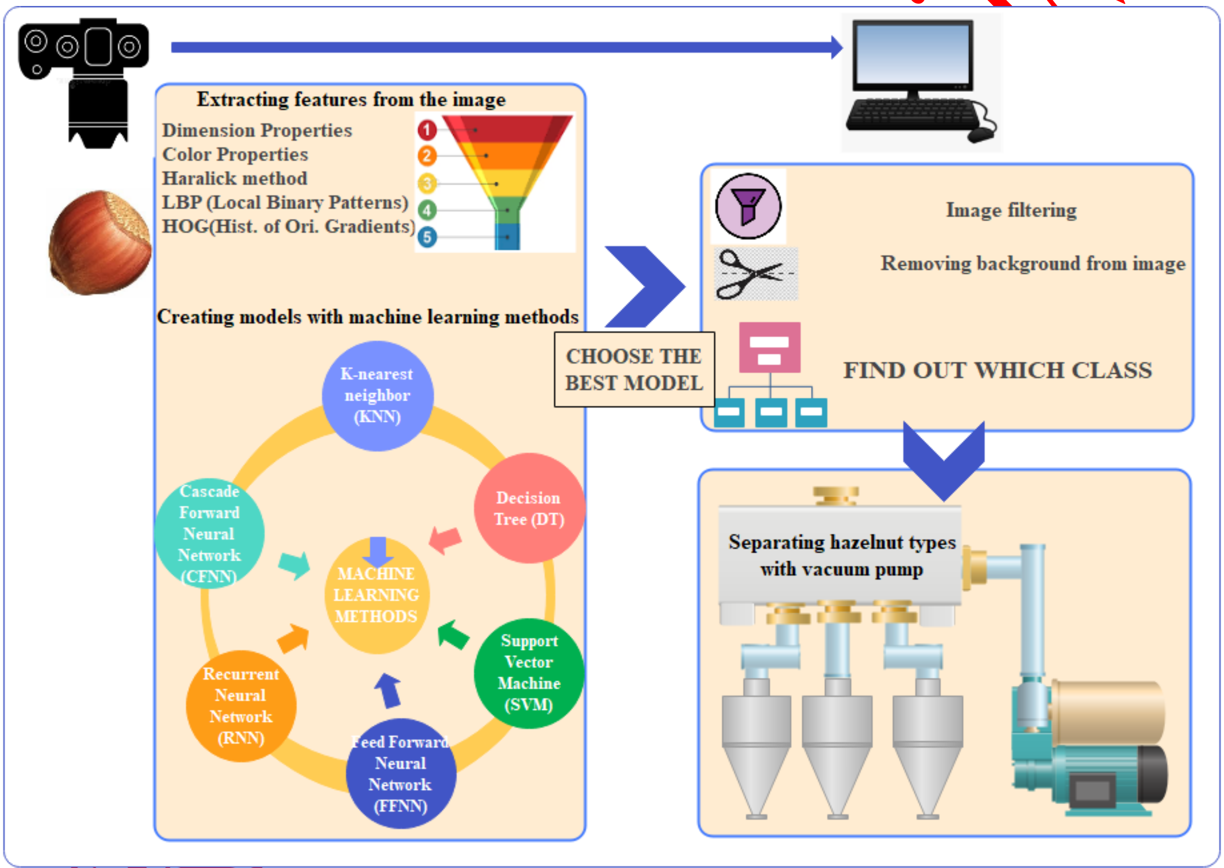


Figure 9. Basic flow chart of the study that separates hazelnut types using machine learning methods (Makine öğrenmesi yöntemlerini kullanarak fındık türlerini ayıran çalışmanın temel akış şeması).

2.5. Machine Learning Algorithms Using Hazelnut Classification (Fındık Sınıflandırmasını Kullanan Makine Öğrenmesi Algoritmaları)

Machine learning consists of a variety of computational algorithms capable of learning from data and making insightful predictions. The primary goal of these methods is to accelerate the iterative training process and facilitate the application of large, complex models as datasets continue to grow in size and complexity [24]. In the proposed study, classification is conducted according to the flowchart presented in Figure 9. Among the commonly used classical machine learning algorithms, K-Nearest Neighbor (k-NN), Naive Bayes Algorithm (NBA), and Decision Tree (DT) were selected. Additionally, from the family of artificial neural networks, Feedforward Neural Networks (FFNN), Recurrent Neural Networks (RNN), and Cascade Forward Neural Networks (CFNN) were employed. k-NN, introduced by Cover and Hart is determined by its nearest neighbors, with the number of neighbors (k value) playing a crucial role [25]. A key factor affecting the algorithm's classification accuracy is the appropriate selection of the k value. If this value is too high or too low, patterns that should belong to the same class may be misclassified into different classes. Additionally, the size of the data patterns can influence the processing speed of k-NN [26].

