ESKİŞEHİR TECHNICAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
A- APPLIED SCIENCES AND ENGINEERING

2020, 21(4), pp. 499-513, DOI: 10.18038/estubtda.658077

RESEARCH ARTICLE

WEIGHT-BASED CLASSIFICATION OF EGGS USING SEVERAL STATE-OF-THE-ART CLASSIFIERS ON A MECHANICAL WEIGHING SYSTEM INTEGRATED WITH A DSP MICROCONTROLLER

Gülin ELİBOL SEÇİL [1], Mehmet YUMURTACI [2,*], Semih ERGİN [1], İsmail YABANOVA [3]

[1] Department of Electrical-Electronics Engineering, Eskisehir Osmangazi University, Eskisehir, Turkey
[2] Department of Electrical-Electronics Engineering, Afyon Kocatepe University, Afyonkarahisar, Turkey
[3] Department of Mechatronics Engineering, Afyon Kocatepe University, Afyonkarahisar, Turkey

ABSTRACT

In this study, a feature vector is determined in order to classify chicken eggs into four different weight groups by using the dynamic weighing system, and then the success rates of different classifiers in the process of weight classification are analyzed. The class numbers and labels of the egg weight groups are determined as 1-extra large, 2-large, 3- middle, and 4-small, respectively. The dynamic weighing system is made of three components; mechanic system, electronic control board, and software. Firstly, a data set is created on the basis of analogue egg weight data obtained from the dynamic weighing system. From the obtained data set, three different feature vectors are extracted by using Time-Domain (TD), Power Spectral Density (PSD) and Discrete Wavelet Transform (DWT) based methods. The extracted feature vectors are then applied to Linear Bayes Normal Classifier (LBNC), Fisher's Linear Discriminant Analysis (FLDA), Support Vector Machine (SVM), Decision Tree (DT), and K-Nearest Neighborhood (k-NN) classifiers respectively and egg weight classes are determined. A five-fold cross-validation is carried out in order to confidentially test the performance of classification. As can be seen from the experimental results, both feature vectors and classifiers are highly successful in determining the weight classes of eggs. It is observed that the most successful features are the entropy values of DWT with a classification rate of 97.01% for the k-NN classifier.

Keywords: Dynamic weighing system, Measurement, Signal processing, Feature extraction, Egg classifying

1. INTRODUCTION

An egg is a protein, nutrient, and vitamin-rich resource [1]. Egg, used in various foodstuffs, is a significant part of the food sector [2]. Egg weight is one of the important parameters in marketing; in addition to this, it has a significant role in determining quality indexes such as albumen ratio, eggshell thickness, and hatchability [3]. The weight of the albumen varies according to the spawning time, age, and size, nutrition of the chicken and the light of the environment [4]. Eggs are generally classified according to their sizes, but it is more appropriate and economic to classify them on the basis of their weight. This method necessitates obtaining information about the relationship between the weight and some geometrical aspects of the egg [5].

At this point, weighing systems are used in order to obtain this significant information. Weighing systems are divided into two; static and dynamic systems. In a static weighing system, object weights are measured when they are on a weighing platform in a fixed position until they reach steady-state value. In this weighing method, although high reliability and renewability are guaranteed, there are some disadvantages such as limited weight because of the long operation cycle. The alternative solution to this method is dynamic weighing which is widely used in industrial production processes. Dynamic weighing systems can weigh more products with accuracy; objects are weighed when they are in motion on weighing platforms. Weight can be determined in a short time according to the temporary response of the measurement [6]. Measuring signal obtained from the dynamic weighing

*Corresponding Author: mehmetyumurtaci@aku.edu.tr
Received: 12.12.2019 Published: 28.12.2020

systems is distorting as load cells have an oscillatory response and there are vibrations in the system [7]. Measurement time can be much shorter than the period of system settlement [8]. Dynamic weighing systems are commonly used in different fields such as chemistry, medicine, pharmacology, food, transportation. These systems are designed and produced in different shapes and features as they are used in different fields. In general, they can be classified into two systems as "checkweigher" and "weigh-in-motion (WIM)". While checkweigher is generally used in industrial practices, WIM is used in transportation. Precision, time-saving, and the opportunity to weigh vehicles without disrupting traffic flow are some of the advantages of WIM systems [9]. A waiting period of vehicles is shortened by the use of these systems, which makes a significant economic contribution to drivers and weighing stations [10]. Halimic et al. (2008) combined fuzzy logic and Artificial Neural Network (ANN) in their study in order to filter the signals measured by the dynamic weighing system and they obtained fast and accurate measurement results [11]. Xiaoyan et al. (2008) thought that increase in velocity will decrease accuracy in dynamic weighing systems and designed a controller card by combining ANN and fuzzy logic methods in order to prevent accuracy problems in high velocity [12]. In their study, Bahar and Horrocks (1998) made weight estimation by ANN before the weighing system became stable [13]. In another study, Fuzzy Logic Estimator was used as a filter for the dynamic weighing system [14]. In their study, Almodarresi Yasin et al. presented a method in which feature extractor and ANN was used for dynamic weighing. As feature extractor was used before the ANN, neuron number of the network and calculation decreased complication [15].

Appearance and size are critical factors that are taken into consideration by consumers while determining main quality features. Geometrical features of eggs such as volume and surface area have a significant role in studies in the field of fowl industry and biology [16]. Image processing systems are also commonly used in the studies about egg weighing systems. In Ref. [17], a method that functions apart from the lightening level of the environment and the distance between the camera and the egg is suggested. A reference object, whose size is known, is placed next to an egg, whose weight is determined with an android device; after determining the size of the pixel, the size of the egg to be weighed is calculated and the weight class of the egg is determined by artificial intelligence. Alikhanov et al. carried out a study about weighing eggs indirectly; they suggested the use of an algorithm based on image processing with regression analysis in determining the relationship between geometric parameters such as perimeter, area, vertical and horizontal axis and volume [18]. In Ref. [19], through real-time image processing, the width and length of the egg are measured; after that, through the design and organization of ANFIS model, egg weight estimation is completed. Li Sun et al. suggested a method for evaluating the freshness of egg; the method that they suggested is based on a real-time process combining machine vision and dynamic weighing. They used machine vision and a dynamic weighing system in order to determine the physical characteristics of egg such as width, length, weight; they made estimation by creating a multivariable and linear model [20]. Omid et al. made a research to design and develop smart systems based on the combination of fuzzy logic and machine vision techniques; the ultimate goal of their study was to classify eggs through the use of different egg parameters such as size and defects [21].

