

BS Journals

Volume: 3

Issue: 4

Black Sea Journal of Engineering and Science



ISSN: 2619-8991



BLACK SEA JOURNAL OF ENGINEERING AND SCIENCE
(BSJ ENGIN SCI)



Black Sea Journal of Engineering and Science (BSJ Eng Sci) is a double-blind peer-reviewed, open-access international journal published electronically 4 times (January, April, July and October) in a year by since January 2018. It publishes, in English and Turkish, full-length original research articles, innovative papers, conference papers, reviews, mini-reviews, rapid communications or technical note on advances in a wide range of scientific disciplines from all fields of engineering and science and from any source.

ISSN 2619 - 8991

Phone: +90 362 408 25 15

Fax: +90 362 408 25 15

Email: bsjsci@blackseapublishers.com

Web site: <http://dergipark.gov.tr/bsengineering>

Sort of publication: Periodically 4 times (January, April, July and October) in a year

Publication date and place: October 01, 2020 - Samsun, TURKEY

Publishing kind: Electronically

OWNER

Assoc. Prof. Uğur ŞEN

DIRECTOR IN CHARGE

Prof. Dr. Hasan ÖNDER

EDITOR BOARDS

EDITOR IN CHIEF

Prof. Dr. Hasan ÖNDER
Assoc. Prof. Uğur ŞEN

Ondokuz Mayıs University, TURKEY
Ondokuz Mayıs University, TURKEY

SECTION EDITORS*

Prof. Dr. Erkan YALÇIN, Ondokuz Mayıs University, TURKEY
Prof. Dr. Hasan TANAK, Amasya University, TURKEY
Prof. Dr. Hasan TEMİZ, Ondokuz Mayıs University, TURKEY
Prof. Dr. Mustafa ŞAHİN, Kahramanmaraş Sutcu Imam University, TURKEY
Assoc. Prof. Dr. Esmeray KÜLEY BOĞA, Cukurova University, TURKEY
Assoc. Prof. Dr. Kadyrbay CHEKİROV, Kyrgyz Turkish Manas University, KYRGYZSTAN
Assoc. Prof. Dr. Özgür Hakan AYDOĞMUŞ, Social Sciences University of Ankara, TURKEY
Assoc. Prof. Dr. Roswanira Binti AB WAHAB, University of Technology, MALAYSIA
Asst. Prof. Dr. Fatma Sevinç KURNAZ, Yıldız Technical University, TURKEY
Asst. Prof. Dr. Sedat KARADAVUT, Trakya University, TURKEY

* The ranking is arranged alphabetically within the academic title

EDITORIAL - ADVISORY BOARD*

Prof. Dr. Aglaia (Litsa) LIOPA-TSAKALIDI, Institute of Western Greece, GREECE
Prof. Dr. Ercan EFE, Kahramanmaraş Sutcu Imam University, TURKEY
Prof. Dr. Mohammad Masood TARIQ, University of Balochistan, PAKISTAN
Prof. Dr. Mustafa Çağatay TUFAN, Ondokuz Mayıs University, TURKEY
Prof. Dr. Özkan GÖRGÜLÜ, Ahi Evran University, TURKEY
Prof. Dr. Soner ÇANKAYA, Ondokuz Mayıs University, TURKEY
Assoc. Prof. Dr. Taner TUNÇ, Ondokuz Mayıs University, TURKEY
Asst. Prof. Dr. Emil OMURZAK, Kyrgyz-Turkish Manas University, KYRGYZSTAN
Asst. Prof. Dr. Yılmaz KAYA, Ondokuz Mayıs University, TURKEY

* The ranking is arranged alphabetically within the academic title

STATISTIC EDITOR

Prof. Dr. Mehmet TOPAL, Kastamonu University, Turkey

ENGLISH EDITOR

Assoc. Prof. Dr. Sezen Ocak YETISGIN, Ondokuz Mayıs University, TURKEY

TURKISH EDITOR

Prof. Dr. Serkan ŞEN, Ondokuz Mayıs University, TURKEY

REVIEWERS OF THE ISSUE*

Prof. Dr. Fulya ÖZDİL, Tekirdağ Namık Kemal University, Turkey

Prof. Dr. Nalan TÜRKOĞLU, Van Yüzüncü Yıl University, Turkey

Prof. Dr. Serap GÖNCÜ, Çukurova University, Turkey

Assoc. Prof. Dr. Arzu ÇIĞ, Siirt University, Turkey

Assoc. Prof. Dr. Bahri Devrim ÖZCAN, Çukurova University, Turkey

Assoc. Prof. Dr. Mehmet Serdar GÜZEL, Ankara University, Turkey

Assoc. Prof. Dr. Tolga YUKSEL, Bilecik Şeyh Edebali University, Turkey

Asst. Prof. Dr. Aycan Mutlu YAĞANOĞLU, Atatürk University, Turkey

Asst. Prof. Dr. Emre ALARSLAN, Bandırma Sheep Research Institute, Turkey

Asst. Prof. Dr. Erhan ERGÜN, Ondokuz Mayıs University, Turkey

Asst. Prof. Dr. Gökhan KAYHAN, Ondokuz Mayıs University, Turkey

Asst. Prof. Dr. Hüseyin İrfan BALIK, Sakarya Uygulamalı Bilimler University, Turkey

Asst. Prof. Dr. Mikail KARA, Çankırı Karatekin University, Turkey

Asst. Prof. Dr. Nezihe KÖKSAL, Çukurova University, Turkey

Asst. Prof. Dr. Nurettin ŞENYER, Ondokuz Mayıs University, Turkey

Dr. Ebru ERGÜNEŞ BERKİN, İslahiye Directorate of District Agriculture and Forestry, Turkey





Dr. Thobela Louis TYASI, University of Limpopo, South Africa

* The ranking is arranged alphabetically within the academic title




Volume 3, Issue 4 (2020)

Table of Contents

Research Article

- 1. A COMPARATIVE STUDY FOR INDOOR PLANAR SURFACE SEGMENTATION VIA 3D LASER POINT CLOUD DATA** 
(3D lazer noktası bulut verileri ile iç mekan planar yüzey segmentasyonu için karşılaştırmalı bir çalışma)
Eyüp Eymen ERUYAR, Metehan YILMAZ, Berat YILMAZ, Onur AKBULUT, Kaya TURGUT, Burak KALECİ.....128-137
- 2. ESTIMATING OF BIRTH WEIGHT USING PLACENTAL CHARACTERISTICS IN THE PRESENCE OF MULTICOLLINEARITY** 
(Çoklu bağlantı durumunda plasenta özelliklerini kullanarak doğum ağırlığının tahmini)
Cem TIRINK.....138-145
- 3. EFFECTS OF SOME ENVIRONMENTAL FACTORS ON THE SOME GROWTH AND LACTATION CHARACTERISTICS OF ANATOLIAN WATER BUFFALO IN BALIKESİR, DÜZCE AND KÜTAHYA** 
(Balıkesir, Düzce ve Kütahya'daki Anadolu mandalarında bazı çevre faktörlerinin büyüme ve laktasyon özelliklerine etkisi)
Ebru ERGÜNEŞ BERKİN, Mustafa KÜÇÜKKEBAPÇI, Cüneyt KAPTAN.....146-150
- 4. ETHNOBOTANY IN THE DEVELOPMENT OF SUSTAINABLE TOURISM IN MUDURNU** 
(Mudurnu'da sürdürülebilir turizm gelişiminde etnobotanik)
Nail HOŞCAN, Hamdi ZENGİNBAL.....151-159

Review Article

- 5. RESEARCH AND CONSERVATION STUDIES ON ENDANGERED SNOWFLAKE (LEUCOJUM AESTIVUM) IN TURKEY** 
(Türkiye'nin nesli tehlike altına olan göl soğanı (*Leucojum aestivum*) üzerine araştırma ve koruma çalışmaları)
Fisun Gürsel ÇELİKEL, Sevim DEMİR.....160-164
- 6. ONCOGENIC VIRUSES AND THEIR MOLECULAR DIAGNOSIS IN POULTRY** 
(Kanatlı hayvanlarda onkojenik virüsler ve moleküler tanısı)
Berkant İsmail YILDIZ, Mustafa ERSAL, Kemal KARABAG.....165-172
- 7. OVERVIEW OF DIFFERENT METHODS USED IN CLUSTERING ALGORITHMS** 
(Kümeleme algoritmalarında kullanılan farklı yöntemlere genel bakış)
Tohid YOUSEFİ, Mehmet Serhat ODABAŞ, Recai OKTAŞ GÖKDAL.....173-189
- 8. INVESTIGATION OF STUDIES ON NATURAL GAS CONSUMPTION FORECASTING BY ARTIFICIAL NEURAL NETWORKS** 
(Yapay sinir ağları ile doğal gaz tüketim tahminini üzerine yapılmış çalışmaların irdelenmesi)
Bedia KANT, Mehmet Serhat ODABAŞ.....190-197



Research Article

Volume 3 - Issue 4: 128-137 / October 2020

A COMPARATIVE STUDY FOR INDOOR PLANAR SURFACE SEGMENTATION VIA 3D LASER POINT CLOUD DATA

Eyüp Eymen ERUYAR¹, Metehan YILMAZ¹, Berat YILMAZ¹, Onur AKBULUT¹, Kaya TURGUT¹, Burak KALECİ^{1*}

¹Eskisehir Osmangazi University, Faculty of Engineering, 26000, Eskisehir, Turkey

Received: May 11, 2020; **Accepted:** July 20, 2020; **Published:** October 01, 2020


Abstract

In recent years, point cloud data generated with RGB-D cameras, 3D lasers, and 3D LiDARs have been employed frequently in robotic applications. In indoor environments, RGB-D cameras, which have short-range and can only describe the vicinity of the robots, generally are opted due to their low cost. On the other hand, 3D lasers and LiDARs can capture long-range measurements and generally are used in outdoor applications. In this study, we deal with the segmentation of indoor planar surfaces such as wall, floor, and ceiling via point cloud data. The segmentation methods, which are situated in Point Cloud Library (PCL) were executed with 3D laser point cloud data. The experiments were conducted to evaluate the performance of these methods with the publicly available Fukuoka indoor laser dataset, which has point clouds with different noise levels. The test results were compared in terms of segmentation accuracy and the time elapsed for segmentation. Besides, the general characteristics of each method were discussed. In this way, we revealed the positive and negative aspects of these methods for researchers that plan to apply them to 3D laser point cloud data.

Keywords: 3D laser, Segmentation, Planar surfaces, Indoor, Fukuoka indoor dataset, PCL

***Corresponding author:** Eskisehir Osmangazi University, Faculty of Engineering, 26000, Eskisehir, Turkey

E mail: burakaleci@gmail.com (B. KALECİ)

Eyüp Eymen ERUYAR  <https://orcid.org/0000-0003-2775-1798>
Metehan YILMAZ  <https://orcid.org/0000-0002-3083-2460>
Berat YILMAZ  <https://orcid.org/0000-0003-0214-2662>
Onur AKBULUT  <https://orcid.org/0000-0003-2918-2832>
Kaya TURGUT  <https://orcid.org/0000-0003-3345-9339>
Burak KALECİ  <https://orcid.org/0000-0002-2001-3381>

Cite as: Eryar EE, Yilmaz M, Yilmaz B, Akbulut O, Turgut K, Kaleci B. 2020. A comparative study for indoor planar surface segmentation via 3D laser point cloud data. BSJ Eng Sci, 3(4): 128-137.

1. Introduction

In past years, the researchers generally exploited range data acquired from ultrasonic sensors and 2D laser

scanners and/or visual data captured with cameras to provide information to robots for tasks they are expected to perform. Although the range data is generally accurate,

it can only present planar information at the height where the sensor is located. On the other hand, the visual data yields the color information about the scene. However, its reliability heavily relies on the lighting conditions of the environment. Besides, the range and/or visual data may not describe the vicinity of the robot sufficiently due to their natural characteristics. For this reason, in recent years, point cloud data has been employed frequently in robotic applications because of its powerful ability to describe the shape, size, position, and orientation of objects (Grilli et al., 2017). 3D point cloud data can be acquired with various sensors such as RGB-D cameras, 3D lasers, and 3D LiDARs. The most significant advantage of the RGB-D cameras is their low cost compared with the 3D lasers and 3D LiDARs. The point cloud data generated with RGB-D cameras has the same structure with RGB images since it is constructed by regarding camera's position and depth information in each pixel. RGB-D cameras are frequently preferred in robotic applications due to these advantages. However, the foremost disadvantage of these cameras is their short-range. On the other hand, 3D lasers and LiDARs generally can capture long-range measurements when they are compared with RGB-D cameras. For this reason, these sensors are often used in outdoor applications such as autonomous vehicles and urban mapping (Xie et al., 2019).

The segmentation of point clouds can be defined as the process of separating the points that have the same characteristics into homogeneous parts (Nguyen and Le, 2013; Grilli et al., 2017). Point cloud segmentation is an essential preprocessing or post-processing step for activities such as semantic information extraction, object recognition, place classification, and human tracking. Thus, it is a very active research area in a wide range of applications from outdoor to indoor. The point cloud segmentation was used in different outdoor applications such as building roof plane segmentation with the Airborne Laser Scanning (ALS) point cloud data (Tarsha-Kurdi et al., 2007; Xu et al., 2016), building facade segmentation with the Terrestrial Laser Scanning (TLS) point cloud data (Ning et al., 2009; Vo et al., 2015), forest monitoring (Morsdorf et al., 2004; Ferraz et al., 2010), autonomous vehicles (Himmelsbach et al., 2010; Zermas et al., 2017), and 3D silhouette extraction of a street (Mutlu et al., 2014). The point cloud segmentation was also applied in indoor applications such as Building Information Modeling (BIM) (Anil et al., 2013; Qu et al., 2014), robotic applications (Rusu et al., 2008; Koppula et al., 2011), and object detection (Mattausch et al., 2014; Cadena and Košečka, 2015). In the reviews, Nguyen and Le (2013) and Grilli et al. (2017) classified the point cloud segmentation methods into five categories: edge-based, region-based, model-based, clustering-based, and graph-based. Edge-based methods first aim to detect the edges where point characteristics are changed. Then, they group the points between the edges. The edge-based methods are generally weak against noise and they may produce low segmentation accuracy although they provide fast

segmentation. Region-based methods exploit the neighbor relation to segment the point clouds. Despite their long segmentation time, the region-based methods present high segmentation accuracy. Model-based methods try to fit a model for each segment. They are fast and robust against noise. The disadvantage of these methods is to produce wrong models when the point cloud has nearly coplanar surfaces and the points in the cloud have uneven density. Clustering-based methods are appropriate for irregular object segmentation since they do not depend on a specific model. Also, they can combine different criteria for segmentation. The main disadvantage of these methods is that they have high computational complexity. The graph-based methods can be considered as the subcategory of the clustering-based methods.