The Decision Tree , inspired by human decision-making processes, was first introduced in the 1960s and has since remained one of the most effective methods for predictive modeling [27]. This model can handle both numerical and categorical data, transforming a data example into a set of manageable rules that can be used for classification or prediction [28]. Structured in a hierarchical tree format, it comprises root nodes, internal nodes, and leaf nodes [29].

SVM, developed in the early 1990s, is a powerful classification and regression method widely used in machine learning. Its goal is to identify the optimal hyperplane that best separates the data points. Classification is achieved by leveraging the gaps between points on this hyperplane, and SVM has demonstrated the ability to yield effective results even with complex data [30].

In the proposed study, the first artificial neural network model used is the FFNN. Introduced in 1998, FFNN is a fundamental form of artificial neural network. Despite its relatively simple architecture, it possesses a highly effective learning capability [31]. In this model, information flows unidirectionally, moving forward through the network. It is composed of three primary layers: the input layer, hidden layers, and the output layer. Data enters through the input layer, passes through the hidden layers, and the final classification results are produced at the output layer [32].

The second model utilized in this study is the RNN algorithm. RNN generates bounding box proposals from images and operates by analyzing these regions. It uses a selective search method to identify the bounding box regions, extracting key features such as color, texture, perimeter, and scale-changing patterns. While RNN delivers high accuracy, it suffers from limitations related to processing speed and significant memory requirements [33].

While CFNN shares similarities with FFNN, it includes additional weight connections. These extra connections help accelerate the learning of input-output relationships [34]. Both CFNN and FFNN use the backpropagation algorithm for weight updates. However, the key distinction in cascade networks is that each layer's neurons are directly connected and updated with neurons from all preceding layers, providing more comprehensive and efficient weight connections [35].

2.6. Metrics for Evaluating Hazelnut Variety Classification Results (Fındık Çeşit Sınıflandırma Sonuçlarının Değerlendirilmesine Yönelik Ölçütler)

Confusion Matrix is used for the performance evaluation of the classification methods used in the study. This matrix provides detailed insights into the relationship between actual and predicted classifications.

Table 3. The confusion matrix (Karışıklık matrisi).

		Normal Condition		i. Fault
True Class	Normal Condition	n_{11}	. . .	n_{1j}
		. :		
	i. Fault	n_{i1}		n_{ij}
	Predicted Class			

In Table 3, the notation n_{ij} is used to represent the classification outcome. In this notation, the subscript i indicates the class predicted by the classifier, while the subscript j corresponds to the actual class. The elements that lie on the diagonal of the matrix (where $i = j$) signify instances that have been correctly classified. In contrast, the elements that are off-diagonal (where $i \neq j$) represent instances that have been misclassified.

When it comes to evaluating a specific class i, the examples can be categorized into four distinct groups. These groups include true positives (TP) and true negatives (TN), which are instances where the model has made correct predictions. Additionally, there are false positives (FP) and false negatives (FN), which are instances where the model has made incorrect predictions. This categorization aids in assessing the performance of the classifier for each individual class.

$$Acc = \frac{n_{TP} + n_{TN}}{n_{TP} + n_{TN} + n_{FP} + n_{FN}} \quad 2.1$$

$$Sen = \frac{n_{TN}}{n_{TN} + n_{FP}} \quad 2.2$$

$$Pre = \frac{n_{TP}}{n_{TP} + n_{FP}} \qquad 2.3$$

$$F - Scor = 2 + \frac{Pre * Sen}{Pre + Sen} \qquad 2.4$$

$$G - Mean = \sqrt{\left(\frac{n_{TP}}{n_{TP} + n_{FN}} * \frac{n_{TN}}{n_{TN} + n_{FP}}\right)} \qquad 2.5$$

The criteria and formulas for evaluating the classifier performance are given in Equations 2.1-2.5, respectively. Accuracy is the ratio of correctly predicted instances to the total dataset. Precision measures how many predicted positives are actually positive. Sensitivity shows the proportion of actual positives correctly identified. The F-Score is the harmonic mean of precision and sensitivity, ranging from 0 to 1. The G-Mean reflects the balance between performance in majority and minority classes [36].