In this study, weight data of eggs were collected from an egg production farm using a developed dynamic weighing system. Weight signal data of 4643 eggs from 4 different weight classes were collected at 100 Hz sampling frequency through LabVIEW software. The eggs were randomly chosen. A data set involving 1x376737 dimensional weights was obtained at the end of the process. The signals were combined in the first phase; the weight signal data of every single egg was separated with the code written in MATLAB. As weight data of each egg varies according to the size and process of passing from the load cell, data sizes were equalized with z-score normalization and outlier rejection methods. After that, TD, PSD, and Wavelet transforms were completed and three feature vector groups were formed. Feature vector groups were classified according to LBCN, FLDA, and SVM, DT, and k-NN classifiers; finally, weight classes were determined. As can be seen from the results, both feature vectors and classifiers are efficient in determining weight classes of eggs on the basis of the measuring signal in dynamic egg weighing systems.

2. PREPROCESSING AND FEATURE EXTRACTION

Data that were obtained from the dynamic measurement are required to be prepared before the classification process. This operation of preliminary preparation is called as preprocessing. Preprocessing includes the operations that make the raw data set to be suitable for being classified. Weights of 4643 eggs were measured via the dynamic measurement system and a raw data set is generated from these weights measurements. This generated raw data set consists of measurements from the sensors and also includes missing, inaccurate and inconsistent measurements. The raw data sets have been including a few data types that are lower the quality of data such that noise, outliers, missing values, duplicate data [22]. The generated data set includes errors due to wrong, inconsistent measurements and if this data set is used in classification operations directly, this would be affecting the results of these classification operations negatively. For this reason, to improve the quality and efficiency of and ease the classification operation, preprocessing has been applying to the raw data set.

The methods in preprocessing are cleaning, integration, normalization, and transformation [22]. Data cleaning is the process of extracting the noisy or inconsistent data from the raw data set. While data integration is unifying the data sets that are from more than one source; data transformation makes the data becoming more suitable for the used classification algorithm, this method provides more accurate data for the used algorithm. The other method is data normalization and this method unifies and scales the data.

The unnecessary, inconsistent and inaccurate data were found in the data set that is generated for this study and used. For this reason, to be overcome this unwanted situation data cleaning method that is one of the preprocessing methods was used. Data cleaning was applied via removing measurements of two eggs from the data set that is obtained from dynamic measurements due to the inconsistency of them. As a result of this process, the data set consists of 4641 egg measurements. In the dynamic measurement system, a different number of weight measurements were taken for each egg. However, in some of these weight measurements, the system has been showing the egg's weight zero. Therefore, the egg weights in the generated data set have been including unnecessary measurements. Applying the data cleaning process to the generated data set again, zero one from measurements, that were taken for each egg, has been removed. As a result of this process, for a total of 4641 eggs, a data set of at least 23 and most 47 measurements were obtained. In addition to that, the real weights of six different eggs were not detected so measurements of these eggs were removed and the generated data set includes measurements of 4635 different eggs finally.

2.1. Outlier Rejection

After removing egg data, which reduces the quality of the data set, when the measurements taken for each one egg were checked, it has been found that some of the measurements in this data set are different from what is expected. Like this, the measurements which behave differently than expected are called outliers. If outlier data are not excluded from the data set, they negatively would affect the result of the classification process. For this reason, it has to be necessary to detect all outliers and to remove all of them from the data set. The methods used to determine the data that are outliers; are separated as statistical, parametric, non-parametric, distance-based, clustering-based and neural network [23]. In this study, the z-Score method which is one of the parametric methods was used.

The z-Score value is a quantity calculated according to the proportion between the data and the whole data set's standard deviation and average. After calculating the mean (μ_{egg_i}) and standard deviation (σ_{egg_i}) using all measurements of an egg in the data set, z-score values have been calculated with the formula given in (1).

$$z\_Score_j = \frac{Measurement_{weight_j} - \mu_{egg_i}}{\sigma_{egg_i}} \quad where\ i = 1, \cdots, 4635$$

$$\mu_{egg_i} = \frac{1}{n} \sum_{j=1}^{n} Measurement_{weight_j}$$

$$\sigma_{egg_i} = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} Measurement_{weight_j}} \tag{1}$$

After calculating the z-Score values, a threshold value must be selected to determine which measurements are outliers. Z-Score values of measurements that belong to an egg for this process had been examined. As a result of this examination, the measurements whose z-Score values were smaller than zero were determined to be outliers for that egg (Figure 1). To summarize, measurements whose z-Score values are negative are outliers.

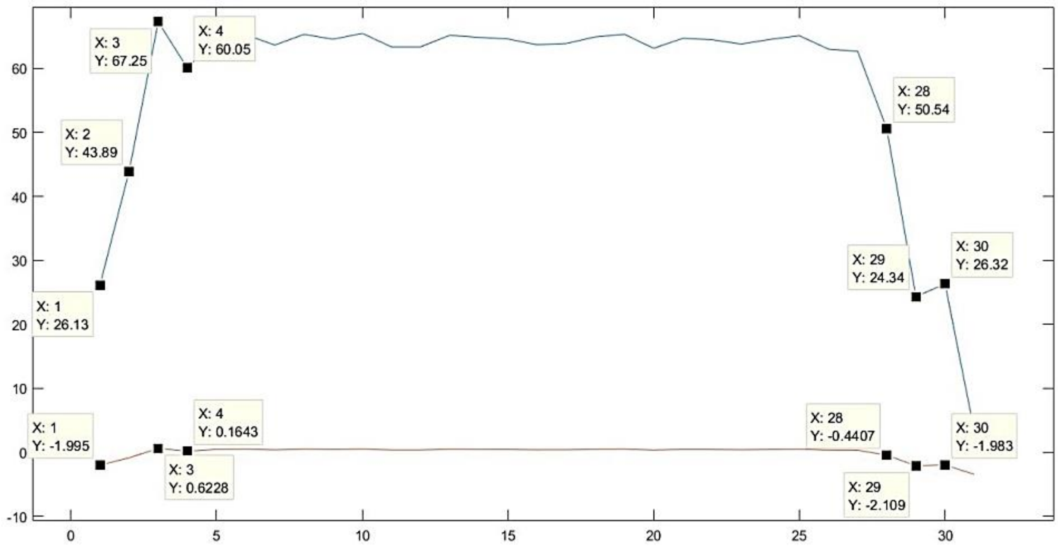

Figure 1. Threshold selection from z-score values of an egg

Implementation of outlier rejection to all measurements, which were taken for whole eggs, has been obtained by extracting measurements that have negative z-Score values.