In this study, we cope with the segmentation of indoor planar surfaces such as wall, floor, and ceiling. Although each segmentation category has advantages and disadvantages, most of the existing studies that address the segmentation of planar surfaces in indoor environments applied the well-known model-based RANSAC and region growing methods. Besides, these studies generally used the point cloud data captured with RGB-D cameras. On the other hand, 3D lasers and LiDARs can offer better performance in indoor robotic applications, especially for BIM, mapping, and semantic information extraction, due to their long-range measurements. In this study, we aim to examine the performance of the segmentation methods that are situated in PCL (Rusu and Cousins, 2011) with 3D laser point cloud data for planar surfaces in the indoor environment. In this way, we investigate the potential of edge-based, clustering-based, and graph-based (supervoxels) methods for planar surfaces with 3D laser point cloud data. To achieve this, the experiments were conducted with Fukuoka indoor laser dataset containing point cloud data, which has different noise levels (Martinez Mozos et al., 2019). After some preprocessing operations were applied to the dataset, point-wise labeling was performed with the RViz cloud annotation tool (Monica et al., 2017). The test results were compared in terms of segmentation accuracy and the time elapsed for segmentation. Besides, the general characteristics of each method were discussed.

The rest of the paper is organized as follows: In section 2, a detailed explanation of the segmentation methods is given. We present the experimental setup and results in section 3 and we conclude with section 4.

2. Material and Methods

The segmentation methods in the PCL are listed as follows: plane model segmentation, cylinder model segmentation, Euclidean cluster extraction, region growing segmentation, color-based region growing segmentation, min-cut based segmentation, conditional Euclidean clustering, the difference of normals based

segmentation, supervoxel clustering and progressive morphological filtering (PCL-S, 2020).

In this study, we focus on the segmentation of structural planar surfaces such as wall, floor, and ceiling and we intend to use only range data for this purpose. Therefore, the cylinder model segmentation method (Rabbani et al., 2006) that detects cylinders and spheres is out of the scope of this paper. The Euclidean cluster extraction approach (PCL-ECE, 2020) successfully segments the objects since it separates the point cloud into smaller parts. However, this approach is not appropriate for structural planar surfaces. The color-based region growing segmentation approach (PCL-CBRG, 2020) uses color instead of curvature and normal features while segmenting the point clouds. Hence, this method is also out of the scope of this paper. Since the min-cut based segmentation approach (PCL-MCBS, 2020) merely is designed for object segmentation, we did not examine the method. Lastly, progressive morphological filter segmentation (PCL-PMFS, 2020) is also out of the scope of this paper because it aims to segment the roof of the buildings with LiDAR sensors mounted in aircraft.

As a result, in this study, we mainly concentrate on the following methods: plane segmentation (PCL-PS, 2020), region growing (PCL-RG, 2020), conditional Euclidean clustering (PCL-CEC, 2020), the difference of normals based segmentation (PCL-DONBS, 2020), and supervoxel clustering (PCL-SC, 2020). For this reason, in this section, we briefly discuss the previous studies that employ these methods while considering the positive and negative aspects. Then, the details of the methods are explained.

2.1. RANSAC

The main idea behind the model-based segmentation methods is to group the points that fit the same mathematical model for primitive shapes such as plane, sphere, cone, cylinder, cubes, and torus. Fischler and Bolles (1981) introduced the well-known model-based RANSAC algorithm used in segmenting planar surfaces in point cloud data. Since the RANSAC does not need to know the relation between neighbor points, it could be applied to both organized point clouds acquired by RGB-D camera (Lu and Song, 2015) and unorganized point cloud data captured with 3D LiDAR or laser sensors (Xu et al., 2016) without any preprocessing step. Moreover, Schnabel et al. (2007) presented a RANSAC-based method for automatically segmenting fundamental shapes such as plane, cylinder, and cones in both mesh and point cloud data. They optimized the method for time complexity and they showed the robustness of the proposed method against outliers. There is an enormous number of studies that apply the RANSAC method for segmentation both in indoor and outdoor applications. In the reviews, Nguyen and Le (2013) and Grilli et al. (2017) explained most of these studies. Besides, several reviews for RANSAC-based methods were published (Kim and Yu, 2009; Raguram et al., 2012). RANSAC is an iterative and fast method and also robust against the noise and the outliers in the point cloud. However, RANSAC-based methods can lead to

inaccurate results when the point cloud includes complex-shaped objects and the points in the scene do not evenly distribute. Also, the RANSAC-based method may not recognize coplanar surfaces since they only utilized the mathematical model of the planes.

The RANSAC is mainly a prediction approach that is executed iteratively to segment point cloud data. The algorithm starts with the selection of the mathematical model. Then, a small set of feasible points is selected randomly instead of searching the large set of points that fit the model. The small set is enlarged regarding the DISTANCE_THRESHOLD parameter. The distance between a point ($p=(x_p, y_p, z_p)$) and the model is calculated as given in Equation 1. In the equation, we assume that the model is plane and the a , b , c and d describe plane parameters. The distance value can be considered as the error. If the error is less than the threshold, the point is added to that model and the model is updated. After a model is entirely segmented, the points that belong to the model are extracted from the point cloud and the process is repeated until the number of remaining points reaches the predefined number.

$$distance = \frac{ax_p + by_p + cz_p + d}{\sqrt{a^2 + b^2 + c^2}} \quad (1)$$

2.2. Region Growing

The region-based methods utilize the local information to cluster the points into regions. To determine the points to be added to a region, the features such as surface orientation, curvature, normal, etc. are investigated for points in a predetermined radius or a certain number of neighbors (Rabbani et al., 2006; Jagannathan and Miller, 2007). Therefore, a preprocessing step is required to define neighborhood relationships before using these methods with unorganized point cloud data (Vo et al., 2015). In the reviews, Nguyen and Le (2013) and Grilli et al. (2017) divide the region-based methods into two categories: Top-down methods (unseeded) and bottom-up (seeded) methods. The top-down methods start with one region that includes all points in the point cloud data. Then, they separate the region into subregions according to a criterion. The success of these methods highly depends on the selected criterion. On the contrary, the bottom-up methods first, select some seed points, and then points that satisfy a predefined condition, are joined to seeds to construct the regions (Besl and Jain, 1988). The region-based methods are generally robust against the noise. The success of these methods relies on selecting seed points and the merging criterion (Xie et al., 2019). Besides, these methods are sensitive to inaccurate normal and curvature values and this situation can cause incorrect results at points in the boundary of the regions where normal and curvature values change very quickly. To avoid these incorrect results, the search radius or number of neighbors around a point can be increased. In that case, the elapsed segmentation time increases. Therefore, there is a trade-off between the success of

segmentation and the segmenting time in region-based approaches (Nguyen and Le, 2013).

Although the region-based methods are separated into two categories, the researchers generally prefer to use bottom-up (seeded) methods. Besl and Jain (1988) introduced the first seeded region growing algorithm. The algorithm consists of two steps: 1) Determining the seed points, 2) growing the seed to form regions. The selection of seed points is an essential step for the success of the region-based methods. Besl and Jain (1988) determined the seed points according to the curvature value of the points. However, seed selection may vary for other methods. For example, Rabbani et al. (2006) and Ning et al. (2009), first, determined an appropriate model for a point and its neighbors to select a seed point. Then, the point that has the smallest distance with the plane created with the neighbor points is chosen as a seed point. Once the seed point is specified, a search list is created as pushing the seed point and its neighbors. At that point, the search begins to examine the similarities between the seed point and other list items in terms of local features such as curvature, surface orientation, smoothness, normal, etc. Besl et al. (1988) preferred normal of the points as the local feature and the angle between normal vectors (n_i and n_j) was calculated. If the angle is less than a predefined threshold value, the neighbor point is added to the region corresponding to the seed point (Equation 2). After that, the seed point is removed from the search list, and the recently added point becomes the new seed. The process is repeated for new seed points until no seed point for that region is left. The method continues with a new seed point for a new region.

$$\cos^{-1}(\mathbf{n}_i \cdot \mathbf{n}_j) < threshold \quad (2)$$

2.3. Conditional Euclidean Clustering

Conditional Euclidean clustering is a variant of the region growing approach, in which local features that are employed to cluster the neighbor points can be customized according to the application. In this way, different features such as color, normal, smoothness, and distance can be aggregated for segmentation (PCL-CEC, 2020). For example, Wu et al. (2019) combined color and distance features in conditional Euclidean clustering to segment the juicy peaches on trees. Zhou et al. (2018) generated topological maps by means of signed surface variation and distance-based conditions. Besides, conditional Euclidean clustering allows determining the number of points in a cluster and applying different local features for different regions of the point cloud or same clustering criterion with varying parameters to segment different-sized objects in a point cloud data. For these reasons, conditional Euclidean clustering method is appropriate for object segmentation and planar surface segmentation. The main disadvantage of the method is that its success depends on the merging condition of region growing.

In this study, apart from the region growing method that uses similarities between normal vectors of neighbors as a clustering constraint, we opted to consider the x, y, or z components of the point's normal to merge the neighbor points into the same segment in conditional Euclidean clustering method. To do that, for each point, we first determined the largest component of its normal vector according to the Equation 3. We called that component as the dominant axis of the normal and employed it for planar surface segmentation. If neighbor points have the same dominant axis, these points are added to the same segment. The idea behind this is that the normal vectors of the planar surfaces generally parallel to one of the x, y, or z-axis. In this way, we aimed to reduce the segmentation errors that occur at the boundary regions of planar surfaces when the similar normal vectors criterion is applied for segmentation like in the region growing method.

$$\text{dominant axis} = \arg \max_n(n_x, n_y, n_z) \quad (3)$$

2.4. The Difference of Normal Based Segmentation

In computer vision applications, the operators such as the difference of Gaussians (DoG) and Laplacian of the Gaussian (LoG) have been widely used to detect edges, to find salience points, and to pre-segment images, before complex algorithms are executed. In a similar fashion, the difference of normals (DON) operator, which calculates normal vectors for large-scale and small-scale support radiuses, is mainly designed to identify essential and distinguishing points needed for object identification within large unorganized the point cloud data for outdoor applications (Ioannou et al., 2012). For example, Su et al. (2018) applied the difference of normals method to extract corn leaves. The main advantage of the method is that the support radiuses can be adjusted depending on the object to be segmented. However, the size of the support radius should be selected carefully since the calculation time of the normal vectors highly depends on this selection. Besides, apart from the region-based methods, the difference of normals does not need the local information to determine the segments.

The main idea behind the difference of normals method is to observe the surface normal vectors that describe the surface geometry. The researchers generally prefer to use the support radius or a fixed number of neighbors to determine the surface normal in spite of the existence of many different methods. The radius or the number of neighbors describes the size of the surface that the normal represents. The difference of normals method calculates surface normal vectors for the small support radius (r_s), which reflects small changes on the surface geometry, and the large support radius (r_l) that depicts the general character of the surface. If the small support radius normal and large support radius normal for a point are similar to each other, it indicates that the normal vectors reflect the character of the surface. However, the difference between small support radius normal and large

support radius normal increases when a point lies at the edge of a planar surface due to the rapid changes in small support radius normal. The difference of normals based segmentation method utilizes from that idea. First, the method executes the difference of normals operator to calculate the arithmetic average of the small and large support radius normal vectors for each point (Equation 4). Then, conditional removal filtering (PCL-CRF, 2020) is applied to remove the points that the difference between two normal vectors is greater than a predefined threshold value (FILTER_THRESHOLD). In other words, the filtering step eliminates the points located at the edge of the planar surfaces. Finally, the Euclidean cluster extraction method is employed to cluster the remaining points (Ioannou et al., 2012).

$$\Delta n(p, r_s, r_l) = \frac{n(p, r_s) - n(p, r_l)}{2} \quad (4)$$

2.5. The Supervoxel Clustering

In recent years, graph-based methods have gained popularity in 2D image segmentation due to their pixel-based strategy that significantly increases the segmentation success (Egger et al., 2012). These methods construct graphs, in which each pixel is represented with a node. Then, the nodes are processed to enhance pixel-based low-level data to object-scale high-level information. On the other hand, the graph-based methods suffer from high time complexity since they handle all pixels in the image. To reduce the computational time of the graph-based methods, the superpixel concept, which considers a group of pixels instead of each pixel as a node, is revealed. After the successful implementation of the superpixel concept for 2D image segmentation, the supervoxel concept has been begun to be used for segmentation of point cloud data. For example, Lin et al. (2018) proposed an improved supervoxel implementation to diminish the segmentation errors at the supervoxel boundaries. They showed the effectiveness of their work on three different datasets, including different types of point cloud data.

The supervoxel clustering method first creates the voxel octree structure regardless of whether the point cloud data is organized or unorganized. Moreover, an adjacency graph with a resolution of R_{voxel} is constructed, this is essential for segmentation of the point cloud data. Although these steps seem to be time-consuming and can be considered a negative aspect of the method, they allow utilizing neighbors relation and speeding up the segmentation process. Then, the seed points must be selected because the supervoxel clustering method is a variant of the region-based methods. It is clear that the selection of the seed points profoundly affects the segmentation accuracy. For that reason, R_{seed} parameter, which should be chosen greater than R_{voxel} value, is used to determine the evenly distributed seed points in the voxelized grid. The seed voxels and its neighbors form the supervoxels that substantially are the cluster of features

such as color, spatial distance, and normal. After supervoxels are determined, the growing process is executed with the k-means clustering algorithm and breadth-first search. First, the nearest voxel to the supervoxel and its neighbors are handled. Then, the similarity between these voxels and the supervoxel in the feature domain is investigated. Since the supervoxels consist of color, distance, and normal features, the distance between supervoxel and the voxels is calculated with a weighted distance function as given in Equation 5:

$$D = \sqrt{w_c D_c^2 + \frac{w_s D_s^2}{3R_{\text{seed}}^2} + w_n D_n^2} \quad (5)$$

where D_c , D_s , and D_n represent the color, spatial, and normal distance, respectively. Also, weights of these distances are described with w_c , w_s , and w_n , respectively. The voxel that has the smallest distance value is joined to the supervoxel and its neighbors in the adjacency graph are added to the search queue. The same process is performed for all supervoxels and the depth of the examined voxels remains the same. The search is terminated when the boundaries of supervoxel are reached or no neighbors are available for search. Lastly, the k-means clustering algorithm is applied to update the cluster centers. The process continues until either the supervoxels stabilize or a predefined number of iterations is performed (Papon et al., 2013).

3. Results

3.1. Experimental Setup

The publicly available Fukuoka indoor dataset was used to analyze the performance of the segmentation methods that are situated in PCL. The dataset consists of panoramic point cloud data captured with the SICK LMS-151 laser sensor. The point clouds are separated into five categories: Corridor, kitchen, laboratory, study room, and office. In the laboratory, kitchen, and study room categories, samples generally include objects such as tables, computer screens, and chairs. However, in this study, we intend to segment structural planar surfaces such as wall, floor, and ceiling. For this reason, we selected 30 samples from the corridor and office categories that mainly include points representing these planar surfaces. Besides, we elaborated on the selected samples containing different levels of noise. Each of the selected point clouds has approximately 474200 points. Some preprocessing steps were performed to prepare the point cloud data for segmentation. First, the points that belong to non-planar surfaces such as flowerpot, air conditioning parts, and fire alarms were removed because these points may adversely affect the performance of the methods. For this reason, the natural characteristic of these methods can be misunderstood. Then, the normal vector of each point in the point cloud data was calculated since all methods use normal information as a local feature for segmentation, excluding the RANSAC. For each

point, to determine the normal vector, firstly, a surface is fitted with the neighbors around the point. Then, the surface normal is considered the normal vector of the point. Thus, this process allowed us to compare the methods fairly because all methods employed the same normal vectors (PCL-SS, 2020). The experiments were conducted on a PC with the Intel i7-7700HQ processor, 16GB RAM, and Ubuntu 16.04 operating system.