3. RESULTS (BULGULAR)

In this study, a machine capable of classifying hazelnuts by variety using artificial intelligence and machine learning algorithms was successfully designed and manufactured. Images were initially captured from the conveyor belt of the system, and the original image was divided into three equal sections, with each section containing a visual of one slot on the belt. Prior to feature extraction, object extraction and background removal processes were applied to the images to streamline the feature extraction process.

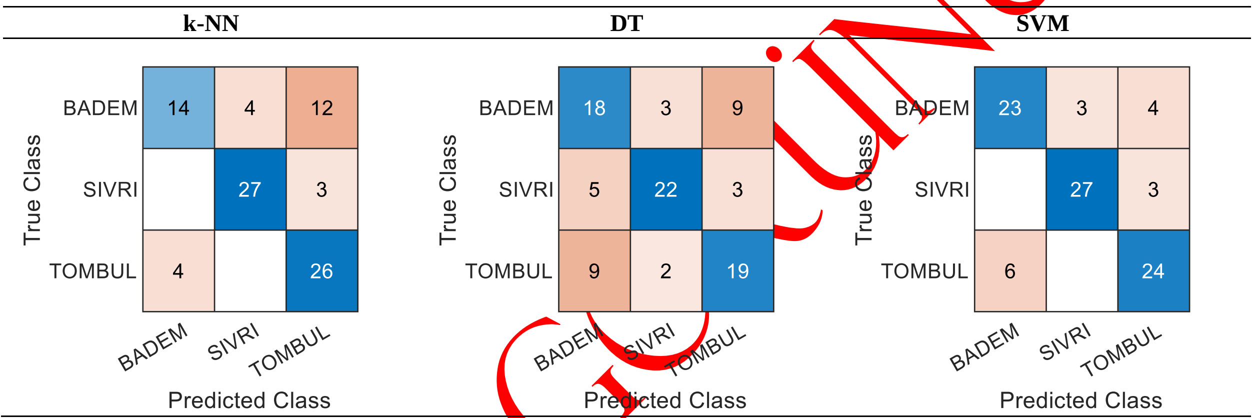


Figure 10. Confusion Matrix Results for Classical Machine Learning Algorithms (Klasik Makine Öğrenmesi Algoritmaları için Karışıklık Matrisi Sonuçları).

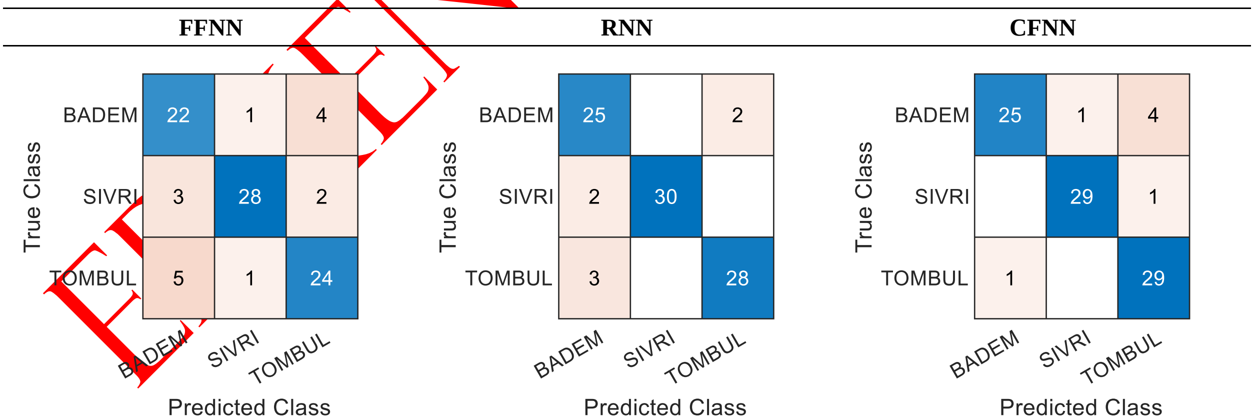


Figure 11. Confusion Matrix Results for Neural Network Based Machine Learning Algorithms (Sinir Ağı Tabanlı Makine Öğrenmesi Algoritmaları için Karışıklık Matrisi Sonuçları).