2.2 Normalization

After the outlier rejection process, normalization that is one of the preprocessing methods was applied to the data set. Since the normalization is the scaling of the data between 0-1, the 2-norm method was used in this study. In this method, all measurements of an egg via the formula given in (2) have been used to calculate the L2-norm value and then all measurement values have been divided to this calculated value.

$$\| Measurement_{weight_{egg_i}} \|_2 = \sqrt{\left[\sum_{j=1}^{n} \left| Measurement_{weight_j} \right|^2 \right]}$$

$$Normalized\ Measurement_{weight_j} = \frac{Measurement_{weight_j}}{\| Measurement_{weight_{egg_i}} \|_2}$$

$$where\ j = 1, \cdots, \#\ of\ measurements \tag{2}$$

The data set, which is obtained after processes of data cleaning, outlier rejection, and normalization, has no wrong and outlier data, and also all measurements for each egg in this data set have been normalized.

2.3 Feature Extraction

Feature extraction can be defined as expressing raw data as attributes (feature) or converting them into attributes. Instead of reducing the size of the data set, whose dimension is still too large after preprocessing, and using the data as measurements only, the use of this measurement information in the classification by being converted it into new attributes provides a reduction in computational cost [24].

In this study, a total of 30 features were extracted for 4635 eggs. These extracted features are TD, PSD-based, and DWT based features.

2.3.1 TD features

TD features are energy, mean, standard deviation, maximum, minimum, skewness, kurtosis, difference, and entropy. The formula of each feature is shown in Table 1. The first TD feature, Energy, is the sum of the squares of all normalized measurements taken for an egg. For the second feature, all the normalized measurements taken for an egg are summed and the summation is divided by the number of measurements taken for that egg, thus the average is calculated. For the third feature, the mean value of the normalized measurements is subtracted from normalized measurements; the squares of them are taken and summed. The summation is divided by the number of measurements and the square root of it is calculated, so the standard deviation is obtained. The fourth and fifth features are the maximum and minimum values of the normalized measurements respectively. Skewness, which is a measure of symmetry, is the degree of distortion that measures the lack of symmetry according to normal distribution or the symmetric bell curve. Kurtosis, according to the normal distribution of the data determines that the tail is heavy or light-tailed. The eighth feature, Difference, is the difference between the maximum and minimum normalized measurements. Entropy, which is the last feature in the time domain, is a measure of disorder or unpredictability in a system [25].

Table 1.TD features

Feature	Abbv.	Equation
Energy	TD_1	$\sum_{j=1}^{n}\left(Normalized\ Measurement_{weight_j}\right)^2$
Mean	TD_2	$\frac{\sum_{j=1}^{n} Normalized\ Measurement_{weight_j}}{n}$
Standard Deviation	TD_3	$\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(Normalized\ Measurement_{weight_j} - TD_2\right)^2}$
Maximum	TD_4	$\max\{Normalized\ Measurement_{weight_j}, 1 \le j \le n\}$
Minimum	TD_5	$\min\{Normalized\ Measurement_{weight_j}, 1 \le j \le n\}$
Skewness	TD_6	$\frac{\sum_{j=1}^{n}\left(Normalized\ Measurement_{weight_j} - TD_2\right)^3}{(n-1)TD_3^3}$
Kurtosis	TD_7	$\frac{\sum_{j=1}^{n}\left(Normalized\ Measurement_{weight_j} - TD_2\right)^4}{(n-1)TD_3^4}$
Difference	TD_8	$TD_4 - TD_5$
Entropy	TD_9	$\sum_{j=1}^{n}\log_2(Normalized\ Measurement_{weight_j}^2)$ & $if \log_2\left(Normalized\ Measurement_{weight_j}^2\right) = \infty,$ $then \log_2\left(Normalized\ Measurement_{weight_j}^2\right) = 0$

2.3.2. PSD based features

PSD based features were obtained by using the "pwelch.m" function on the MATLAB platform. With the help of this function, the PSD estimate value of each egg was calculated by Welch's method. This function generates the PSD estimate values by separating the time series data into segments, calculating the modified periodogram (which is the estimation of the spectral density of a signal) for each segment and averaging it [26]. Using these generated PSD estimate values, 7 features are extracted and these extracted features are shown in Table 2. These features are the same as the TD features (except for Entropy and Difference) and the PSD estimate values of the measurements have been used instead of the eggs' normalized weight measurements.

Table 2. PSD-based features

Feature	Abbv.	Equation
Energy	PSD_1	$\sum_{j=1}^{n}\left(PSD_{weight_j}\right)^2$
Mean	PSD_2	$\frac{\sum_{j=1}^{n} PSD_{weight_j}}{n}$
Standard Deviation	PSD_3	$\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(PSD_{weight_j} - PSD_2\right)^2}$
Maximum	PSD_4	$\max\{PSD_{weight_j}, 1 \le j \le n\}$
Minimum	PSD_5	$\min\{PSD_{weight_j}, 1 \le j \le n\}$
Skewness	PSD_6	$\frac{\sum_{j=1}^{n}\left(PSD_{weight_j} - PSD_2\right)^3}{(n-1)PSD_3^3}$
Kurtosis	PSD_7	$\frac{\sum_{j=1}^{n}\left(PSD_{weight_j} - PSD_2\right)^4}{(n-1)PSD_3^4}$

2.3.3. DWT based features

Wavelets are mathematical functions that separate data into different frequency components. Although the Wavelet transform method is similar to Fourier Transform (FT), however, it provides information about both frequency and localization in time-space instead of frequency information as in FT. Wavelet transforms are separated as continuous and discrete-time. The DWT decomposes the signal into mutually orthogonal wavelet sets in the transformation process. There are low and high-frequency components of the signal in this decomposition. If the decomposition process is more than one level, there are frequency components according to the pyramid algorithm.