The test results were compared in terms of segmentation accuracy and the time elapsed for segmentation. In order to calculate the segmentation accuracy, point-wise labeling was performed for each point cloud data through the RViz cloud annotation tool (Monica et al., 2017). Then, we determined paired segments between the segmentation result of a method and the labeled data. The segmentation accuracy of a sample is calculated by dividing the number of correctly segmented points to the total number of points. The average of all samples is considered the average segmentation accuracy.

In the RANSAC method, the `DISTANCE_THRESHOLD` parameter that determines whether a point is in the plane. The parameter should be selected as high as possible to cope with the noise. However, all points could be clustered into a segment when the parameter is too high. The parameter was selected empirically as 0.03 meters to consider these circumstances. In the region growing method, we used the default values at the PCL. The number of neighbors that examined in segmentation was 30. Also, the curvature and smoothness threshold were set to 1 and 3 degrees, respectively. In the conditional Euclidean clustering method, the neighborhood relationship is established with the search radius. We tried to set the radius of the conditional Euclidean clustering method, in which the number of neighbors to be approximately 30. In the experiments, the radius was selected as 0.015 meters. Also, the minimum and the maximum number of points that a segment can possess were 3000 and 1000000, respectively. In the difference of normal based segmentation method, small support radius (r_s) and large support radius (r_l) were 0.015 and 0.075 meters, respectively. Notice that, we selected the small support radius the same as the radius in the conditional Euclidean clustering method to compare the methods fairly. Besides, Ioannou et al. (2012) recommended selecting the large support radius five times the small support radius. There are five main parameters used in supervoxel clustering. Voxel size value (R_{voxel}), which determines octree leaf resolution, was selected empirically as 0.01 meters. The seed size value (R_{seed}) that describes the maximum size of the supervoxel was chosen 5.0 meters to contain all points in the point cloud data. In this study, we used only normal values in the supervoxel clustering method. Therefore, weights for color (w_c), spatial distance (w_s), and normal (w_n) were 0.0, 0.0, and 1.0, respectively.

3.2. Results

3.2.1. Segmentation time

The segmentation time is one of the essential metrics to evaluate the segmentation methods. The segmentation methods mentioned in Section 2 were applied to the selected 30 scenes for examining the performance of these methods in terms of segmentation time. Table 1 shows the average elapsed time to segment a scene for each method.

Table 1. Average segmentation time of a scene

Methods	Average Time (Seconds)
RANSAC	0.4285
Region Growing	6.2855
Conditional Euclidean Clustering	5.6124
Difference of Normal Based Segmentation	3.2140
Supervoxel Clustering	5.2726

The conditional Euclidean clustering and the supervoxel clustering methods are variants of the region-based methods. These methods utilize the local information to form a region; therefore, segmentation time for these methods was longer when they are compared with the RANSAC and the difference of normals based segmentation. This is an expected result because searching for neighbors is a time-consuming operation and the RANSAC and the difference of normals based segmentation do not perform the operation for segmentation. The segmentation time of region-based methods was also different from each other. The slight difference between the region growing and the conditional Euclidean clustering could be explained with different search approaches that these methods use. Generally, region growing based methods search their vicinity with a radius or the number of neighbors. As applying the region growing, the number of neighbors was selected 30. On the other hand, the conditional Euclidean clustering method uses the radius, which was chosen as 0.015 meters. Although we tried to set the radius of the conditional Euclidean clustering method, in which the number of neighbors to be approximately 30, the number of neighbors may vary according to the density of the points. Apart from the region growing and the conditional Euclidean clustering, which are utilized point-wise neighborhoods, the supervoxel clustering method uses voxel-wise neighborhoods. Therefore, in the worst-case scenario, each voxel includes only one point and the performance of the region growing and the supervoxel method could be the same in terms of the segmentation time. As a result, the number of points in a supervoxel that is adjusted with the R_{voxel} parameter determines the segmentation time. The difference of normals based segmentation method could be considered as the edge-based methods since the method essentially is designed for the detection of the edges. In theory, the main advantage of the edge-based methods is their low

computational complexity. Once the segmentation time performance of the difference of normals based segmentation method is analyzed, we separated elapsed time into three stages. In the first stage, the difference of normal vectors was calculated in 0.1265 seconds. Then, conditional removal filtering took 0.016 seconds. In the last stage, Euclidean clustering was applied and it required 3.0715 seconds. Thus, the time-consuming part of the method is clustering and if an appropriate clustering method is chosen, the segmentation time could be decreased. Finally, RANSAC was the fastest one among the methods. It is an expected result because the RANSAC tries to fit a mathematical model and it does not utilize the local information to segment the planes.

3.2.2. Segmentation accuracy

In order to examine the success of methods that are situated in PCL, the segmentation accuracy was calculated for each method. Table 2 shows the average segmentation accuracy for a scene. Although RANSAC, region growing, and the conditional Euclidean clustering methods yielded approximately the same segmentation accuracy, they behaved differently in different scenes. Also, the reasons why the difference of normals based method and the supervoxel clustering method produce low segmentation accuracy when they are compared to the other methods were discussed.

Table 2. Average segmentation accuracy of a scene

Methods	Average Accuracy
RANSAC	94.33%
Region Growing	94.19%
Conditional Euclidean Clustering	93.54%
Difference of Normal Based Segmentation	89.50%
Supervoxel Clustering	64.49%

To analyze the characteristics of each method, we selected four scenes among the 30 scenes, which are shown in columns of Figure 1. Another criterion of the selection of these scenes was the noise level of the data. We selected scenes that include different noise levels. Scene 1 and scene 2 include low-level noise when they are compared with scene 3 and scene 4. The first row in the figure shows the ground truth of the scenes. The remaining rows depict the results for each segmentation method. Also, the segmentation accuracy and time for each scene is given at the right of the scenes. In figures, the incorrect or unsegmented regions are represented with white ellipses. The RANSAC method suffers from combining the coplanar surfaces that are located in different locations on the same surface. For example, in scene 1, there are two vertical planes, which lie on two sides of the door plane. These planes approximately fit the same mathematical model and the RANSAC does not utilize the local information. Thus, the RANSAC added these planes into one segment, which is shown with

yellow color. A similar problem also occurred in scene 4 and the points in the ellipses were incorrectly joined into the wall plane that is depicted in yellow color. In order to deal with the high noise, the `DISTANCE_THRESHOLD` parameter was selected as high as possible. Therefore, the high value provides accurate segmentation at intersection points as the noise level increases and scene 3 and scene 4 show examples of this situation. On the other hand, the incorrect results were produced at the intersection points of two planar surfaces in scene 2 due to the high value in the selection `DISTANCE_THRESHOLD` parameter. The well-known negative aspect of the region growing methods is that they may lead to incorrect results at points in the boundary of the regions where normal and curvature values change very quickly. The ellipses in the figures show the unsegmented regions. As the noise level increases just like scene 3 and scene 4, the normal and the curvature values are more corrupted than the values in scene 1 and scene 2. Thus, the segmentation accuracy for scene 3 and scene 4 are low since the number of unsegmented points directly proportional to the noise level. Besides, the required time for segmentation increases as the noise level rises. In this study, the dominant axis concept was suggested for the merging constraint of the conditional Euclidean clustering method. Therefore, the success of the method heavily depends on the accuracy of the normal. The method offered promising results for scene 1 and scene 2. Moreover, for these scenes, the method outperformed the RANSAC and region growing methods in terms of segmentation accuracy. However, as the noise level increases, the success of the method dramatically decreases. Nevertheless, the potential of the method should not be underestimated. The difference of normals based segmentation method employs the filtering process after small and large support radius normal vectors are calculated. Thus, the success of the method heavily depends on the filtering process. The filtering algorithm uses a threshold to determine whether a point is filtered. If the threshold is low, all points that the normal values are changed larger than the threshold are eliminated. All points could be combined into one segment when the threshold is high. As a result, we observed that the best accuracy success is obtained when the threshold was set to 0.1. However, the threshold could be adaptively adjusted according to the noise level to increase the segmentation accuracy. The method can always be considered for the applications that need fast segmentation and when the data contains coplanar surfaces that the RANSAC prone to error. The supervoxel clustering method was not appropriate for the segmentation of planar surfaces since it only allows adjusting the R_{voxel} parameter. The parameter determines the number of supervoxel and their boundaries. Then, the method clusters the points in each supervoxel according to the normal vectors. The points that have similar normal vectors are added to the same region and this causes incorrect segmentation results.

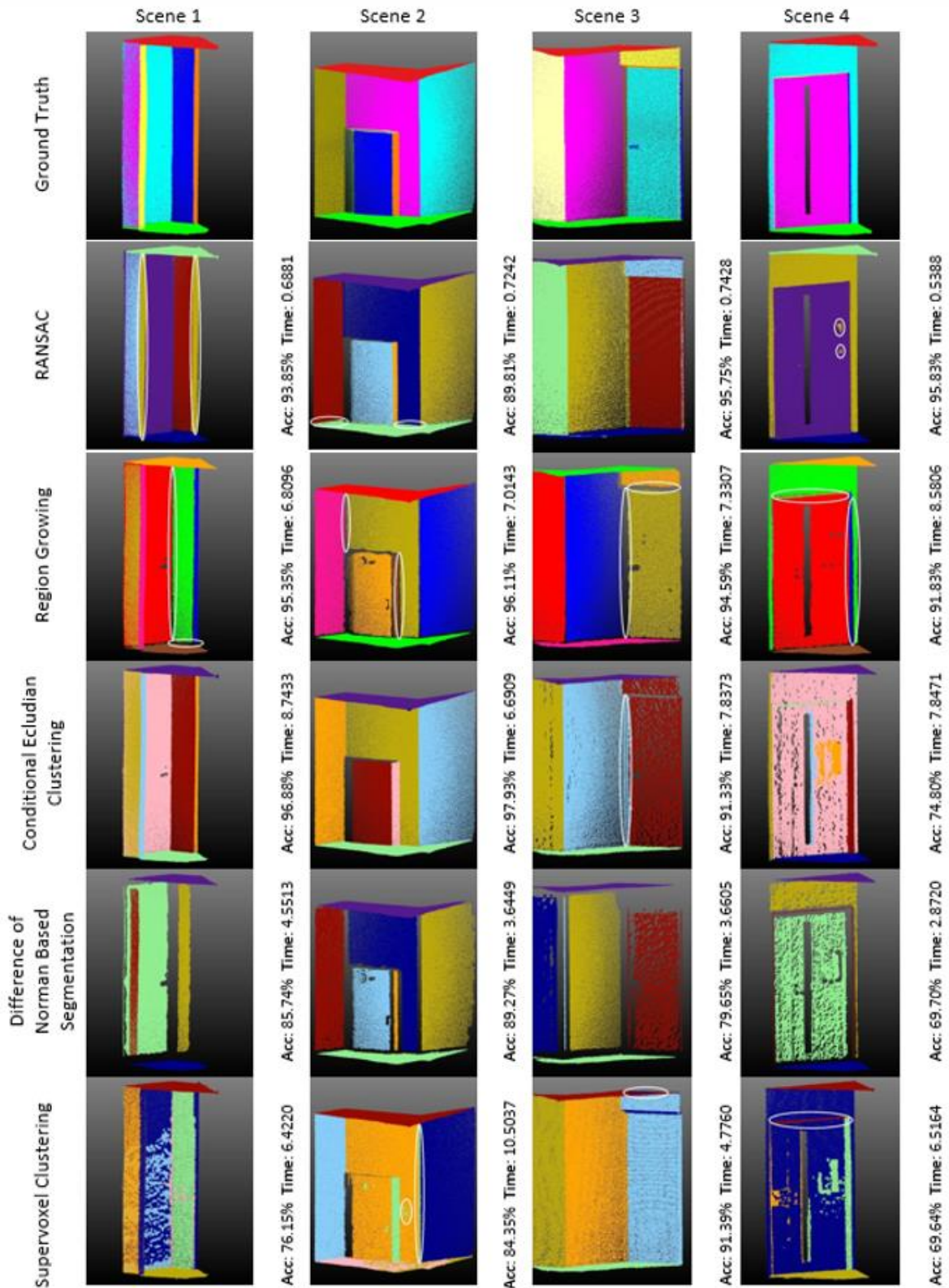


Figure 1. The segmentation results.

4. Conclusion

In this study, the segmentation of indoor structural planar surfaces such as wall, floor, and ceiling via point cloud data that is captured with the 3D laser range scanner was considered. In previous studies, the researchers generally used RANSAC and region-growing methods. However, other segmentation methods such as conditional

Euclidean clustering, the difference of normals based segmentation, and the supervoxel clustering may produce satisfactory results for indoor applications. For this reason, we examined the performance of these methods in terms of segmentation accuracy and segmentation time to reveal the positive and negative aspects of these methods for researchers that plan to apply them to 3D

laser point cloud data. To do that, the publicly available Fukuoka indoor laser dataset, which has point clouds with different noise levels, was used. The test results were evaluated regarding the noise level of point clouds and the behaviors of each method against the noise were discussed.

Conflict of interest

The authors declare that there is no conflict of interest.