A total of 90 hazelnut images, consisting of 30 images from each of the three hazelnut varieties, were analyzed focusing on color, size, and texture characteristics, resulting in 100 feature columns.The dataset generated from these features was then used to train both traditional machine learning algorithms and Artificial Neural Networks models, yielding classification results.

The study was conducted on a computer with an Intel(R) Core (TM) i7-10750H CPU @ 2.60 GHz processor and 16.00 GB of RAM, utilizing MATLAB version 2021a.

Figures 10 and 11 display the results for the k-NN, DT, SVM, FFNN, RNN, and CFNN algorithms, respectively. The same feature extraction methods were applied across all algorithms, with the CFNN algorithm demonstrating the highest performance. For instance, in Figure 10, the k-NN classifier produced correct classification results of 14 for Badem, 27 for Sivri, and 26 for Tombul. In contrast, Figure 11 illustrates that the CFNN classifier achieved 25 correct classifications for Badem, 29 for Sivri, and 29 for Tombul.

Table 4. Evaluation criteria results for measuring the success of machine learning methods (Makine öğrenmesi yöntemlerinin başarısını ölçmek için değerlendirme kriteri sonuçları).

	Class. Metod	**Accuracy**	**Sensitivity**	**Precision**	**F-Score**	**G-Mean**	**Time**
Classical Machine Learning Algorithms	k-NN	0.778	0.867	0.619	0.722	0.797	2.184
	DT	0.667	0.700	0.500	0.583	0.675	0.821
	SVM	0.822	0.800	0.706	0.750	0.816	1.657
Neural Network Based Machine Learning Algorithms	FFNN	0.878	0.900	0.771	0.831	0.883	13.130
	RNN	0.922	0.933	0.848	0.889	0.925	23.809
	CFNN	0.922	0.967	0.829	0.892	0.933	39.617

In Table 4, the performance metrics for all classifiers, derived from the selected features of the experimental dataset and depicted in Figures 12 and 13, are outlined. The metrics considered are accuracy, sensitivity, precision, F-Score, and G-Mean. The CFNN classifier demonstrated impressive values: 0.922 for accuracy, 0.967 for sensitivity, 0.829 for precision, 0.892 for F-Score, 0.933 for G-Mean, and 39.617 as an additional metric. These results clearly indicate that the CFNN algorithm outperforms other classifiers in terms of accuracy and other metrics. However, it is also important to note that the CFNN algorithm takes a longer time to complete the classification process compared to other methods. Therefore, for conveyor belts operating at higher speeds, the FFNN and RNN algorithms may be more suitable, whereas the CFNN algorithm is better suited for applications where the belt speed is slower. This trade-off between accuracy and processing time is crucial when selecting the appropriate algorithm for a specifie application.