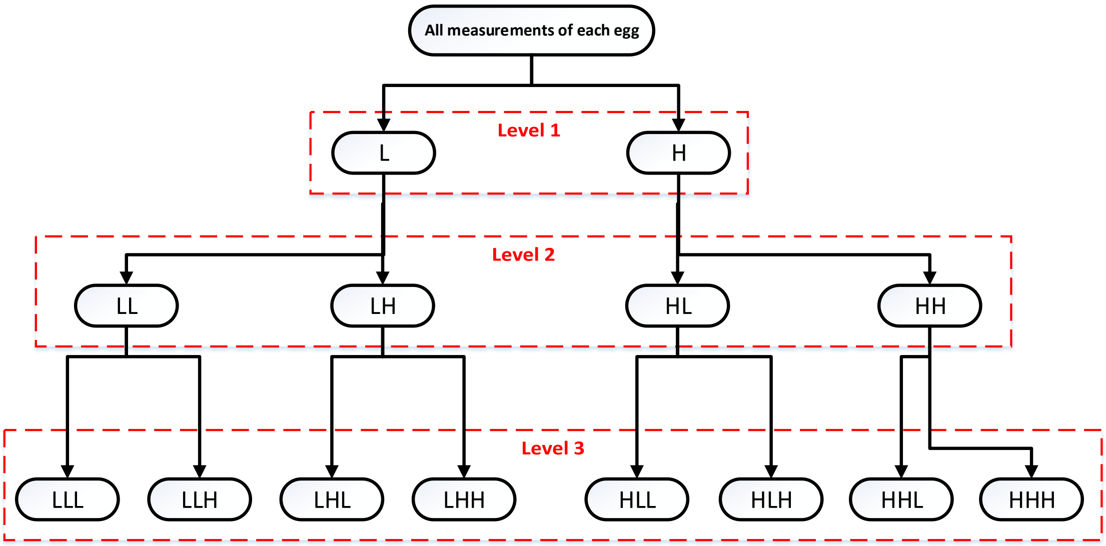

Figure 2. 3-Level decomposition of DWT

In this study, 3-level decomposition was performed as shown in Figure 2. For the decomposition process, "wpdec.m" function in the MATLAB platform was used [27] and one of the most known orthogonal wavelet sets named as Daubechies 20 was used. After each decomposition, wavelet packet coefficients are formed with the "wprcoef.m" [27] that is in the MATLAB platform, and entropy of these calculated coefficients are considered as features. DWT based features are shown in Table 3.

Table 3. DWT based features

Wavelet Packet Components	Abbv.	Feature	Abbv.
Low-Frequency Components	W_{1_1}	Entropy of W_{1_1}	$Ent_{W_{1_1}}$
High-Frequency Components	W_{1_2}	Entropy of W_{1_2}	$Ent_{W_{1_2}}$
Low-Low Frequency Components	W_{2_1}	Entropy of W_{2_1}	$Ent_{W_{2_1}}$
Low-High Frequency Components	W_{2_2}	Entropy of W_{2_2}	$Ent_{W_{2_2}}$
High-Low Frequency Components	W_{2_3}	Entropy of W_{2_3}	$Ent_{W_{2_3}}$
High-High Frequency Components	W_{2_4}	Entropy of W_{2_4}	$Ent_{W_{2_4}}$
Low-Low-Low Frequency Components	W_{3_1}	Entropy of W_{3_1}	$Ent_{W_{3_1}}$
Low-Low-High Frequency Components	W_{3_2}	Entropy of W_{3_2}	$Ent_{W_{3_2}}$
Low-High-Low Frequency Components	W_{3_3}	Entropy of W_{3_3}	$Ent_{W_{3_3}}$
Low-High-High Frequency Components	W_{3_4}	Entropy of W_{3_4}	$Ent_{W_{3_4}}$
High-Low-Low Frequency Components	W_{3_5}	Entropy of W_{3_5}	$Ent_{W_{3_5}}$
High-Low-High Frequency Components	W_{3_6}	Entropy of W_{3_6}	$Ent_{W_{3_6}}$
High-High-Low Frequency Components	W_{3_7}	Entropy of W_{3_7}	$Ent_{W_{3_7}}$
High-High-High Frequency Components	W_{3_8}	Entropy of W_{3_8}	$Ent_{W_{3_8}}$

3. CLASSIFIERS

3.1. LBNC

LBNC is a linear classifier and has been assuming that the covariance matrices of the classes to be classed are equal and their density is normal [28]. By using the average of covariance matrices of the classes, the joint covariance matrix is weighted and is regulated according to the R or S parameters expressed in PRtools. These parameters were used to find a new covariance matrix. In fact, small variance directions are eliminated due to the regulation process and then a new covariance matrix is calculated via the leading main components and the smallest eigenvalues [29].

3.2. FLDA

FLDA is used as a linear classifier optimizing class separation [25] and it is also utilized for size reduction operations. When used as a classifier, it projects the data first. The purpose of this process is to maximize the distance between the averages of the classes and to minimize the covariance within the class. After the projection process, it categorizes the classes by finding linear combinations, which include large ratios, between within-class and between-class distributions. Fisher criterion is used to find this ratio and so this ratio is obtained from the multiplication of the inverse of within-class and between-class covariance matrices. As a result of this process, the covariance matrix is obtained and the difference between classes is maximized by this calculated matrix.

3.3. SVM

SVM method fundamentally involves selecting a plane that will separate the classes. It is mapping data to a high-dimensional area and by finding a hyperplane with a maximum margin separates these classes [30]. The classification process is carried out by the hyperplane. This plane is selected to be on the largest distance to the nearest data sample of any classes to be separated. This distance is called functional margin and this margin may be linear or non-linear. When creating a linear margin, SVM finds the optimal hyperplane by maximizing function margin, in other words, maximizing the perpendicular distance between the nearest samples that belong to different classes. For the non-linear ones, it maps another space via non-linear transform and linearizes it or selects the hyperplane that gives the least error.