References

- Anil EB, Tang P, Akinci B, Huber D. 2013. Deviation analysis method for the assessment of the quality of the as-is Building Information Models generated from point cloud data. *Aut in Construc*, 35: 507-516.
- Besl PJ, Jain RC. 1988. Segmentation through variable-order surface fitting. *IEEE Transact on Pat Analy Machine Intel*, 10(2): 167-192.
- Cadena C, Košečka J. 2015. Semantic parsing for priming object detection in indoors RGB-D scenes. *Int J Robotics Res*, 34(4-5): 582-597.
- Egger J, Colen RR, Freisleben B, Nimsky C. 2012. Manual refinement system for graph-based segmentation results in the medical domain. *J Medic Sys*, 36(5): 2829-2839.
- Ferraz A, Bretar F, Jacquemoud S, Gonçalves G, Pereira L. 2010. 3D segmentation of forest structure using a mean-shift based algorithm. *IEEE Int Conference on Image Processing*, 1413-1416.
- Fischler MA, Bolles RC. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun of the ACM*, 24(6): 381-395.
- Grilli E, Menna F, Remondino F. 2017. A review of point clouds segmentation and classification algorithms. *The Intl Archives of Phot, Remote Sensing and Spatial Inf Sci*, 42: 339.
- Himmelsbach M, Hundelshausen FV, Wuensche HJ. 2010. Fast segmentation of 3D point clouds for ground vehicles. *IEEE Intelligent Vehicles Symposium*, 560-565.
- Ioannou Y, Taati B, Harrap R, Greenspan M. 2012. Difference of normals as a multi-scale operator in unorganized point clouds. *Second Int Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, 501-508.
- Jagannathan A, Miller EL. 2007. Three-dimensional surface mesh segmentation using curvedness-based region growing approach. *IEEE Transactions on Pat Analy Machine Intel*, 29(12): 2195-2204.
- Kim T, Yu W. 2009. Performance evaluation of ransac family. In *Proceedings of the British Machine Vision Conference (BMVC)*, 1-12.
- Koppula HS, Anand A, Joachims T, Saxena A. 2011. Semantic labeling of 3d point clouds for indoor scenes. In *Adv in Neural Inf Processing sys*, 244-252.
- Lin Y, Wang C, Zhai D, Li W, Li J. 2018. Toward better boundary preserved supervoxel segmentation for 3D point clouds. *ISPRS J Photogrammetry and remote sensing*, 143: 39-47.
- Lu Y, Song D. 2015. Robustness to lighting variations: An RGB-D indoor visual odometry using line segments. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 688-694.
- Martinez Mozos O, Nakashima K, Jung H, Iwashita Y, Kurazume R. 2019. Fukuoka datasets for place categorization. *Int J Robotics Res*, 38(5): 507-517.
- Mattausch O, Panozzo D, Mura C, Sorkine-Hornung O, Pajarola R. 2014. Object detection and classification from large-scale cluttered indoor scans. *Computer Graphics Forum*, 33(2): 11-21.
- Monica R, Aleotti J, Zillich M, Vincze M. 2017. Multi-label point cloud annotation by selection of sparse control points. *International Conference on 3D Vision (3DV)*, 301-308.
- Morsdorf F, Meier E, Kötz B, Itten KI, Dobbertin M, Allgöwer B. 2004. LIDAR-based geometric reconstruction of boreal type forest stands at single tree level for forest and wildland fire management. *Remote Sensing of Envir*, 92(3): 353-362.
- Mutlu B, Hacıömeroğlu M, Serdar GM, Dikmen M, Sever H. 2014. Silhouette extraction from street view images. *Int J Advanced Robotic Sys*, 11(7): 114.
- Nguyen A, Le B. 2013. 3D point cloud segmentation: A survey. *6th IEEE conference on robotics, automation and mechatronics (RAM)*, 225-230.
- Ning X, Zhang X, Wang Y, Jaeger M. 2009. Segmentation of architecture shape information from 3D point cloud. *Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry*, 127-132.
- Papon J, Abramov A, Schoeler M, Worgotter F. 2013. Voxel cloud connectivity segmentation-supervoxels for point clouds. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2027-2034.
- PCL-Color-Based Region Growing (PCL-CBRG). 2020. Point Cloud Library Color-Based Region Growing Code. URL: https://pcl.readthedocs.io/projects/tutorials/en/latest/region_growing_rgb_segmentation.html#region-growing-rgb-segmentation (access date: 01.05.2020)
- PCL- Conditional Euclidean Clustering (PCL-CEC). 2020. Point Cloud Library Conditional Euclidean Clustering Segmentation Code. URL: https://pcl.readthedocs.io/projects/tutorials/en/latest/conditional_euclidean_clustering.html#conditional-euclidean-clustering (access date: 01.05.2020).
- PCL- Conditional Removal Filtering (PCL-CRF). 2020. Point Cloud Library Conditional Removal Filtering Code. URL: https://pcl.readthedocs.io/projects/tutorials/en/latest/conditional_removal.html#conditional-removal (access date: 01.05.2020).
- PCL- Difference of Normal Based Segmentation (PCL-DONBS). 2020. Point Cloud Library Difference of Normal Based Segmentation Codes. URL: https://pcl.readthedocs.io/projects/tutorials/en/latest/don_segmentation.html#don-segmentation (access date: 01.05.2020).
- PCL-Euclidean Clustering Extraction (PCL-ECE). 2020. Point Cloud Library Euclidean Clustering Extraction Code. URL: https://pcl.readthedocs.io/projects/tutorials/en/latest/cluster_extraction.html#cluster-extraction (access date: 01.05.2020).
- PCL-Min-Cut Based Segmentation (PCL-MCBS). 2020. Point Cloud Library Min-Cut Based Segmentation Code. URL: https://pcl.readthedocs.io/projects/tutorials/en/latest/min_cut_segmentation.html#min-cut-segmentation (access date: 01.05.2020).
- PCL-Plane Segmentation (PCL-PS). 2020. Point Cloud Library Plane Segmentation Code. URL: https://pcl.readthedocs.io/projects/tutorials/en/latest/planar_segmentation.html#planar-segmentation (access date: 01.05.2020).
- PCL-Progressive Morphological Filter Segmentation (PCL-PMFS). 2020. Point Cloud Library Progressive Morphological Filter Segmentation Code. URL: [https://pcl.readthedocs.io/projects/tutorials/en/latest/progressive-](https://pcl.readthedocs.io/projects/tutorials/en/latest/progressive_morphological_filtering.html#progressive-)

- morphological-filtering (access date: 01.05.2020).
- PCL-Region Growing (PCL-RG). 2020. Point Cloud Library Region Growing Segmentation Code. URL: https://pcl.readthedocs.io/projects/tutorials/en/latest/region_growing_segmentation.html#region-growing-segmentation (access date: 01.05.2020).
- PCL-Segmentation (PCL-S). 2020. Point Cloud Library Segmentation Codes. URL: <https://pcl.readthedocs.io/projects/tutorials/en/latest/index.html#segmentation> (access date: 01.05.2020).
- PCL- Supervoxel Clustering (PCL-SC). 2020. Point Cloud Library Supervoxel Clustering Segmentation Code. URL: https://pcl.readthedocs.io/projects/tutorials/en/latest/supervoxel_clustering.html#supervoxel-clustering (access date: 28.04.2020).
- PCL-Surface Smoothing (PCL-SS). 2020. Point Cloud Library Surface Smoothing Code. URL: <https://pcl.readthedocs.io/projects/tutorials/en/latest/resampling.html?highlight=resampling#smoothing-and-normal-estimation-based-on-polynomial-reconstruction> (access date: 02.05.2020).
- Qu T, Coco J, Rönnäng M, Sun W. 2014. Challenges and trends of implementation of 3D point cloud technologies in building information modeling (BIM): case studies. *Computing in Civil and Build Eng*, 809-816.
- Rabbani T, Van Den Heuvel F, Vosselmann G. 2006. Segmentation of point clouds using smoothness constraint. *Int archives of Photogrammetry, Remote Sens and spatial Inf Sci*, 36(5): 248-253.
- Raguram R, Chum O, Pollefeys M, Matas J, Frahm JM. 2012. USAC: a universal framework for random sample consensus. *IEEE Trans on Pattern Analy and Machine Int*, 35(8): 2022-2038.
- Rusu RB, Cousins S. 2011. 3d is here: Point cloud library (pcl). *IEEE international conference on robotics and automation*, 1-4.
- Rusu RB, Marton ZC, Blodow N, Dolha M, Beetz M. 2008. Towards 3D point cloud based object maps for household environments. *Robotics and Autonomous Sys*, 56(11): 927-941.
- Schnabel R, Wahl R, Klein R. 2007. Efficient RANSAC for point-cloud shape detection. *Comp Graphics Forum*, 26(2): 214-226.
- Su W, Zhang M, Liu J, Sun Z. 2018. Automated extraction of corn leaf points from unorganized terrestrial LiDAR point clouds. *Int J Agri and Biol Eng*, 11(3): 166-170.
- Tarsha-Kurdi F, Landes T, Grussenmeyer P. 2007. Hough-transform and extended ransac algorithms for automatic detection of 3d building roof planes from lidar data. *ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007*, Sep 2007, Espoo, Finland, 407-412.
- Vo AV, Truong-Hong L, Laefer DF, Bertolotto M. 2015. Octree-based region growing for point cloud segmentation. *ISPRS J Photoy and Remote Sensing*, 104: 88-100.
- Wu G, Zhu Q, Huang M, Guo Y, Qin J. 2019. Automatic recognition of juicy peaches on trees based on 3D contour features and colour data. *Biosystems Eng*, 188: 1-13.
- Xie Y, Tian J, Zhu XX. 2019. A review of point cloud semantic segmentation. *arXiv preprint arXiv:1908.08854*.
- Xu B, Jiang W, Shan J, Zhang J, Li L. 2016. Investigation on the weighted ransac approaches for building roof plane segmentation from lidar point clouds. *Remote Sens*, 8(1): 5.
- Zermas D, Izzat I, Papanikolopoulos N. 2017. Fast segmentation of 3d point clouds: A paradigm on lidar data for autonomous vehicle applications. *IEEE International Conference on Robotics and Automation (ICRA)*, 5067-5073.
- Zhou W, Peng R, Dong J, Wang T. 2018. Automated extraction of 3D vector topographic feature line from terrain point cloud. *Geocarto Int*, 33(10): 1036-1047.



Research Article

Volume 3 - Issue 4: 138-145 / October 2020

ESTIMATING OF BIRTH WEIGHT USING PLACENTAL CHARACTERISTICS IN THE PRESENCE OF MULTICOLLINEARITY

Cem TIRINK^{1*}

¹Igdir University, Faculty of Agriculture, Department of Animal Science, Igdir, Turkey

Received: September 01, 2020; **Accepted:** September 10, 2020; **Published:** October 01, 2020


Abstract

In this study, it was aimed to compare the performance of proposed estimators in the presence of multicollinearity that will be used in regression analysis as an alternative to Least Squares. Birth weight was estimated by using placental features such as sex, placental efficiency, total cotyledon numbers, large cotyledon weight, medium cotyledon weight, small cotyledon weight, large cotyledon number, medium cotyledon number, small cotyledon number, large cotyledon width, medium cotyledon width, small cotyledon width, large cotyledon length, medium cotyledon length, small cotyledon length, large cotyledon depth, medium cotyledon depth, small cotyledon depth for Bafra sheep breed. In the presence of multicollinearity, more reliable models can be obtained by using some estimator. The performances of the Ridge and Liu estimators, which are suggested methods for this situation, were compared. MSE, RMSE, rRMSE, MAPE, R², and AIC were used as model comparison criteria. As a result of, in the presence of multicollinearity; Liu estimator is recommended as an alternative method to Least Squares.

Keywords: Least squares, Ridge estimator, Liu estimator, Multicollinearity, Placental characteristics

***Corresponding author:** Igdir University, Faculty of Agriculture, Department of Animal Science, Igdir, Turkey

E mail: cem.tirink@gmail.com (C. TIRINK)

Cem TIRINK  <https://orcid.org/0000-0001-6902-5837>

Cite as: Tirink C. 2020. Estimating of birth weight using placental characteristics in the presence of multicollinearity. BSJ Eng Sci, 3(4): 138-145.

1. Introduction

The main purpose of animal breeding is to genetically improve populations of livestock and thus produce more efficient production from future populations (Onder and Abaci, 2015; Sen et al., 2019). In general, newborn lambs with higher birth weight are considered as selection criteria that have higher birth weight are selected as breeding material for the future population (Sen et al., 2019).

Reproductive parameters have major effects in small ruminant breeding. In particular, the success of the

gestation period effect the health of the newborn lamb (Brzozowska et al., 2020). Postnatal mortality of offspring in sheep breeding depends on placental features (Dwyer et al., 2005). Many studies showed that the placental features have positive and strong correlations between birth weight for newborn lambs (Echternkamp, 1993; Konyali et al., 2007; Alkass et al., 2013; Ozyurek, 2019; Osgerby et al., 2003; Dwyer et al., 2005; Sen et al., 2013; Sen and Onder, 2016). Placental features are more important indicators for the birth weight in sheep and goat breeding (Ozyurek and Turkyilmaz, 2020).

Morphometric parameters of the placenta affect fetal growth and thus, placental size play vital roles for the birth weight and detecting the postnatal viability (Sen et al., 2013; Brzozowska et al., 2020).

Placental features give much information for the birth weight, postnatal viability. Researches using many properties are the subject of multivariate statistics. One of the multivariate statistical methods used to reveal the relationships between placental morphological features and birth weight of animals is regression analysis. In multivariate statistical modeling, regression analysis is a process to estimate the relationship between explanatory variables and response variable (Ari and Önder, 2013). Many methods are used to estimate the response variable and the most common of which is the Least Squares (LS) method (Uckardes et al., 2012). LS method requires some assumptions to make an effective model estimation. When in the presence of multicollinearity between explanatory variables is provided from these assumptions, alternative methods such as Ridge estimator, Liu estimator are proposed (Hoerl and Kennard, 1970; Liu, 1993).

The aim of this study, to compare the performance of proposed estimators in the presence of multicollinearity that will be used in regression analysis as an alternative to Least Squares using some placental characteristics in Bafra sheep.

2. Material and Method

This study was carried out on 40 Bafra sheep kept on Ondokuz Mayıs University research farm unit in Samsun Province of Turkey. For this aim, some placental measurements such as sex (S), placental efficiency (PE), total cotyledon numbers (TCN), large cotyledon weight (LCW), medium cotyledon weight (MCW), small cotyledon weight (SCW), large cotyledon number (LCN), medium cotyledon number (MCN), small cotyledon number (SCN), large cotyledon width (LCW_i), medium cotyledon width (MCW_i), small cotyledon width (SCW_i), large cotyledon length (LCL), medium cotyledon length (MCL), small cotyledon length (SCL), large cotyledon depth (LCD_e), medium cotyledon depth (MCD_e), small cotyledon depth (SCD_e) were used as explanatory variables. The birth weight of lamb is a response variable. All statistical analyses were performed using the RStudio software (R Core Team, 2020). The “stats”, “lmridge”, “liureg” packages were used to perform for LS method, Ridge estimator and Liu estimator, respectively (R Core Team, 2019; Imdad and Aslam, 2018a; Imdad and Aslam, 2018b). To perform the model selection criteria was used to “ahaGoF” package (Eyduvan, 2019).

In the matrix form of multiple regression is:

$$Y = X\beta + \varepsilon$$

where; Y is a response variable, X is a matrix for the explanatory variables and β is a vector for regression coefficients and ε is an error term. The most common

method used to estimate the model in regression analysis is the LS method. The main purpose of the LS minimizes the sum of squares of error terms (Kutner et al., 2004).

$$\hat{\beta}_{LS} = \operatorname{argmin} \sum (y_i - \hat{y}_i)^2 \quad i = 1, 2, \dots, n$$

Some assumptions need to be provided to estimate the optimum model with LS. The first assumption is that the regression model must be linear. The second assumption is that the value of the expected error for the regression model must be zero. Further, the variance of the errors must be constant and the errors must be independent (Sarstedt and Mooi, 2014). Besides these assumptions, there should not be a linear relationship (multicollinearity) between the explanatory variables (Tirink et al., 2020). In the presence of multicollinearity, the results of the model to be obtained will not reliable (Cankaya et al. 2019).