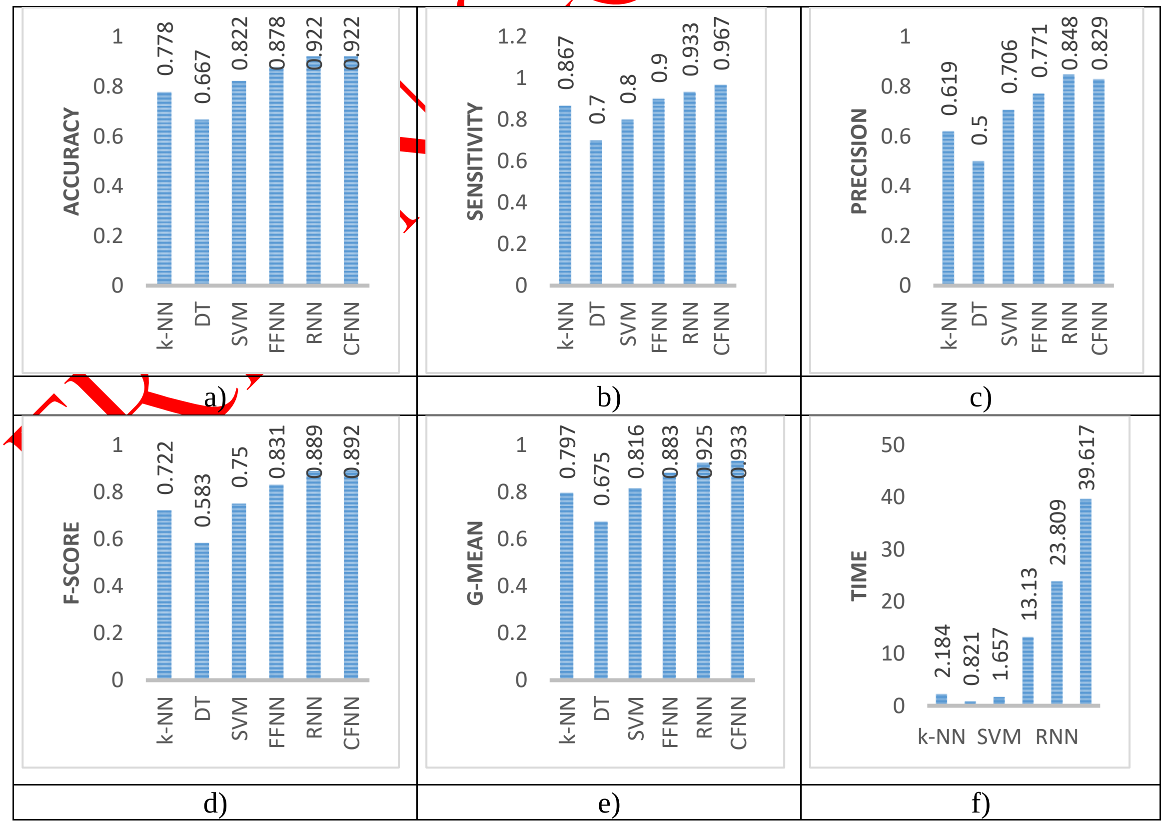


Figure 12. Metric evaluation graphs by classifiers **(a)** Accuracy; **(b)** Sensivitiy; **(c)** Precision; **(d)** F-score; **(e)** G-mean; **(f)** Time (Sınıflandırıcılara göre metrik değerlendirme grafikleri (a) Doğruluk; (b) Duyarlılık; (c) Kesinlik; (d) F puanı; (e) G ortalaması; (f) Zaman).

Based on the graphical metric evaluation results presented in Figure 12, RNN and CFNN achieved the highest accuracy, both with a value of 0.922, while DT emerged as the least successful classifier with an accuracy of 0.667. In terms of sensitivity, CFNN outperformed the other classifiers with a value of 0.967, whereas DT had the lowest score at 0.7. For precision, RNN led with a value of 0.848, while DT again recorded the poorest result at 0.5. This trend is consistent across the F-Score and G-Mean metrics as well. The poor performance of DT can be attributed to the fact that the dataset is not based on linear and easily separable relationships. Given the complex structure of the dataset, neural networks such as RNN and CFNN were better suited to learn these intricate relationships, resulting in their superior performance compared to other classifiers.

4. CONCLUSIONS (SONUÇLAR)

This study presents the design and development of a machine capable of classifying hazelnuts by variety using machine learning algorithms and Artificial Neural Networks models. The images were processed through object extraction and background removal, followed by feature extraction focused on color, size, and texture attributes. The extracted features were then used as input for classification and performance evaluation through k-NN, DT, SVM, FFNN, RNN, and CFNN algorithms.

The key findings from the study are as follows:

- Extracting size, color, and texture attributes from the outer shell of hazelnuts significantly enhanced the performance of the classifiers.
- In terms of performance metrics, including accuracy, CFNN and RNN algorithms outperformed FFNN, k-NN, DT, and SVM, with values of 0.922, 0.967, 0.829, 0.892, 0.933, and 39.617 for CFNN, and 0.922, 0.933, 0.848, 0.889, and 23.809 for RNN. Although classification time is a consideration, CFNN remains a viable classification method in this study due to its superior performance, which is attributed to its better results compared to the RNN algorithm in the metrics of Sensitivity, Precision, F-Score, and G-Mean.

The proposed method is the first in the literature to report a design and manufacturing approach for identifying hazelnut varieties. The study demonstrates that the classification results can be effectively used to distinguish different hazelnut types.

DECLARATION OF ETHICAL STANDARDS
(ETİK STANDARTLARIN BEYANI)

The author of this article declares that the materials and methods they use in their work do not require ethical committee approval and/or legal-specific permission.

Bu makalenin yazarı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler.