3.4. DT

DT is a classifier in the form of a hierarchical tree structure using the divide and seizes strategy [25]. The node (root) to which all nodes in this tree structure are connected is called the root node and the classification process is started from this node. In addition to this, there are decision nodes, branches, and leaf nodes in the tree structure. While the decision nodes contain the conditions that test the features, the branches contain the possible responses to these conditions and the leaf nodes contain the class information. In the working principle of the DT; for the data, whose class will be assigned, there is searching between the nodes until the root node reaches the leaf node. This search process is done through decision nodes. As the decision nodes contain rules in the If-Then-Else structure, the decision is made according to the answers to these rules and also the next node is determined by these answers. The conditions presented in these nodes test a specific feature and the classifier follows a branch if the condition is met, and another branch if the condition is not met. The process of following the nodes terminates when a node reaches the leaf node and the class information is assigned to that data.

3.5. k-NN

k-NN is one of the similarity-based methods, and a "lazy learner" classification algorithm. Within this algorithm, the data sample, whose class is unknown is classified by the ratio of similarities between it and some memorized samples [31]. It uses different distance measurements as similarity ratios and also it is more suitable for classification of numerical data [30], [31], [32]. This algorithm basically determines the nearest K neighbors of the data sample whose class will be assigned and detect the most common class label among these neighbors and then assigns it to that sample [33]. The determination of the K value depends on the user, but if this value is set to one, the sample to be assigned to a class is assigned to the class of the nearest neighbor.

4. EXPERIMENTAL WORK

Dynamic weighing systems are designed to weigh circular products (such as egg, orange, etc.). The system is made of two main parts as hardware and software. The hardware system involves a mechanic weighing system and electronic control board; a software system is made of three components. The first component is electronic card software which transforms the data obtained from load cell into weight data. The second software component is interface software which is prepared in the LabVIEW environment, in which the egg weight data are observed and recorded. The third component software is designed to process the obtained egg weight data, feature vectors are extracted and egg weight is determined through classifiers. The flow chart of the experimental work is presented in Figure 3.

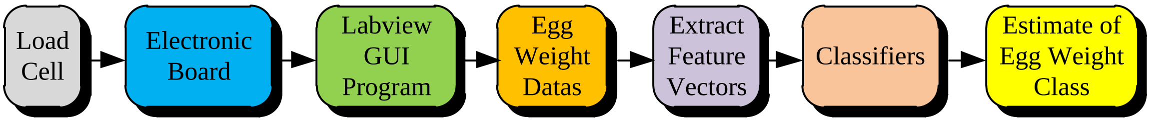

Figure 3. Flow chart of experimental work

4.1. Mechanical Weighing System

Mechanic details of the dynamic weighing system are presented in Figure 4 in detail. Products to be weighed roll on an inclined surface through carrier pipes. When they reach the load cell platform located horizontally, they are separated from the carrier pipes without any decrease in speed. Products pass from the load cell platform and they are weighed without being exposed to any force that may affect the measurement process. As products are weighed without stopping on the load cell platform, necessary weighing speed in industrial processes can be obtained [34].

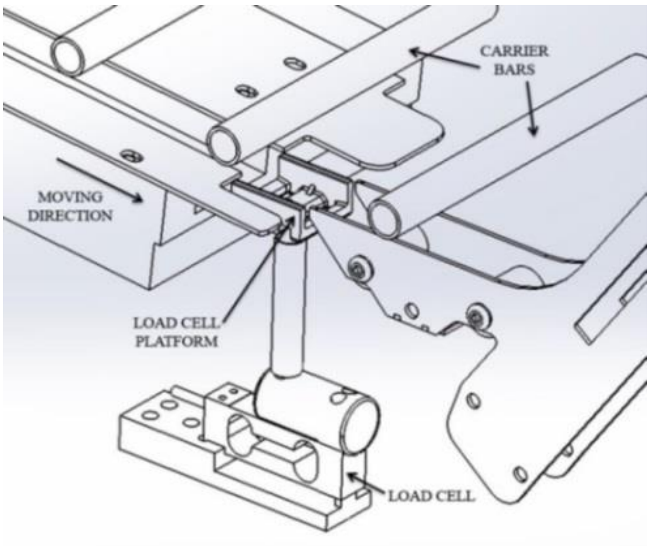

Figure 4. Mechanical weighing system [34]

4.2 Electronic Control Board

The electronic control board is made of basic electronic circuit components; mainly 24-bit Sigma-Delta ADC and a microcontroller. The electronic control board is designed to transform the measuring signal obtained from the load cell into a digital signal. The content of the board is presented in Figure 5. Analogue weight data transformed into the digital signal is transmitted to the DSP microcontroller through SPI serial communication interface. DSP microcontroller is designed to make necessary adjustments of Sigma-Delta ADC and to send the digital weight data from ADC to the interface designed in the LabVIEW software through the CAN network according to the desired sampling frequency.

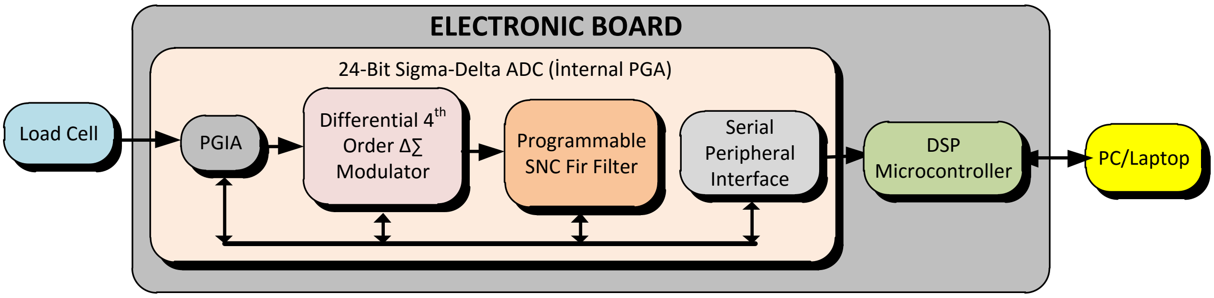

Figure 5. Dynamic weighing system block scheme

4.3. Data Acquisition Program

An interface is created in the LabVIEW program in order to graphically observe the changes in digital weight data sent from the DSP microcontroller through the CAN network and to save the data into a file. It is possible to observe immediate digital weight data and immediate weight data in terms of grams. The data observed at this point is sent from the designed interface when the load cell is empty and full. The data is a 24-bit resolution. It is also possible to graphically draw the change of these data according to time. The data collection program interface is presented in Figure 6. Data is saved into the desired file by noting the name and location of the file and pressing the save button. As data is recorded in a file, it is possible to make a detailed analysis with the desired method.