There are many methods for determining multicollinearity. One of them is the determination of multicollinearity with Variance Inflation Factor (VIF) value. When the VIF value is greater than 10, it can be mentioned that there is multicollinearity between the explanatory variables (Albayrak, 2005; Topal et al., 2010).

$$VIF = c_{ij} = \frac{1}{1 - R_j^2}$$

2.1. Ridge Estimator

The ridge estimator, whose main purpose is to obtain more reliable models by eliminating the multicollinearity between the explanatory variables, was proposed by Hoerl and Kennard (1970). In the case of multicollinearity between the explanatory variables, variance and covariances will increase in the X'X matrix. (Vupa and Gurunlu Alma, 2008; Uckardes et al., 2012). To reduce the variance and covariances in the X'X matrix, the bias coefficient k is added to the diagonal elements of the matrix (Tirink et al., 2020). k should be between 0 and 1. If the bias coefficient of k is zero, regression parameter estimation is the same as LS (Uckardes et al., 2012). Regression parameter estimation as a matrix notation is given below with the ridge estimator.

$$\hat{\beta}_{Ridge} = (X'X + kI)^{-1}X'Y$$

It is important to calculate optimum k value. Many researchers suggested a lot of method for calculating the optimum k value. The equation is given below that proposed by Kurtulus (2001) for calculating the optimum bias coefficient of k value based on eigenvalue:

$$k \leq \frac{\lambda_{max} - 100\lambda_{min}}{99}, k \neq 0$$

2.2. Liu Estimator

The Liu estimator, whose main purpose is to eliminate the multicollinearity between explanatory variables. Liu estimator was proposed by Liu (1993). To eliminate the multicollinearity, the Liu estimator was proposed by combining Stein and Ridge estimator (Alpu and Samkar, 2010). Regression parameter estimation as a matrix notation is given below with the Liu estimator.

$$\hat{\beta}_{Liu} = (X'X + I)^{-1}(X'Y + dI)\hat{\beta}_{LS}$$

d is a biasing parameter used to overcome the multicollinearity. d parameter should be between 0 and 1. $\hat{\beta}_{Liu}$ is a linear function of biasing parameter of d so that to calculate d is easier than k (Alpu and Samkar, 2010). The equation is given below that for calculating the optimum biasing parameter of d based on the matrix of eigenvectors (Q) (Alpu and Samkar, 2010):

$$\hat{d} = 1 - \hat{\sigma}^2 \left[\frac{\sum_{i=1}^p \frac{1}{\lambda_i(\lambda_i + 1)}}{\sum_{i=1}^p \frac{\hat{\alpha}_i^2}{(\lambda_i + 1)^2}} \right], \quad \hat{\alpha} = Q'\hat{\beta}_{LS}$$

2.3. Model Selection Criteria

Mean square error (MSE), root mean square error (RMSE), relative root mean square error (rRMSE), mean absolute percentage error (MAPE), determination of coefficient (R^2) and Akaike information criteria (AIC) were used as a model selection criteria as To measure the model accuracy of estimators used model comparison criteria as given below.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$

$$rRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}}{\bar{y}} * 100$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100$$

$$R^2 = \left[1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right]$$

$$AIC = n * \ln \left[\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right] + 2p$$

Within the scope of model selection criteria, the lowest RMSE, MAPE, AIC values and the highest R^2 values were considered for the criteria used in the selection of the best model. (Tatliyer, 2020).

2.4. Ethical Consideration

The experiment design was approved by the Local Animal Care and Ethics Committee of Ondokuz Mayıs University with an approval number of 2018-31.

3. Results and Discussion

In this study, different estimators that are in the presence of multicollinearity were used to estimate the birth weight of Bafra sheep. Descriptive statistics for response and explanatory variables are in Table 1 and Table 2 and correlations among them was given in Figure 1.

3.1. Results of Least Squares

Regression coefficients, its standard error, test statistics and significance level obtained from the LS method are given in Table 3. According to the Table 3, intercept, S, SCW, MCL and LCDe were determined to be statistically significant ($p < 0.05$).

3.2. Results of Ridge Estimator

VIF values for each biasing parameter of k were given in Table 5. In Table 5, the optimum k value was determined as 0.013. Regression coefficients, its standard error, test statistics and significance level obtained from the ridge estimator were given in Table 4. According to the Table 4, S and SCW were determined to be statistically significant ($p < 0.05$).

3.3. Results of Liu Estimator

Regression coefficients, its standard error, test statistics and significance level obtained from the ridge estimator were given in Table 6. According to the Table 6, S, SCW and LCDe were determined to be statistically significant ($p < 0.05$).

Table 1. Descriptive statistics for explanatory variables

		n	Median	Minimum	Maximum
Male	LCN		9	1	39
	MCN	22	22	5	38
	SCN		18	4	66
Female	LCN		8	18	21.50
	MCN	18	0	2	1
	SCN		26	51	36

Table 2. Descriptive statistics for response and explanatory variables

		n	Mean±Standard Deviation	Minimum	Maximum
Male	BW		5.19 ± 0.72	3.50	6.53
	PE		240.99 ± 104.24	2.32	505.71
	TCN		53.55 ± 17.79	26.00	89.00
	LCW		28.36 ± 23.01	6.94	100.00
	MCW		30.57 ± 12.38	10.19	59.00
	SCW		15.43 ± 9.79	1.00	34.00
	LCW _i		28.18 ± 7.75	15.20	47.08
	MCW _i	22	21.9 ± 4.16	16.40	30.22
	SCW _i		15.01 ± 3.89	9.40	23.36
	LCL		39.34 ± 9.04	26.60	61.24
	MCL		28.4 ± 4.86	21.00	37.20
	SCL		19.44 ± 3.58	14.80	27.56
	LCD _e		5.34 ± 1.51	3.00	9.88
	MCD _e		4.85 ± 1.61	3.00	9.81
SCD _e		3.58 ± 1.25	2.10	6.74	
Female	BW		4.23 ± 0.57	3.00	5.25
	PE		249.82 ± 93.07	3.25	418.12
	TCN		49.94 ± 20.52	15.00	74.00
	LCW		28.64 ± 16.52	0.00	64.26
	MCW		28.76 ± 12.16	7.34	49.00
	SCW		20.68 ± 15.54	4.00	70.13
	LCW _i		26.84 ± 7.93	0.00	37.06
	MCW _i	18	22.81 ± 4.34	16.57	31.44
	SCW _i		16.06 ± 3.39	10.30	22.78
	LCL		36.41 ± 11.67	0.00	55.68
	MCL		28.87 ± 5.15	22.72	36.87
	SCL		20.84 ± 3.89	15.19	26.10
	LCD _e		5.01 ± 1.79	0.00	7.71
	MCD _e		5.04 ± 1.73	2.75	9.85
SCD _e		4.43 ± 1.76	2.18	8.60	

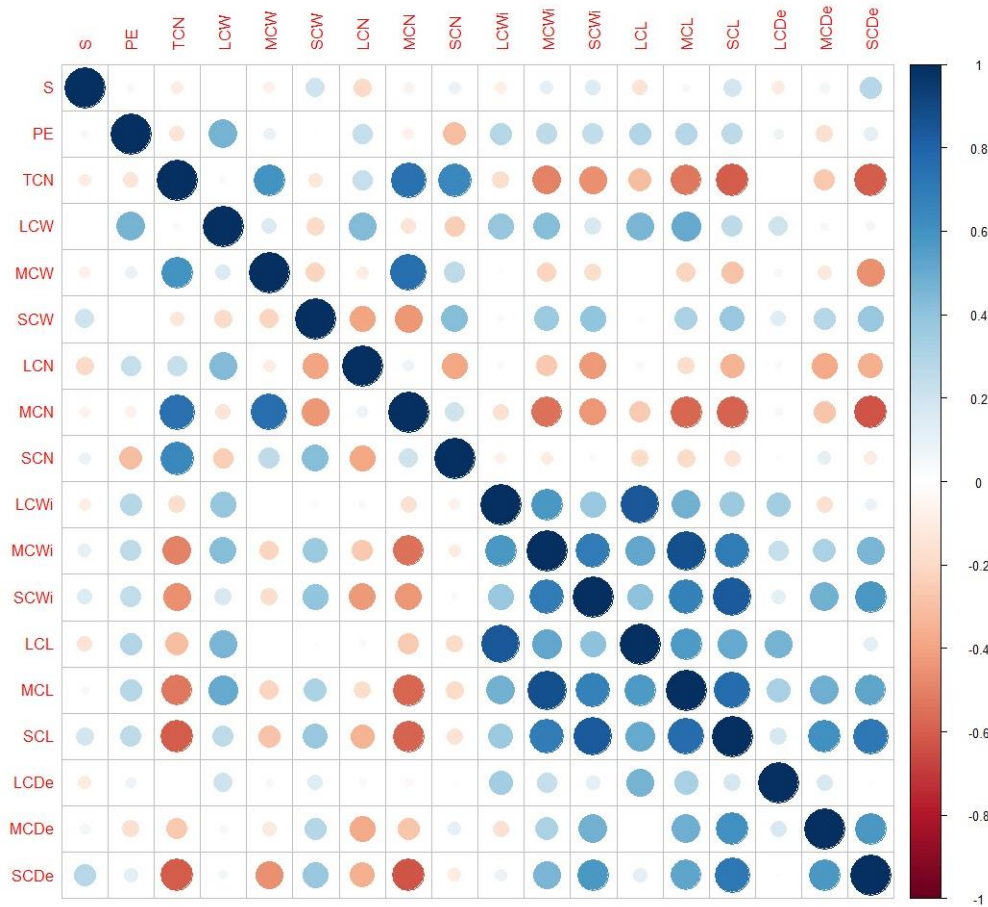


Figure 1. Correlation coefficients for explanatory variables.

Table 3. Results of LS method

	Estimate	Standard Error	t-value	p-value
(Intercept)	7.16	1.57	4.56	0.00
S	-0.98	0.24	-4.14	0.00
PE	0.00	0.00	-1.28	0.21
CN	0.06	0.04	1.50	0.15
LCW	0.02	0.01	1.75	0.09
MCW	-0.01	0.02	-0.80	0.43
SCW	0.03	0.01	2.43	0.02
LCN	-0.06	0.05	-1.18	0.25
MCN	-0.02	0.04	-0.61	0.55
SCN	-0.08	0.05	-1.63	0.12
LCW _i	0.00	0.04	-0.01	0.99
MCW _i	0.06	0.07	0.84	0.41
SCW _i	0.01	0.06	0.17	0.86
LCL	0.02	0.03	0.60	0.55
MCL	-0.12	0.06	-1.89	0.07
SCL	0.00	0.09	-0.02	0.98
LCD _e	-0.18	0.08	-2.19	0.04
MCD _e	0.10	0.13	0.81	0.43
SCD _e	-0.03	0.13	-0.19	0.85

Table 4. Results of Ridge estimator

	Estimate	Standard Error	t-value	p-value
Intercept	7.100	249.908	0.486	0.632
S	-0.985	0.698	-4.439	0.000
PE	-0.002	0.795	-1.307	0.204
CN	0.025	1.941	1.547	0.136
LCW	0.015	1.129	1.718	0.100
MCW	-0.008	1.173	-0.503	0.620
SCW	0.027	0.946	2.312	0.030
LCN	-0.016	1.265	-0.643	0.527
MCN	0.003	1.568	0.138	0.892
SCN	-0.032	1.492	-1.720	0.099
LCW _i	0.002	1.414	0.052	0.959
MCW _i	0.037	1.506	0.652	0.521
SCW _i	0.028	1.193	0.536	0.598
LCL	0.011	1.526	0.443	0.662
MCL	-0.094	1.625	-1.788	0.087
SCL	-0.008	1.642	-0.117	0.908
LCD _e	-0.151	0.752	-2.040	0.053
MCD _e	0.067	1.153	0.594	0.558
SCD _e	-0.033	1.124	-0.278	0.783

Table 5. Optimum k values according to the VIF values

S	PE	CN	LCW	MCW	SCW	LCN	MCN	SCN	LCWI	MCWI	SCWi	LCL	MCL	SCL	LCDe	MCDe	SCDe	
k=0	1.410	1.802	64.484	4.097	4.825	2.716	16.460	20.726	34.292	8.146	9.149	5.035	9.996	10.220	10.278	1.729	4.560	4.079
k=0.001	1.394	1.789	50.950	4.028	4.670	2.677	13.482	17.403	27.244	7.826	8.778	4.837	9.545	9.831	9.922	1.691	4.436	4.004
k=0.002	1.380	1.778	41.321	3.962	4.537	2.641	11.354	15.005	22.228	7.528	8.442	4.678	9.134	9.484	9.597	1.660	4.326	3.932
k=0.003	1.368	1.766	34.225	3.898	4.419	2.608	9.777	13.208	18.533	7.251	8.133	4.544	8.755	9.169	9.297	1.634	4.226	3.863
k=0.004	1.356	1.754	28.846	3.835	4.313	2.577	8.575	11.820	15.730	6.992	7.847	4.428	8.404	8.879	9.018	1.613	4.133	3.798
k=0.005	1.346	1.743	24.670	3.775	4.215	2.548	7.636	10.720	13.555	6.749	7.582	4.325	8.079	8.610	8.755	1.593	4.047	3.736
k=0.006	1.337	1.732	21.363	3.716	4.125	2.520	6.887	9.829	11.831	6.522	7.333	4.232	7.775	8.357	8.508	1.576	3.965	3.676
k=0.007	1.327	1.721	18.698	3.659	4.040	2.493	6.278	9.094	10.443	6.308	7.101	4.147	7.491	8.120	8.273	1.560	3.888	3.619
k=0.008	1.319	1.710	16.520	3.604	3.960	2.467	5.776	8.477	9.307	6.106	6.882	4.068	7.225	7.896	8.051	1.546	3.815	3.564
k=0.009	1.310	1.700	14.716	3.550	3.884	2.442	5.355	7.953	8.366	5.916	6.676	3.994	6.976	7.683	7.839	1.533	3.744	3.511
k=0.01	1.302	1.690	13.205	3.498	3.811	2.417	5.000	7.501	7.578	5.736	6.482	3.925	6.742	7.481	7.637	1.520	3.677	3.461
k=0.011	1.295	1.679	11.926	3.447	3.742	2.393	4.695	7.107	6.911	5.567	6.297	3.859	6.521	7.289	7.445	1.508	3.613	3.412
k=0.012	1.287	1.669	10.834	3.397	3.676	2.370	4.432	6.761	6.340	5.406	6.123	3.797	6.312	7.106	7.260	1.497	3.550	3.365
k=0.013	1.280	1.660	9.895	3.349	3.612	2.348	4.203	6.454	5.849	5.253	5.957	3.737	6.115	6.931	7.084	1.486	3.490	3.319
k=0.25	0.666	0.719	0.340	0.702	0.683	0.742	0.623	0.602	0.497	0.624	0.685	0.801	0.614	0.654	0.658	0.694	0.685	0.813
k=0.3	0.606	0.639	0.291	0.595	0.578	0.643	0.538	0.493	0.436	0.524	0.564	0.669	0.510	0.520	0.521	0.625	0.584	0.684
k=0.4	0.511	0.518	0.228	0.452	0.439	0.502	0.420	0.357	0.349	0.394	0.409	0.492	0.377	0.358	0.356	0.516	0.448	0.507
k=0.5	0.438	0.432	0.188	0.362	0.352	0.408	0.343	0.277	0.291	0.315	0.381	0.297	0.297	0.267	0.263	0.437	0.362	0.395
k=0.6	0.380	0.368	0.160	0.300	0.292	0.340	0.288	0.224	0.249	0.261	0.253	0.305	0.243	0.209	0.205	0.375	0.302	0.318
k=0.75	0.314	0.296	0.130	0.237	0.231	0.269	0.230	0.173	0.204	0.206	0.191	0.230	0.189	0.155	0.151	0.306	0.240	0.241
k=1	0.237	0.219	0.099	0.172	0.169	0.195	0.170	0.123	0.154	0.150	0.131	0.157	0.136	0.105	0.101	0.228	0.175	0.165