AUTHORS' CONTRIBUTIONS (YAZARLARIN KATKILARI)

***Rabia KAYMAK*:** She conducted the experiments, analyzed the results and performed the writing process.

Deneyleri yapmış, sonuçlarını analiz etmiş ve makalenin yazım işlemini gerçekleştirmiştir.

***Ferzan KATIRCIOĞLU*:** He carried out the creation phase of the method and took part in the design and manufacturing phase.

Yöntemin oluşturulma safhasını gerçekleştirmiştir. Tasarım ve imalat safhasında yer almıştır.

CONFLICT OF INTEREST (ÇIKAR ÇATIŞMASI)

There is no conflict of interest in this study.

Bu çalışmada herhangi bir çıkar çatışması yoktur.

REFERENCES (KAYNAKLAR)

[1] T.C. Tarım ve Orman Bakanlığı Fındık Araştırma Enstitüsü Müdürlüğü., Savran E. (2023). Gıda Olarak Fındığın Değeri. [Online]: https://arastirma.tar.gov.tr/findik/Sayfalar/Sayfa=29

[2] Özdemir F., Topuz A., Doğan Ü., Karkacıer M.. Fındık Çeşitlerinin Bazı Fiziksel ve Kimyasal Özellikleri. Gıda, 23(1), 37-41, (1998).

[3] Acıbadem Hastanesi (2022). Çiğ fındığın sağlığımıza kattığı 10 fayda. [Online]: https://www.acibadem.com.tr/hayat/cig-findigin-sagliginiza-kattigi-10-fayda.

[4] Aktaş A., Öztürk E., Hatırlı S. Dünya Fındık Piyasasında Türkiye'nin Rolü. Süleyman Demirel Üniversitesi Vizyoner Dergisi, 1(1), 36-54, (2016).

[5] Toprak mahsulleri Ofisi Genel Müdürlüğü. (2021). 2020 Yılı Fındık Sektör Raporu Ankara, 2021.

[Online]:https://www.tmo.gov.tr/Upload/Document/sektorraporlari/findik2020.pdf

[6] Giresun Ziraat Odası. (2022). Fındığın kullanım alanları. [Online]: https://www.giresunziraatodasi.org.tr/findigin-kullanim-alanlari

[7] Pallottino, F., Menesatti, P., Costa, C., Paglia, G., De Salvador, F. R., & Lolletti, D. Image analysis techniques for automated hazelnut peeling determination. Food and Bioprocess Technology, 3, 155-159, (2010). DOI 10.1007/s11947-009-0211-1

[8] Bayrakdar, S., B. Çomak, D. Başol, İ. Yücedag D. Determination of type and quality of hazelnut using image processing techniques," 2015 23nd Signal Processing and Communications Applications Conference (SIU), Malatya, Türkiye, (2015).

[9] Sayıncı, B., Kara, M., Ercişli, S., Duyar, Ö., Ertürk, Y. Elliptic Fourier analysis for shape distinction of Turkish hazelnut cultivars. Erwerbs-Obstbau, 57(1), 1-11, (2015). DOI 10.1007/s10341-014-0221-7.

[10] Guvenc, S. A., Senel F. A., B. Cetisli. Classification of processed hazelnuts with computer vision," 2015 23nd Signal Processing and Communications Applications Conference (SIU), Malatya, Türkiye, (2015).

[11] Solak, S., Altınışık, U. Görüntü işleme teknikleri ve kümeleme yöntemleri kullanılarak fındık meyvesinin tespit ve sınıflandırılması. Sakarya University Journal of Science, 22(1), 56-65, (2018). DOI 10.16984/saufenbilder.303850.

[12] Ishikawa, T., Hayashi, A., Nagamatsu, S., Kyutoku, Y., Dan, I., Wada, T., Oku, K., Saeki, Y., Uto, T., Tanabata, T., Isobe, S., & Kochi, N. Classification of strawberry fruit shape by machine learning. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 42(2), 463–470, (2018).

[13] Kıvrak, O., Gürbüz, MDetection of Defective Hazelnuts by Image Processing and Machine Learning. Natural and Engineering Sciences, 4(3), Supplemenet of Symposium Articles, 100-106, (2019).

[14] Koklu, M., Kursun, R., Taspinar, Y. S., Cinar, I. Classification of Date Fruits into Genetic Varieties Using Image Analysis. Mathematical Problems in Engineering, 2021. DOI 10.1155/2021/4793293.