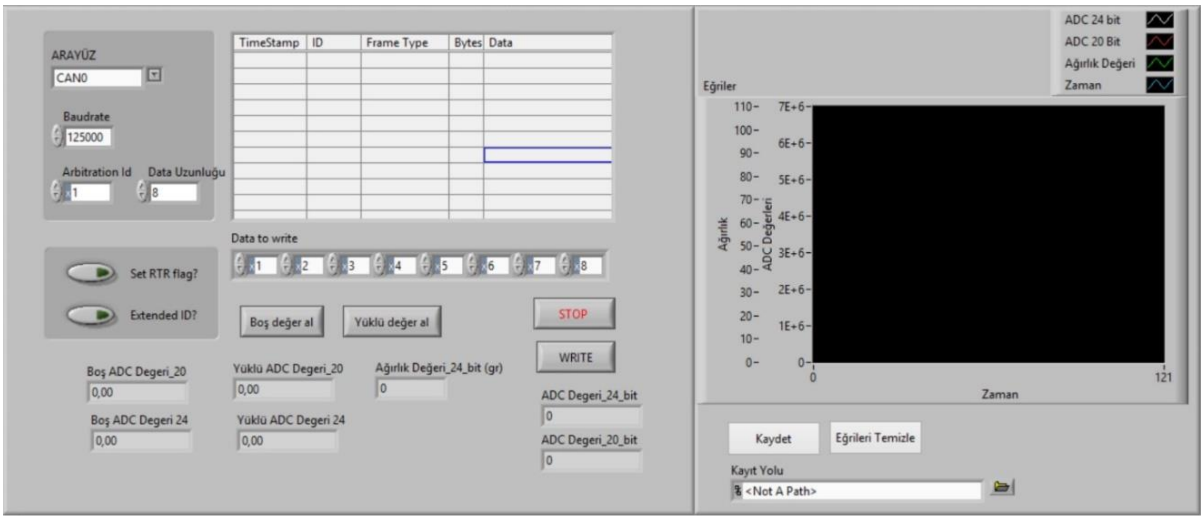

Figure 6. Data collection program interface

4.4. Database

According to the Ministry of Food, Agriculture, and Livestock Turkish Food Codex Eggs Communique, eggs are divided into four classes according to their weight. The class labels we determined in order to be used in classification are presented in Table 4 together with the classes of eggs according to their weight [35].

Table 4. Egg class label and weight variations

Class Label	**1**	**2**	**3**	**4**
Weight Class	XL – Extra Large	L - Large	M - Middle	S - Small
Weight Range	≥73 gr	≥63 – <73 gr	≥ 53 – <63 gr	<53 gr

The data set to be used in this study was created by the writers. Data were obtained from a poultry farm in which dynamic weighing system has been used; the data were collected at 100 Hz sampling frequency through LabVIEW software and electronic board. Randomly collected data is presented in Table 5 and it involves weight signal data of 4635 eggs from 4 weight classes. Change in egg weight data obtained from load cell is presented in Figure 7.

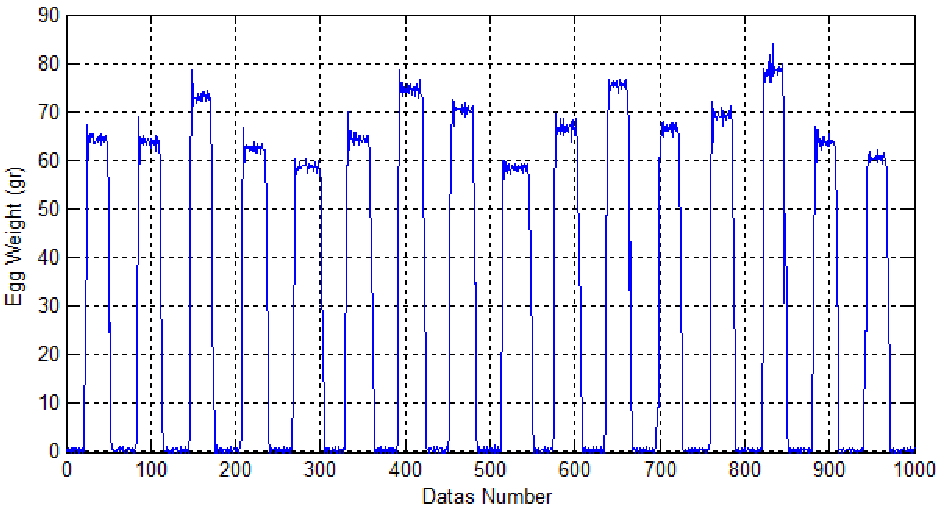

Figure 7. Change in egg weight data

The signals were combined in the first phase; the weight signal data of every single egg was separated with the code written in MATLAB. The weight data of each egg varied according to its size and process of passing from the load cell; variation is presented in Figure 8. Score normalization and outlier rejection were used in order to equalize the size of the separated egg data.

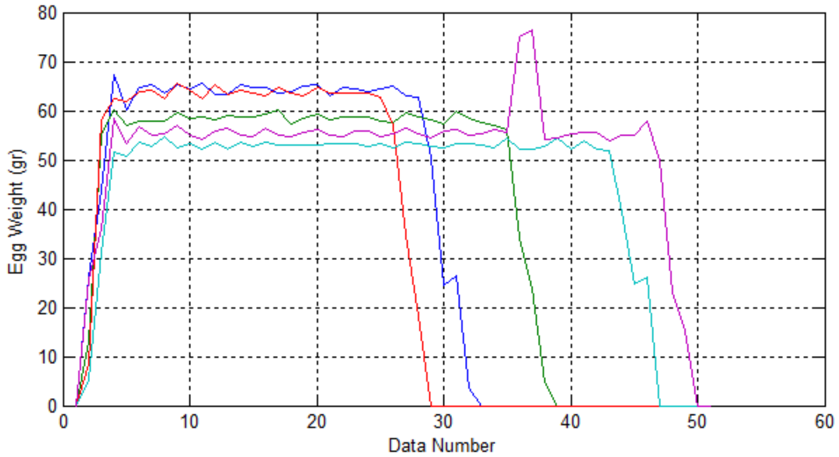

Figure 8. Egg weight data of eggs with different weight

The amount of data used by the classifiers for training and test processes is presented in detail in Table 5.