Table 6. Results of Liu estimator

	Estimate	Standard Error	t-value	p-value
Intercept	7.07E+00	4.37E+00	1.618	0.1056
S	-9.09E-01	2.14E-01	-4.252	2.12E-05
PE	-1.75E-03	1.33E-03	-1.309	0.1906
CN	6.46E-02	4.11E-02	1.575	0.1153
LCW	1.70E-02	9.81E-03	1.733	0.0831
MCW	-1.37E-02	1.76E-02	-0.775	0.4383
SCW	3.07E-02	1.26E-02	2.444	0.0145
LCN	-5.84E-02	4.82E-02	-1.211	0.2261
MCN	-2.58E-02	3.78E-02	-0.684	0.4942
SCN	-7.59E-02	4.46E-02	-1.701	0.0889
LCW _i	5.49E-05	3.56E-02	0.002	0.9988
MCW _i	5.88E-02	6.95E-02	0.846	0.3976
SCW _i	1.26E-02	5.93E-02	0.213	0.8314
LCL	1.94E-02	2.97E-02	0.653	0.5141
MCL	-1.21E-01	6.25E-02	-1.927	0.0539
SCL	-6.35E-03	8.22E-02	-0.077	0.9384
LCD _e	-1.76E-01	7.81E-02	-2.249	0.0245
MCD _e	1.09E-01	1.23E-01	0.889	0.374
SCD _e	-2.94E-02	1.25E-01	-0.236	0.8133

3.4. Comparison of the Estimators

If the assumptions are not provided in the model estimates made with LS, the reliability of the model estimation decreases. In this study, the Ridge and Liu estimator, which is one of the proposed estimators, was used as an alternative method to the LS in cases where the multicollinearity assumption, which is one of the assumptions, was not provided. MSE, RMSE, rRMSE, MAPE, R², and AIC were used as model comparison

criteria. According to the model comparison criteria used in Table 7, it is possible to interpret the best model for the lowest MSE, RMSE, rRMSE, MAPE, R², and AIC and the highest R² value (Tatliyer, 2020). In table 7 examined, Liu estimator, used as an alternative estimator to LS, has the lowest MSE, RMSE, rRMSE, MAPE and AIC and the highest R² value. Tirink et al. (2019) were determined similar results.

Table 7. Model comparison criteria

Model Comparison Criteria	Least Squares	Ridge Estimator	Liu Estimator
Mean square error (MSE)	0.205	0.214	0.206
Root mean square error (RMSE)	0.453	0.463	0.454
Relative root mean square error (rRMSE)	9.51	9.726	9.53
Mean absolute percentage error (MAPE)	7.522	7.785	7.508
Coefficient of determination (R ²)	0.679	0.664	0.677
Akaike's information Criterion (AIC)	-59.398	-57.602	-59.236

4. Conclusion

In the evaluation of the data for agricultural studies according to cause and effect relationships, in the case of multicollinearity between the explanatory variables, Ridge and Liu estimators were used as an alternative to LS. For this purpose, it was aimed to estimate birth weight by using placental features in Bafra sheep. The results showed that Liu estimator is more reliable model estimation than LS and Ridge estimator. On account of in

case of the multicollinearity, the Liu estimator is an alternative to LS.

Conflict of interest

The author declared that there is no conflict of interest.

References

Albayrak AS. 2005. An alternative bias estimation technique and an application of the least-squares technique in multiple

- linear connections. Zonguldak Karaelmas Univ J Soc Sci, 1: 105-126.
- Alkass JE, Merkhan KY, Hamo RAH. 2013. Placental traits and their relation with birth weight in Meriz and Black goats. *Sci J Anim Sci*, 2: 168-172.
- Alpu O, Samkar H. 2010. Liu estimator based on an m estimator. *Turkiye Klinikleri J Biostat*, 2 (2): 49-53.
- Ari A, Onder H. 2013. Regression models used for different data structures. *Anadolu J Agr Sci*, 28 (3): 168-174.
- Brzozowska A, Wojtasiak N, Błaszczyk B, Stankiewicz T, Wieczorek-Dąbrowska M, Udała J. 2020. The effects of non-genetic factors on the morphometric parameters of sheep placenta and the birth weight of lambs. *Large Anim Rev*, 26: 119-126.
- Cankaya S, Eker S, Abaci SH. 2019. Comparison of Least Squares, Ridge Regression and Principal Component Approaches in the Presence of Multicollinearity in Regression Analysis. *Turkish J Agriculture-Food Sci and Tech*, 7(8): 1166-1172. DOI: 10.24925/turjaf.v7i8.1166-1172.2515.
- Dwyer CM, Calvert SK, Farish M, Donbavand J, Pickup HE. 2005. Breed, litter and parity effects on placental weight and placental number and consequences for the neonatal behavior of the lamb. *Theriogenology*, 63: 1092-1110.
- Echternkamp SE. 1993. Relationship between placental development and calf birth weight in beef cattle. *Anim Reprod Sci*, 32; 1-13.
- Eyduran E. 2020. ehaGoF: calculates goodness of fit statistics. R package version 0.1.1. URL: <https://CRAN.R-project.org/package=ehaGoF> (erişim tarihi: 17.08.2020).
- Hoerl AE, Kennard RW. 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12 (1): 55-67. DOI: 10.2307/1267351.
- Imdad MU, Aslam M. 2018a. lmr: linear ridge regression with ridge penalty and ridge statistics. URL: <https://CRAN.R-project.org/package=lmr>, R package version 1.2. (erişim tarihi: 17.08.2020).
- Imdad MU, Aslam M. 2018b. liureg: liu regression with liu biasing parameters and statistics. URL: <https://CRAN.R-project.org/package=liureg>, R package version 1.1.2. (erişim tarihi: 15.08.2020).
- Konyali A, Tölü C, Daş G, Savaş T. 2007. Factors affecting placental traits and relationships of placental traits with neonatal behaviour in goat. *Anim Reprod Sci*, 97: 394-401.
- Kurtulus M. 2001. A study on ridge regression. Master Thesis. Gazi University, Graduate School of Natural and Applied Sciences, p. 42, Ankara.
- Kutner MH, Nachtsheim CJ, Neter J, Li W. 2004. *Applied linear statistical models*. 5th edition, McGraw-Hill/Irwin, New York.
- Liu K. 1993. A new class of biased estimate in linear regression. *Comm Statist Theory Methods*, 22(2): 393-402.
- Onder H, Abaci SH. 2015. Path analysis for body measurements on body weight of saanen kids. *Kafkas Univ Vet Fak Derg*, 21(3): 351-354, DOI: 10.9775/kvfd.2014.12500.
- Osgerby JC, Gadd TS, Wathes DC. 2003. The effects of maternal nutrition and body condition on placental and foetal growth in the ewe. *Placenta*, 24: 236-247. DOI: 10.1053/plac.2002.0902.
- Ozyurek S, Turkyilmaz D. 2020. Determination of relationships between placental characteristics and birth weight in Morkaraman sheep. *Arch Anim Breed*, 63: 39-44. DOI: 10.5194/aab-63-39-2020.
- Ozyurek S. 2019. Investigation of the relationship between kids vitality and placental characteristics in Hair goats. *Large Anim Rev*, 25: 173-177.
- R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Sarstedt M, Mooi E, 2014. A concise guide to market research. Heidelberg: Springer Texts in Business and Economics. Chapter 7: Regression Analysis; p. 193-233. DOI: 10.1007/978-3-642-53965-7_7, #Springer-Verlag Berlin Heidelberg.
- Sen U, Sirin E, Kuran M. 2013. The effect of maternal nutritional status during mid-gestation on placental characteristics in ewes. *Animal Reprod Sci*, 137: 31-36.
- Sen U, Sirin E, Onder H. 2019. Establishment of effective placental characteristics on birth weight in karayaka sheep breed with path analysis. In: *International Congress on Domestic Animal Breeding, Genetics and Husbandry 2019 (ICABGEH-19)*; 2019 September 11-13; Prague, CZECHIA.
- Sen U, Onder H. 2016. Poor placental traits reduce kid birth weight in young Saanen dams at the first parity. *Turkish J Vet and Anim Sci*, 40: 554-561.
- Tatliyer A. 2020. The effects of raising type on performances of some data mining algorithms in lambs. *KSU J Agric Nat*, 23(3): 772-780. DOI:10.18016/ksutarimdogu.vi.651232.
- Tirink C, Onder H, Kurnaz FS. 2019. Comparison of least square method, ridge and liu estimators in saanen goats. In: *International Congress on Domestic Animal Breeding Genetics and Husbandry 2019*; 2019 September 11-13, Prague, CZECHIA. p. 130.
- Tirink C, Abacı SH, Önder H. 2020. Comparison of ridge regression and least squares methods in the presence of multicollinearity for body measurements in saanen kids. *J Institute of Sci and Tech*, 10(2): 1429-1437, ISSN: 2146-0574, eISSN: 2536-4618.
- Topal M, Eyduran E, Yaganoglu AM, Sonmez AY, Keskin S, 2010. Use of ridge and principal component regression analysis methods in multicollinearity. *J Agri Fac of Atatürk Univ*, 41(1): 53-57. ISSN: 1300-9036.
- Uckardes F, Efe E, Narinc D, Aksoy T. 2012. Estimation of the egg albumen index in the Japanese quails with ridge regression method. *Akademik Ziraat Derg*, 1 (1): 11-20.
- Vupa O, Gurunlu Alma O. 2008. Investigation of multicollinearity problem in small samples included outlier value in linear regression analysis. *Selcuk Univ J Sci Fac*, 31: 97-107.



Araştırma Makalesi (Research Article)

Cilt 3 - Sayı 4: 146-150 / Ekim 2020
(Volume 3 - Issue 4: 146-150 / October 2020)

BALIKESİR, DÜZCE VE KÜTAHYA'DAKİ ANADOLU MANDALARINDA BAZI ÇEVRE FAKTÖRLERİNİN BÜYÜME VE LAKTASYON ÖZELLİKLERİNE ETKİSİ

Ebru ERGÜNEŞ BERKİN^{1*}, Mustafa KÜÇÜKKEBAPÇI², Cüneyt KAPTAN³

¹İslahiye Tarım ve Orman İlçe Müdürlüğü, 27800, İslahiye, Gaziantep, Türkiye

²Karabük Tarım ve Orman İl Müdürlüğü, 78050, Karabük, Türkiye

³Koyunculuk Araştırma Enstitüsü Müdürlüğü, 10245, Bandırma, Balıkesir, Türkiye

Gönderi: 19 Ağustos 2020; **Kabul:** 14 Eylül 2020; **Yayınlanma:** 01 Ekim 2020

(Received: August 19, 2020; **Accepted:** September 14, 2020; **Published:** October 01, 2020)

Özet

Bu çalışmada, 2013-2017 yılları arasında Balıkesir, Düzce ve Kütahya'da yürütülen Anadolu Mandasının Halk Elinde Islahı Ülkesele Projesinden elde edilen veriler kullanılarak doğum, 6. ay ve 12. ay canlı ağırlıkları değerlendirilmiştir. Yıl heterojenite kaynağı, il faktör, cinsiyet ise kovaryet olarak kullanılmıştır. Yıl, il ve kovaryet ilgilenilen ağırlıklar üzerinde istatistiksel olarak anlamlı bulunmuştur ($P < 0,001$). 12. ay canlı ağırlık üzerine ilk 3 yılın etkisi bulunmazken, son iki yılda olan artış ise, istatistiksel olarak anlamlı ($P < 0,001$) bulunmuştur. Aynı illere ait toplam 5721 laktasyon verisi değerlendirilmiş olup il, doğum mevsimi ve yılların laktasyon süt verimi ve laktasyon süresi üzerine etkisi incelenmiş ve tüm interaksiyonların istatistiksel olarak anlamlı olmadığı görülmüştür. Bu durumda sadece ana etkiler yorumlanmıştır. Doğumların %40,61'i ilkbahar, %37,39'u yaz %22,01' i ise diğer aylarda gerçekleşmiştir. İl, doğum mevsimi ve yılların, laktasyon süt verimi ve laktasyon süresi üzerine etkisi istatistiki olarak önemli ($P < 0,001$) bulunmuştur. Laktasyon süt verimi ve laktasyon süresi en uzun olan mevsim, kışın doğum yapan mandalardadır. Bu mandalarda, laktasyonun başında süt verimi düşükken, ilkbaharın gelmesi ile birlikte meranın canlanması, süt verimini arttıran faktörlerden biridir. Yıllar itibariyle laktasyon süt verimi ve süresinde dalgalanmalar görülmektedir. Bu çalışmada bazı çevre faktörlerinin, büyüme özellikleri, laktasyon süt verimi ve süresi üzerine etkisi incelenmiştir. Bu faktörlerin kullanılması ile birlikte oluşturulması muhtemel seleksiyon programının, etkilerinin olumlu yönde olacağı düşünülmektedir. İlerde yapılacak çalışmalarda, diğer verim ve performans parametrelerinin eklenmesi suretiyle de yeni seleksiyon programları da oluşturulabilir.

Anahtar kelimeler: Anadolu Mandası, Bazı büyüme özellikleri, Laktasyon

Effects of Some Environmental Factors on the Some Growth and Lactation Characteristics of Anatolian Water Buffalo in Balıkesir, Düzce and Kütahya


Abstract: In this study, birth, 6th month and 12th month live weights and lactation characteristics were evaluated by using the data obtained from the National Project of Breeding in the Hand of the Anatolian Water Buffalo carried out in


Balıkesir, Düzce and Kütahya between 2013 and 2017. Year was used as the source of heterogeneity, province was as factor and gender as covariate. Year, province and co-variable were found statistically significant on the weights of interest ($P<0.001$). There was no effect of the first 3 years on 12th month live weight and the increase in the last two years was statistically significant ($P <0.001$). 5721 lactation data of the same provinces were evaluated. The effect of province, birth season and years on lactation milk yield and lactation period was examined and results showed that all interactions were not statistically significant. In that case only main effects have been interpreted. 40.61% of births were in spring, 37.39% in summer and 22.01% in other months. The effect of birth season on lactation milk yield and lactation period was statistically significant ($P<0.001$). In buffaloes giving birth in winter, lactation milk yield is higher and lactation period is longer. In these buffaloes, milk yield at the beginning of lactation is low, but the recovery of pasture with the arrival of spring is one of the factors that increase milk yield. Fluctuations in lactation milk yield and duration have been observed over the years. As a result; the effects of some environmental factors on growth characteristics, lactation milk yield and duration were investigated. With the use of these factors, it is thought that the selection program that will be formed together will have positive effects. Also; In future studies, new selection programs can be created by adding other efficiency and performance parameters.