[15] Terzi, İ., Özgüven, M. M., Yağcı, A. Derin Öğrenme Teknikleri ile Bazı Üzüm Çeşitlerinin Tespiti. Turkish Journal of Agriculture - Food Science and Technology, 11(1), 125–130, (2023). DOI 10.24925/turjaf.v11i1.125-130.5722.

[16] Kahya, E., Özdüven, F. F. Robotik Hasat Sistemlerinde Kullanılmak Amacıyla Lahana ve Brokolinin Derin Öğrenme Metodu ile Sınıflandırılması. Turkish Journal of Agriculture- Food Science and Technology, 11(9), 1639–1647, (2023).

[17] Harmancı, S., Yavuz, Ünal., Barış, A. Classification of hazelnut species with pre-trained deep learning models. Intelligent Methods In Engineering Sciences, 2(2), 58-66, (2023). DOI 10.58190/imiens.2023.14.

[18] T.C. Tarım ve Orman Bakanlığı Fındık Araştırma Enstitüsü Müdürlüğü, (2014). Tombul, [Online]:https://arastirma.tarimorman.gov.tr/findik/Sayfalar/Detay.aspx?SayfaId=17

[19] T.C. Tarım ve Orman Bakanlığı Fındık Araştırma Enstitüsü Müdürlüğü, (2014). Sivri, [Online]:https://arastirma.tarimorman.gov.tr/findik/Sayfalar/Detay.aspx?SayfaId=9

[20] T.C. Tarım ve Orman Bakanlığı Fındık Araştırma Enstitüsü Müdürlüğü, (2014). Yassı Badem, [Online]:https://arastirma.tarimorman.gov.tr/findik/Sayfalar/Detay.aspx?SayfaId=16

[21] Haralick, R. M., Shanmugam, K., Dinstein, I. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3(6), 610-621, (1973). DOI 10.1109/TSMC.1973.4309314.

[22] Ojala T., M. Pietikäinen, and D. Harwood. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR 1994), vol. 1, pp. 582- 585, (1994). DOI 10.1109/ICPR.1994.576366.

[23] Dalal, N., Triggs, B. Histograms of Oriented Gradients for Human Detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1, 886-893, (2005). DOI 10.1109/CVPR.2005.177.

[24] Gibert, D., Mateu, C., Planes, J. The rise of machine learning for detection and classification of malware: Research developments, trends and challenges. Journal of Network and Computer Applications, 153, (2020). DOI 10.1016/j.jnca.2019.102526.

[25] Cover, T., & Hart, P. Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1), 21-27, (1967). DOI 10.1109/TIT.1967.1053964

[26] Gou, J., Qiu, W., Yi, Z., Shen, X., Zhan, Y., & Ou, W. Locality constrained representation-based K-nearest neighbor classification. Know.-Bas. Systems, 167, 38-52, (2019). DOI 10.1016/j.knosys.2019.01.016.

[27] Hastie T., Friedman J. , Tibshirani R. The Elements of Statistical Learning, New York, 2009.

[28] Song Y. Y., Ying L. U. Design tree methods: application for classification and prediction. Shangai Arcivhes of psychiatry. (27)2, 130, (2015). DOI 10.11919/j.issn.1002-0829.215044.

[29] Santos I., Brezo F., Ugarte-Pedro X., Bringas P. G. Opcode sequences as representation of executables for data-mining-based unknown malware detection. Information Sciences (231), 64-82, (2013). DOI 10.1016/j.ins.2011.08.020.

[30] Vapnik, V., & Cortes, C. Support-Vector Networks. Machine Learning, 20(3), 273-297, (1995).

[31] Haykin, S. Neural Networks: A Comprehensive Foundation. Prentice Hall, (1998).

[32] Bishop, C. M. Pattern Recognition and Machine Learning. Springer, (2006).

[33]Girshick, R. Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition, 580-587, (2014).

[34] Beale, M. H., Hagan, M. T., Demuth, H. B. Neural network toolbox user’s guide. The MathWorks, 77-81, (2010).

[35] Chayjan, R. A. Modeling of sesame seed dehydration energy requirements by a soft-computing approach. Australian journal of crop science, 4(3), 180-184, (2010).

[36] He, H., Garcia, E. A. Learning from imbalanced data. IEEE Transactions on knowledge and data engineering, 21(9), 1263-1284, (2009). DOI 10.1109/TKDE.2008.239.