Table 5. The data amount used by the classifiers for education and test processes per class

Class Labels	1	2	3	4
Train	60	1012	2080	556
Test	15	253	520	139
Total	75	1265	2600	695

Feature vector extraction was applied to the obtained data set; TD, PSD, and DWT methods were used in the process. The extracted feature vectors are presented together in Table 6.

Table 6. Extracted feature vectors

Number	TD Features	PSD Features	Entropy of eachWavelet Features
1	Energy	Energy	Entropy of Level 1 L
2	Mean	Mean	Entropy of Level 1 H
3	Standard Deviation	Standard Deviation	Entropy of Level 2 LL
4	Max	Max	Entropy of Level 2 LH
5	Min	Min	Entropy of Level 2 HL
6	Skewness	Skewness	Entropy of Level 2 HH
7	Kurtosis	Kurtosis	Entropy of Level 3 LLL
8	Difference		Entropy of Level 3 LLH
9	Entropy		Entropy of Level 3 LHL
10			Entropy of Level 3 LHH
1			Entropy of Level 3 HLL
1			Entropy of Level 3 HLH
12			Entropy of Level 3 HHL
13			Entropy of Level 3 HHH
14			Entropy of Level 3 LLL

5. RESULTS

After feature extraction, feature vectors in three different sizes were obtained. 5-fold cross-validation was applied to the data in order to determine the success of classification made by classifiers and feature vectors. Classification experiments are performed at a computer that has 16 GB RAM and Intel Core i7-4790K/ 4.00 GHz processor. LBNC, FLDA, SVM, DT and k-NN average success results and the average elapsed times for test phases are presented in Table 7 and Table 8, respectively.

Table 7. Classification results of classifiers

Classifies	TD Fea.	PSD Fea.	Entropy of DWT Fea.
DT	0,9456	0,9561	0,9638
FLDA	0,9126	0,9127	0,9124
LBNC	0,9125	0,9125	0,9129
k-NN	0,9692	0,9688	0,9701
SVM	0,9126	0,9124	0,9135

Table 8. Elapsed times for test phase of classifiers (second)

Classifies	TD Fea.	PSD Fea.	Entropy of DWT Fea.
DT	0,3646	0,3858	0,3055
FLDA	0,0856	0,0857	0,0992
LBNC	0,1281	0,1270	0,1283
k-NN	0,3011	0,2948	0,2647
SVM	0,3662	0,3533	0,3206

Five different classifiers have similar classification success according to the Cross-validation process. This result indicates the coherence of data in the data set and shows the performance of classifiers. The highest classification success was obtained from the k-NN classifier. Apparently, both extracted feature vectors and the classifiers used in the classification process are suitable for determining the real weight class of eggs through the dynamic egg weighing system.

When the literature studies are examined, studies on egg classification are mainly based on image processing. In this study, the classification process is based on signal processing. Waranusast et al. (2016) divided the eggs into six classes according to their weights: jumbo, extra large, large, middle, small, pewee. By processing the image taken from the mobile device, they made classification with SVM and achieved 80.4% accuracy [17]. Asadi et al. (2010) estimated the weight of the egg with 95% accuracy using the machine vision technique [1]. Omid et al. (2013) have developed an intelligent system based on a combination of machine vision and fuzzy logic techniques to classify eggs into small, medium, and large eggs according to their size [21]. The classification success rate was determined as 95%. In this study, according to the Ministry of Food, Agriculture, and Livestock Turkish Food Codex Eggs Communique, egg weights were divided into four classes as extra large, large, middle, and small. Egg weight signal data were processed with different methods and weight classes were determined with various classifiers. With the entropy of DWT feature vector and k-NN classifier, the highest performance rate was found to be 97.1%. With this method, the weight class of an egg can be detected in an average of 0.2647 seconds.

6. CONCLUSION

Egg is one of the very nutritious and beneficial foods for humans and it is not only consumed alone but also used in a variety of convenience foods. Almost completely of the eggs are presented and sold with respect to their weights in almost all countries. Therefore, the impeccable measurement of egg weights is so much crucial before selling them in different market points. In this paper, features calculated from the measurement of egg weights by three different feature extraction techniques, and their performances are compared with each other. Their respective successes on egg classification are separately evaluated by five different novel and well-known classifiers. If one categorizes these three feature types, one of them is based on TD where the other two are based on frequency-domain features. The entropy values of DWT subbands are the most successful features especially for SVM, LDA, DT, and k-NN classifiers since they present the importantly preserved information within discrete-time signals obtained from weight measurements and reflect the inherent nature of data classes. Besides, the k-NN classifier gives the highest recognition accuracy among all classifiers. The reason for this confounding success of the k-NN classifier is directly related to the number of classes. If a few numbers of classes are treated in a traditional pattern recognition problem, the taxonomy of data may be generally driven with an easy manner. On the contrary, a more elaborate analysis is needed using more exhaustive classifiers for the cases in which high number of classes is involved. Therefore, the 97.01% accuracy achievement is not a surprising outcome for the k-NN classifier in this study. In addition, its computational cost is much low when a comparison is made with respect to the computational complexity issue. Besides this, the discriminative power of features (especially for entropy of DWT-based ones) extracted in the data of this study are very effective for obviously separating classes so that the k-NN classifier utilizes from this advantageous circumstance compared to the other more complex classifiers such as FLDA and SVM. In future work, the classification of eggs in terms of various aspects such as albumen ratio and eggshell thickness using hyperspectral imaging techniques is planned.

REFERENCES

[1] Asadi V, Raoufat MH, Nassiri SM. Multiple regression analysis results of EGG weight estimation using machine vision technique. In: International Conference on Agricultural Engineering - AgEng 2010: towards environmental Technologies; 6-8 September 2010; Clermont-Ferrand, France, 1-10.