Keywords: Anatolian Water Buffalo, Some growth characteristics, Lactation

*Corresponding author: İslahiye Tarım ve Orman İlçe Müdürlüğü, 27800, İslahiye, Gaziantep, Türkiye

E mail: ebru_berkin01@hotmail.com (E. ERGÜNEŞ BERKİN)

Ebru ERGÜNEŞ BERKİN  <https://orcid.org/0000-0003-0673-2627>

Mustafa KÜÇÜKKEBAPÇI  <https://orcid.org/0000-0003-3750-4328>

Cüneyt KAPTAN  <https://orcid.org/0000-0001-9867-7025>

Cite as: Ergüneş Berkin E, Küçükkebaççı M, Kaptan C. 2020. Effects of some environmental factors on the some growth and lactation characteristics of Anatolian Water Buffalo in Balıkesir, Düzce and Kütahya. *BSJ Eng Sci*, 3(4): 146-150.

1. Giriş

Dünyadaki mandaların %57'si Hindistan'da (112568781 baş) bulunmaktadır. Bu ülkeyi sırası ile Pakistan, Çin, Nepal takip etmektedir. Manda varlığı bakımından Türkiye ise 19. sırada yer almaktadır (142073 baş). Sağılan manda sayısı ve yıllık süt üretimi açısından yine Hindistan, Pakistan ve Çin ilk sıralardadır. Türkiye ise 63329 baş sağılan manda ile 12. sırada yer almakta ve yılda 63085 ton süt elde edilmektedir (FAO, 2016).

Türkiye'de 1998 yılında 176.000 baş olan manda mevcudu, 2009 yılında 87726 başa kadar hızlı bir şekilde düşmüştür (TÜİK, 2016). İlk etapta bu düşüşün önüne geçmek, manda yetiştiriciliğini yaygınlaştırmak, laktasyon süt verimi (LSV) ve laktasyon süresinin (LS) artırılmasını, verim ve performans parametrelerini ortaya koymak gibi amaçlarla, Tarım ve Orman Bakanlığı tarafından, 'Anadolu Mandasının Halk Elinde İslahı Ülkesel Projesi' kapsamında, 19 ilde alt projeler olarak halen yürütülmektedir. Proje kapsamına mümkün olduğunca çok hayvan alınmaya çalışılmıştır. Proje başlangıcı olan 2010 yılından 2018 yılına kadar olan süreçte manda mevcudu ve süt üretimi iki katından fazla artmıştır (TÜİK, 2016).

Bu çalışmanın amacı, Balıkesir, Düzce ve Kütahya illerinde yürütülen bu projelerin elde edilen verilerin değerlendirilmesiyle, canlı ağırlık (CA), LSV ve LS üzerine etkili bazı çevre faktörlerinin etkisini incelemektir.

2. Materyal ve Metot

Bu çalışma, 'Anadolu Mandasının Halk Elinde İslahı Ülkesel Projesi'nin alt projesi olan, Balıkesir, Düzce ve Kütahya'da illerinde yürütülen 2013-2017 yılları arasında elde edilen, toplam 6714 büyüme ve 5721

laktasyon kaydı içermektedir. Elde edilen veriler bahsi geçen proje kapsamında kayıt tutmak amacıyla kullanılan Manda Yıldızı programından (Tekerli, 2015) alınmıştır. Veriler çevre faktörleri bakımından birbirlerine benzer özellikte ve farklı işletmelerden alınmıştır. Bu işletmelerde, manda sürüsü sabah meraya çıkmakta, akşam ise dönmektedir. Doğum yapan mandalar ise 3-4 gün sonra meraya salınmaktadır. Emzirme, sabah-akşam olmak üzere günde iki kez yapılmaktadır. Yetiştirici tercihinin göre malaklar ise 6. aydan sonra anaları ile birlikte meraya çıkmaktadır.

Doğum ağırlığı (DA), 6. ay ağırlığı (6CA) ve 12. ay canlı ağırlıkların (12CA) değerlendirildiği ilk aşamada yıl heterojenite kaynağı, il faktör, cinsiyet ise kovaryet olarak kullanılmıştır. İstatistik analiz ise tesadüf bloklarında kovaryans analizi kullanılarak yapılmıştır. İkinci aşamada ise, il, doğum mevsimi ve yılların LSV ve LS üzerine etkisi, tesadüf bloklarında faktöriyel deneme deseni ile analiz edilmiştir.

Tüm analizler SPSS 23.V. (SPSS, 2015). versiyonunda yapılmıştır. Gruplar arasındaki farklılıkların analizi için Duncan çoklu karşılaştırma testi kullanılmıştır.

2.1. Etik beyan

Bu çalışmada hayvanlara müdahale edilmemiş olup sadece proje kapsamındaki Manda Yıldızı programında kayıtlı veriler kullanıldığından etik kurul iznine ihtiyaç bulunmamaktadır.

3. Bulgular

Üç ilin farklı yaşlardaki CA üzerine etkisi istatistik olarak anlamlı bulunmuştur. En yüksek CA Balıkesir'de gerçekleşirken, Düzce ilinde doğum ağırlığı Kütahya'ya göre düşük, ancak 6CA yüksek bulunmuştur. Kütahya'nın ise 12CA, Düzce'nin 12CA'sına göre yüksek bulunmuştur.

Bu yaşlar arasındaki farklılıkların tamamı istatistiksel olarak önemli bulunmuştur (Tablo 1). Ayrıca; Tablo 1'e göre DA yıllar itibariyle azalırken, 6CA artmaktadır. Yılların, DA ve 6CA üzerine etkisi önemli

bulunmuştur. 12CA üzerine ilk 3 yılın etkisi istatistiksel olarak önemli bulunmazken, son iki yılda olan artış ise, istatistiksel olarak anlamlıdır ($P<0,001$).

Tablo 1. Farklı yaşlardaki malakların canlı ağırlıkları üzerine ilin ve yılın etkisi

Genel ortalama	n	DAort $\pm S_{\bar{X}}$	n	6CAort $\pm S_{\bar{X}}$	n	12CAort $\pm S_{\bar{X}}$
	6,714	29,74 \pm 0,07	4,845	99,66 \pm 0,31	3,561	164,42 \pm 0,57
İl						
Balıkesir	2,987	31,71 \pm 0,10 ^a	1,557	103,71 \pm 0,58 ^a	742	179,84 \pm 1,25 ^a
Düzce	1,540	27,98 \pm 0,15 ^c	1,524	96,71 \pm 0,65 ^b	1,491	150,41 \pm 1,06 ^c
Kütahya	2,187	28,72 \pm 0,13 ^b	1,764	94,18 \pm 0,61 ^c	1,328	163,74 \pm 1,08 ^b
		p<0,001		p<0,001		p<0,001
Yıl						
2013	1,047	30,48 \pm 0,18 ^a	781	92,71 \pm 0,83 ^d	468	154,39 \pm 1,59 ^b
2014	888	30,62 \pm 0,19 ^a	602	92,69 \pm 0,89 ^d	487	151,11 \pm 1,49 ^b
2015	1,086	29,54 \pm 0,17 ^b	774	96,61 \pm 0,81 ^c	653	155,83 \pm 1,35 ^b
2016	1,670	28,42 \pm 0,14 ^c	1,244	101,89 \pm 0,66 ^b	851	182,78 \pm 1,21 ^a
2017	2,023	28,27 \pm 0,13 ^c	1,444	107,09 \pm 0,61 ^a	1,102	179,20 \pm 1,12 ^a
		p<0,001		p<0,001		p<0,001

DAort: Doğum ağırlığı ortalaması, 6CAort: 6. ay canlı ağırlık ortalaması, 12CAort: 12. ay canlı ağırlık ortalaması.

Tablo 2'de ise toplam 5721 laktasyon verisi değerlendirilmiş ve tüm interaksyonlar istatistiksel olarak anlamlı olmadığından sadece ana etkiler yorumlanmıştır.

İller arasındaki farklılıklar istatistik olarak önemli bulunmuştur. LSV ve LS en uzun olan il Balıkesir olarak gözlemlenmiştir. Balıkesir'i Düzce ve Kütahya takip etmektedir.

Doğumların %40,61'i ilkbahar, %37,39'u yaz, %22,01'i ise diğer aylarda gerçekleşmiştir. Doğum mevsiminin LSV ve LS üzerine etkisi istatistiksel olarak önemli ($P<0,001$) bulunmuştur.

Yılların LSV ve LS'ne etkisi istatistik olarak önemlidir ve yıllar itibariyle veri dalgalanmaları Tablo 2'de görüldüğü gibidir. 2014-2015 yıllarında, LSV ve LS'nin düşmektedir. 2015-2016 yıllarında ise, LSV artarken, LS ise azalmıştır, 2017 yılında ise LSV ve LS diğer yıllara göre azalmıştır ve bu azalmalar istatistiksel olarak önemlidir ($P<0,001$).

4. Tartışma ve Sonuç

Bu çalışma ile malakların DA ortalaması 29,74 \pm 0,07 kg, ve 6CA ortalaması ise 99,66 \pm 0,31 kg, 12CA ortalaması ise 164,42 \pm 0,57 kg bulunmuştur. DA ortalaması Kul ve ark.'nın (2018) Amasya'daki Anadolu mandalarında bulunduğu değerden yüksek (29,30 \pm 0,43 kg), Uğurlu ve ark.'nın (2016) Giresunda ki Anadolu Mandalarında bulunduğu değerden düşük (26,95 \pm 0,25 kg), Pandya ve ark. (2015) Surti mandalarında bulunduğu bu üç değerlerden sırasıyla daha yüksektir (24,64 \pm 0,50 kg), (73,41 \pm 1,65 kg), (130,50 \pm 2,52 kg). Bu çalışmada, farklı

yaşlardaki CA'nın istatistiksel olarak anlamlı olması, proje başlangıcında yetiştiricilerin yeterince bilinçli olmaması sebebi ile tartımların sağlıklı yapılamaması, proje kotasındaki değişimlerin etkisi ve bakım ile besleme koşullarının yıllar itibariyle değişiminden kaynaklanabilir. Yine aynı şekilde diğer çalışmalar ile olan farklılığın kaynağı olarak da ırk, farklı bakım ve besleme programları gibi faktörlerin etkisinin olabileceği düşünülmektedir.

LSV ortalaması 906,93 \pm 4,56 kg, LS ortalaması ise 210,34 \pm 0,49 gün bulunmuştur. Soysal ve ark., (2018) İstanbul'daki Anadolu Mandalarında yaptıkları çalışmayla il, doğum mevsimi ve yılın, LSV ve LS üzerine etkisini de aynı şekilde önemli bulmuştur. Ancak bu değerler daha yüksektir (1223 \pm 6,83 kg ve 210,34 \pm 0,49 gün).

Tekerli ve ark., (2016) farklı orjinli Anadolu Mandalarında yaptıkları çalışmada ise bölgenin etkisi önemli iken, yıl ve doğum mevsimi etkisi önemsiz, fakat bu çalışmaya kıyasla yine daha yüksek bulmuştur. (1106,077 kg. ve 229,43 gün). Sidgel ve ark., (2015), Nepal'in Kaski bölgesinde Murrah Mandalarında yaptıkları çalışmada, doğum mevsiminin etkisini önemsiz bulmuştur. Bu mandaların LSV, 1838,45 \pm 32,32 kg iken, LS ortalaması ise 362,77 \pm 4,49 gündür. Aynı zamanda benzer şekilde en yüksek süt verimine kışın mevsiminde doğum yapan mandaların sahip olduğunu belirtmiştir (374,09 \pm 4,96 kg). Rathod ve ark., (2018), Maharastra'da Surti Mandalarında yaptıkları çalışmada doğum mevsiminin etkisini önemsiz bulmuştur. LSV ortalaması, bu çalışmada bulunan ortalamadan düşük iken (898,80 \pm

25,06 kg.), ortalama LS ise daha uzundur (254,30 ± 4,76 gün). Thiruvenkadan ve ark., (2014), Hindistan'ın Tamil Namu eyaletindeki Murrah Mandalarının süt verimi ve üreme performansları üzerine yaptıkları çalışmada,

doğum mevsimin etkisini önemli bulmuşlar, en düşük LSV kuzey batı muson mevsiminde iken (1801,60 ± 19,40 kg), kış, yaz ve kuzey batı muson mevsimleri arasında fark bulunmadığını belirtmiştir.

Tablo 2. Laktasyon süt verimi ve süresi üzerine ilin, doğum mevsiminin ve yılın etkisi

	n	LSVort ± S _x	LSort ± S _x
Genel ortalama	5,721	906,93 ± 4,56	210,34 ± 0,49
İl			
Balikesir	1,806	1150,80 ± 7,84 ^a	229,14 ± 0,91 ^a
Düzce	1,921	780,94 ± 7,12 ^c	199,38 ± 0,79 ^c
Kütahya	1,994	807,43 ± 4,92 ^b	203,86 ± 0,70 ^b
		p<0,001	p<0,001
Mevsim			
İlkbahar	2,323	907,10 ± 6,77 ^c	214,14 ± 0,74 ^b
Kış	398	1023,11 ± 17,69 ^a	220,80 ± 1,94 ^a
Sonbahar	861	942,47 ± 13,06 ^b	208,46 ± 1,47 ^c
Yaz	2,139	870,82 ± 6,76 ^d	205,01 ± 0,75 ^c
		p<0,001	p<0,001
Yıl			
2013	1,262	864,21 ± 9,90 ^d	217,04 ± 0,99 ^{ab}
2014	890	948,82 ± 11,52 ^b	219,13 ± 1,10 ^a
2015	909	921,09 ± 10,93 ^c	215,36 ± 1,30 ^b
2016	1,474	978,03 ± 9,21 ^a	212,05 ± 0,99 ^c
2017	1,186	821,73 ± 7,41 ^e	190,63 ± 0,95 ^d
		p<0,001	p<0,001

LSVort :Laktasyon süt verimi ortalaması, LSort: Ortalama laktasyon süresi

Doğum mevsiminin ise ortalama LS üzerine etkisini önemsiz bulmuştur. Bu çalışmada ise; LSV ve LS en uzun olan mevsim, kışın doğum yapan mandalardadır (1023,11 ± 17,69 kg). Şahin ve Ulutaş (2014), yaptıkları çalışmada; kış mevsiminde doğum yapan mandaların yaza göre, sonbahardakilerin ise ilkbahara göre daha fazla LSV'ne sahip olduğunu, ayrıca kışın doğum yapan mandaların yaza, sonbaharda doğum yapanların ise ilkbahara nazaran daha uzun LS'ne sahip olduklarını bildirmiştir. Bu çalışmada ise LSV bakımından aynı durum söz konusu iken; İlkbaharda doğum yapan mandalar, sonbahara nazaran daha uzun LS'ne sahiptir ve ayrıca LS bakımından yaz ve sonbaharda doğum yapan mandaların aralarındaki fark istatistik olarak anlamlı değildir. Kışın doğum yapan mandaların, diğer mevsimlere göre LSV'nin yüksek olması, gün ışığı süresinin az olması, üreme hormonları düzeylerinin düşük olması (inaktif ovaryumlar), östrus görülme sıklığının azalması, buna karşın prolaktin hormon düzeyinin yüksek olması, laktasyonun başındaki süt veriminin düşüken, ilkbaharın gelmesi ile birlikte meranın canlanması ile ortaya çıkan bir durum olabilir.