[2] Asadi V, Raoufat MH, Nassiri SM. Fresh egg mass estimation using machine vision technique, Int. Agrophys. 2012; 26(3): 229-234.

[3] Asadi V, Raoufat M.H, Estimation of egg weight by machine vision and neural networks technique. International Journal Natural and Engineering Sciences (IJNES) 2010; 4(2): 1-4.

[4] King'ori AM. Poultry egg external characteristics: egg weight, shape and shell colour. Res. J. Poultry Sci. 2012; 5(2): 14-17.

[5] Rashidi M, Gholami M. Prediction of egg mass based on geometrical attributes. Agr. Biol. J. N. Am. 2011; 2(4): 638-644.

[6] Niedźwiecki M, Meller M, Pietrzak P. System identification based approach to dynamic weighing revisited. Mechanical Systems and Signal Processing 2016; 80: 582–599.

[7] Boschetti G, Caracciolo R, Richiedei D, Trevisani A. Model-based dynamic compensation of load cell response in weighing machines affected by environmental vibrations. Mechanical Systems and Signal Processing 2013; 34(1- 2): 116-130.

[8] Yamazaki T, Ono T. Dynamic problems in measurement of mass-related quantities. In: SICE, 2007 Annual Conference; 17-20 Sept. 2007; Takamatsu, Japan, 1183-1188.

[9] Rui Z, Wen-hong L, Yin-jing G. A vehicle weigh-in-motion system based on hopfield neural network adaptive filter. In: 3rd International Communication and Mobile Computing Conference; 12-14 April 2010; Shenzhen, China Shenzhen, 123-127.

[10] Xiao J, Lv P. Application of wavelet transform in weigh-in-motion. In: International Workshop on Intelligent Systems and Applications; 23-24 May 2009; Wuhan, China, 1-4.

[11] Halimic M, Halimic A, Zugail S, Huneitti Z. Intelligent signal processing for electro-mechanical systems. In: Proceeding of the 5th International Symposium on Mechatronics and its Applications; 27-29 May 2008; 1-5.

[12] Xiaoyan C, Zhenliang L. An intelligent dynamic weighing controller. In: IEEE International Conference on Automation and Logistics; 27-29 May 2008; 1609-1612.

[13] Bahar H.B, Horrocks DH. Dynamic weight estimation using an artificial neural network. Artif. Intell. Eng. 1998; 12(1-2): 135-139.

[14] Halimic M, Balachandran W, Enab Y. Fuzzy logic estimator for dynamic weighing system. In: Proceedings of IEEE 5th International Fuzzy Systems; 11 Sept. 1996; New Orleans, LA, USA, 2123-2129.

[15] Yasin SMTA, White NM. Application of artificial neural networks to intelligent weighing systems. IEE Proceedings - Science, Measurement and Technology 1999; 146(6): 265-269. [16]

[16] Soltani M, Omid M, Alimardani R. Egg volume prediction using machine vision technique based on pappus theorem and artificial neural network. J. Food Sci. Technol. 2015; 52(5): 3065–3071.

[17] Waranusast R, Intayod P, Makhod D. Egg size classification on android mobile devices using image processing and machine learning. In: 2016 Fifth ICT International Student Project Conference (ICT-ISPC); 27-28 May 2016; Nakhon Pathom, Thailand, 170-173.

[18] Alikhanov D, Penchev S, Georgieva Ts, Moldajanov A, Shynybaj Z, Daskalov P. Indirect method for egg weight measurement using image processing. International Journal of Emerging Technology and Advanced Engineeering 2015; 5(11): 30-34.

[19] Javadikia P, Dehrouyeh MH, Naderloo L, Rabbani H, Lorestani AN. Measuring the weight of egg with image processing and ANFIS model. In: SEMCCO 2011; 19-21 December; Visakhapatnam, India; 407–416.

[20] Sun L, Yuan L, Cai J, Lin H, Zhao J. Egg freshness on-line estimation using machine vision and dynamic weighing. Food Anal. Methods 2015; 8: 922–928.

[21] Omid M, Soltani M, Dehrouyeh MH, Mohtasebi SS, Ahmadi H. An expert egg grading system based on machine vision and artificial intelligence techniques. J. Food Eng. 2013; 118: 70–77.

[22] Garcia S, Luengo J, Herrera F. Data preprocessing in data mining. Springer, Intelligent Systems Reference Library 2015; 72.

[23] Malik K, Sadawarti H, Kalra G. Comparative analysis of outlier detection techniques. Int. J. Computer Appl. 2014; 97: 12–21.

[24] Hira, ZM, Gillies DF. A review of feature selection and feature extraction methods applied on microarray data. Adv. Bioinformatics 2015; 2015: 1-13.

[25] Dougherty G. Pattern Recognition and Classification. Springer New York, 2013.

[26] Welch's power spectral density estimate. Available at: https://www.mathworks.com/help/signal/ref/pwelch.html; Accessed: 21-Apr.-2019.

[27] Reconstruct wavelet packet coefficients. Available at: https://www.mathworks.com/help/wavelet/ref/wprcoef.html; Accessed: 21-Apr.-2019.

[28] PRTools A MATLAB toolbox for pattern recognition. Available at: http://prtools.tudelft.nl/; Accessed: 21-Apr.-2019.

[29] Elibol G, Ergin S. The assessment of time-domain features for detecting symptoms of diabetic retinopathy", International Journal of Intelligent Systems and Applications in Engineering 2016; 4: 136-140.

[30] Guyon I, Gunn S, Nikravesh M, Zadeh LA. Feature Extraction Foundations and Applications. Springer -Verlag Berlin Heidelberg, 2006.

[31] Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans. Inform. Theor. 1967; 13(1): 21-27.

[32] Bolon-Canedo V, Sanchez-Marono N, Alonso-Betanzos A. Feature Selection for High-dimensional Data. Springer International Publishing, 2015.

[33] Bishop CM. Pattern Recognition and Machine Learning. Springer, 2006.

[34] Yabanova İ. Dynamic mass measurement and appropriate filter analysis. IU-JEEE 2016; 16(2): 3033-3036.

[35] Türk Gıda Kodeksi Yumurta Tebliği, Gıda Tarım Hayvancılık Bakanlığı, Resmi Gazete 2014, Sayı: 29211.