Diğer çalışmalar ile olan farklılığın kaynağı olarak da ırk, sıcaklık, bakım ve besleme koşullarındaki farklılıklar gibi etkenler gösterilebilir.

LSV ve LS bakımından Tablo 2'de görülen dalgalanmaların; ilk etapta projede yer alan mandaların bireysel süt veriminde, her sene projeye dahil edilen veya projeden çıkarılan hayvanların (ölüm, kesim vd.) laktasyon sıraları ve ayrıca işletmelerin bakım ve besleme koşullarındaki değişimlerin etkili olduğu düşünülmektedir.

Bu çalışmada bazı çevre faktörlerinin, büyüme özellikleri, LSV ve LS üzerine etkisi incelenmiştir. Bu faktörlerin kullanılması ile birlikte oluşturulması muhtemel seleksiyon programının, etkilerinin olumlu yönde olacağı düşünülmektedir. İlerde yapılacak çalışmalarda, diğer verim ve performans parametrelerinin eklenmesi suretiyle de yeni seleksiyon programları da oluşturulabilir.

Çıkar İlişkisi

Yazarlar bu çalışmada hiçbir çıkar ilişkisi olmadığını

beyan etmektedirler.

Teşekkür ve Bilgilendirme

Tarım ve Orman Bakanlığı, Tarımsal Araştırmalar Genel Müdürlüğü (TAGEM)' ne, Anadolu Mandasının Halk elinde Islahı Ülkesel Projesi'nin Balıkesir, Düzce, Kütahya illerinde yürütülen alt projelerine vermiş olduğu destekten dolayı teşekkürü borç biliriz (Proje; 10MAN-2011-01, 81MAN-2011-01 ve 43MAN-2012-01). Ayrıca Balıkesir, Düzce ve Kütahya ili Damızlık Manda Yetiştiricileri Birliği'ne teknik desteklerinden dolayı teşekkür ederiz.

Kaynaklar

- FAO. 2016. World Water Buffalo Statistics. URL: <http://www.fao.org/faostat/en/#data/QA>, (erişim tarihi: 26.06.2019).
- Kul E, Filik G, Şahin A, Çayıroğlu H, Uğurlutepe E, Erdem H. 2018. Effects of some environmental factors on birth weight of Anatolian buffalo calves. *TURJAF*, 6(4): 444-446.
- Pandya GM, Joshi CG, Rank DN. 2015. Genetic analysis of body weight traits of surti buffalo. *Buf Bu*, 34(2): 189-195.
- Rathod AS, Vaidya MS, Ali SS. 2018. Genetic Studies of productive and reproductive attributes of surti buffalo in Maharashtra. *IJLR*, 8(18): 309-314.
- Sigdel A, Bhattarai N, Kolachhapati MR, Paudyal S. 2015. Estimation of genetic parameters for productive traits of murren buffaloes in Kaski, Nepal. *IJR*, 2(5): 1-11.
- Soysal Mİ, Genç S, Aksel M, Ünal EÖ, Gürcan EK. 2018. Effect of environmental factors on lactation milk yield, lactation length and calving interval of Anatolian buffalo in Istanbul. *JASP*, 1(1): 93-101.
- SPSS. 2015. SPSS for Windows. Version 23.00. Chicago.
- Şahin A, Ulutaş Z. 2014. Anadolu mandalarının değişik metotlara göre tahmin edilen süt verimleri üzerine bazı çevresel faktörlerin etkilerinin belirlenmesi. *Kafkas Univ Vet Fak Derg*, 20: 79-85.
- Tekerli M, Altuntaş A, Birdane F. 2016. Farklı bölge orijinli Anadolu mandalarından oluşturulan bir sürüde verim özellikleri, beden ölçüleri ve biyokimyasal polimorfizm yönünden ıslah olanaklarının karşılaştırmalı belirlenmesi: laktasyon özellikleri ve genetik polimorfizm. *Lalahan Hay Araş Enst Derg*, 56(1): 7-12.
- Tekerli M. 2015. Manda Yıldızı Veri Kayıt, Hesap ve Proje Takip Programı. Afyon Kocatepe Üniversitesi. Afyonkarahisar.
- Thiruvenkadan AK, Panneerselvam S, Murali N, Selvam S, Saravanakumar VR. 2014. Milk production and reproduction performance of Murrah buffaloes of Tamil Nadu, India. *Buf Bu*, 33(3): 291-300.
- TÜİK. 2016. Yıllara göre hayvan varlığı istatistikleri. URL: <https://biruni.tuik.gov.tr/medas/?kn=101&locale=tr>, (erişim tarihi: 26.06.2019).
- Uğurlu M, Kaya İ, Saray M. 2016. Effects of some environmental factors on calf birth weight and milk yield of Anatolian water buffalo (*Bubalus Bubalis*). *BJAS*, 22(6): 995-998.



Araştırma Makalesi (Research Article)

Cilt 3 - Sayı 4: 151-159 / Ekim 2020
(Volume 3 - Issue 4: 151-159 / October 2020)

MUDURNU'DA SÜRDÜRÜLEBİLİR TURİZM GELİŞİMİNDE ETNOBOTANİK

Nail HOŞCAN^{1*}, Hamdi ZENGİNBAL²

¹Bolu Abant İzzet Baysal Üniversitesi, Mudurnu Meslek Yüksekokulu, Yönetim ve Organizasyon Bölümü, Mudurnu 14800 Bolu, Türkiye

²Bolu Abant İzzet Baysal Üniversitesi, Bolu Teknik Bilimler Meslek Yüksekokulu, Park ve Bahçe Bitkileri Bölümü, Merkez 14300 Bolu, Türkiye

Gönderi: 31 Ağustos 2020; **Kabul:** 15 Eylül 2020; **Yayınlanma:** 01 Ekim 2020

(Received: August 31, 2020; **Accepted:** September 15, 2020; **Published:** October 01, 2020)

Özet

Etnobotanik, yerel halk tarafından kullanılan, korunan ve yönetilen bitkilerin çalışmasıdır. Yerel halkın gastronomik, medikal, dinsel ve diğer amaçlarla bitkilerin kullanımına yoğunlaşmaktadır. Etnobotanik bir yandan, çeşitli bitkilerin potansiyel ekonomik değerini belirleyen diğer yandan da sürdürülebilir turizm pazarı ile bağı olan hızla büyüyen bilimsel bir alandır. Destinasyonlarda yerel halkın kültürü ve kültürel mirası, yerli ve yabancı turistler için önemli bir çekim ögesidir. Günümüzdeki ve yakın gelecekteki destinasyonların, mevcut bir gastronomik kültür taşıdığı varsayıldığından sürdürülebilir turizm olgusunda; yenilebilir bitki türlerinin ve geleneksel etnobotanik bilgisinin kullanılabilmesi açıkça görülmektedir. Son yıllarda, sürdürülebilir turizm anlayışının ortaya çıkmasıyla, yenilebilir bitki türlerinin geleneksel kullanımı ile gastronomi sunumları arasındaki ilişkiyle ilgili olabildiğince çok bilgi toplanması için farklı bilimsel çalışmalar yapılmıştır. Bu çalışmada, Mudurnu'da sürdürülebilir turizm faaliyetleri için özel ilgi gören geleneksel besin ve yerel bitki kaynakları incelenmiş ve özellikle köylerde yaşayan kişilerle ve yöre halkıyla görüşmeler yapılmıştır. Bu çalışma için Mudurnu, ekoturizm arzının ana öğeleri olan doğal, tarihsel ve kültürel kaynaklarının zenginliği nedeniyle tercih edilmiştir. Ayrıca, yöresel gastronomi alanında geleneksel etnobotanik bilgisinin kullanımına ve bunun sürdürülebilir turizm gelişimine olan etkisine özellikle dikkat edilmiştir.

Anahtar kelimeler: Mudurnu, Etnobotanik, Sürdürülebilir turizm

Ethnobotany in the Development of Sustainable Tourism in Mudurnu


Abstract: Ethnobotany is the study of plants used, conserved and managed by the local people. It deals with local people knowledge of a particular use of plants in gastronomical, medical, religious, and other purposes. Ethnobotany is a rapidly growing scientific topic and it is predominantly linked to economic botany: on the one hand to determine the potential economic value of various plants and, on the other, to make a link with a relatively new aspect of sustainable tourism market. The culture and heritage of destinations are significant pull-factors in attracting domestic and international tourists. Sustainable tourism phenomenon today and in the near future is in the direction of the assumption that the relevant destination contains gastronomic culture. In recent years, different studies have been conducted in terms of collecting as much information as possible about the relationship between traditional use of


edible plant species with gastronomic offer and the increase of sustainable tourism development. The traditional food and local plant resources of a particular interest for sustainable tourism activities in Mudurnu were investigated in this paper. Mudurnu was preferred for this study because of its rich natural, historical and cultural resources which contain as main parts of ecotourism supply. To reach the objective of this study, some interviews were utilized on a key important people and local people living in villages. In this study, special attention was paid to the use of traditional ethnobotany knowledge in local gastronomy and its effect to sustainable tourism development.

Keywords: Mudurnu, Ethnobotanic, Sustainable tourism

***Corresponding author:** Bolu Abant İzzet Baysal Üniversitesi, Mudurnu Meslek Yüksekokulu, Yönetim ve Organizasyon Bölümü, Mudurnu 14800 Bolu, Türkiye

E mail: nailhoscan@ibu.edu.tr (N. HOŞCAN)

Nail HOŞCAN  <https://orcid.org/0000-0003-3669-6532>

Hamdi ZENGİNBAL  <https://orcid.org/0000-0001-5041-2298>

Cite as: Hoşcan N, Zenginbal H. 2020. Ethnobotany in the development of sustainable tourism in Mudurnu. BSJ Eng Sci, 3(4): 151-159.

1. Giriş

İnsanların bitkiler ile dinamik ilişkisini açıklayan Etnobotanik terimi (Luković ve Kostić, 2017), ilk kez 1895 yılında, John W. Harshberger tarafından kullanılmıştır (Kendir ve Güvenç, 2010). Botanik kültürünü inceleyen etnobotanik, bitkilerin insanlar tarafından kullanılmasını, botanik ve antropoloji bilimi bakış açılarından inceleyerek bir neden sonuç ilişkisi ortaya koymaya çalışır (Diker ve ark., 2017).

Bir toplumdaki bireylerin yedikleri ya da içtikleri, yiyecek ve içecek yapımında kullandıkları hammadde ve malzemeler toplumların kültürleri ile yakından ilişkilidir. Etnobotanik, yerel halkın gastronomi, medikal, dinsel ve diğer amaçlarla bitkilerin belirli bir kullanımı hakkındaki deneysel yollarla öğrendiği ve kuşaktan kuşağa aktardığı bilgisini ele almaktadır. Etnobotanik bilgisi, biyoçeşitliliğin korunmasında, ekosistemin işleyişini sürdürmesinde önemli bir öge sayılmaktadır (Kayode ve ark., 2017).

Biyoçeşitliliğin, başta besin olmak üzere temel insan gereksinimlerinin karşılanmasında vazgeçilemez bir yeri ve önemi vardır. Besinler, bitkilerden, bitkileri tüketen hayvanlardan veya hayvanlardan elde edilen ürünlerden oluşmaktadır (Karagöz ve ark., 2010). Dünya genelinde besin olarak tüketilen bitkilerin yaklaşık 20 türden elde edildiği, besin olarak tüketilen yabani bitkilerin sayısının ise 10.000'in üzerinde olduğu bilimsel araştırmalarda ortaya çıkmıştır (Yücel ve ark., 2012).

Anadolu halklarının yabani bitkileri ilaç ve yiyecek olarak kullanmaları, çok eski dönemlere kadar uzanmaktadır. Özellikle ilaç ve besin endüstrisinde, ham madde olarak kullanılan ve doğal olarak yetişen bitkilerin yanı sıra çeşitli türlerin kültürleri de yapılmaktadır (Karaca ve ark., 2015). Türkiye toprakları, biyoçeşitlilik açısından küçük bir kıta özelliği göstermektedir. Anadolu, kendi başına ayrı bir kıta olmamasına karşın fitocoğrafik konumu (Avrupa Sibiryası, İran-Turan ve Akdeniz Flora Alanı) nedeniyle, bir kıtanın taşıyabileceği tüm ekosistem ve habitat özelliklerini tek başına taşımaktadır ve bu sayede pek çok bitkinin gen merkezidir. Türkiye, dünyadaki bitki türlerinin yaklaşık %3,6'sına sahip, bitki türleri açısından nadir ülkelerden biridir. Tüm Avrupa kıtasında 12500 açık ve kapalı tohumlu bitki türü varken,

sadece Anadolu'da bu sayıya yakın (yaklaşık 11707) tür olduğu bilinmektedir ve bunların yaklaşık üçte biri endemik türlerdir. Türkiye, endemiklik oranı (%34) ile dünyada önemli bir yere sahiptir (T.C. Tarım ve Orman Bakanlığı Bakanlığı, 2019).

1.1. Sürdürülebilir Turizm ve Etnobotanik

Son yıllarda ekolojik, ekonomik ve sosyal kaynaklar üzerindeki baskıların yoğunlaştığı ve rekabetçiliğin iyice arttığı bir süreçte sürdürülebilirlik, doğal, ekolojik, sosyokültürel kaynaklarla simbiyotik bir ilişki içerisinde olan turizmde de önem kazanmıştır (Hoşcan, 2018). Turizmde diğer sektörlerden farklı olarak turistik ürün talebinde bulunan tüketici konumunda olan turistler, turistik ürün arzının bulunduğu yerlere yani destinasyonlara gitmekte ya da götürülmektedir. Bu durumda, destinasyonlardaki turizm faaliyetleri sonucunda en başta biyoçeşitlilik kaybıyla, yaşam formlarının bileşimindeki sürekli ve güçlü değişimle karşı karşıya kalınmakla beraber, sosyokültürel çevre ve ekonomik çevre üzerinde de bazı olumsuz etkiler ortaya çıkmaktadır (Demir ve Çevirgen, 2006; Hoşcan, 2017).

Proaktif ve aktif bir karakterde olan sürdürülebilirlik kavramı, "bir toplumun, belirli bir ekosistemin ya da sürekliliği olan herhangi bir sistemin işleyişini, kesintisiz, bozulmadan, aşırı kullanımla tüketilmeden, sistemin yaşamsal bağı olan ana kaynaklara aşırı yüklenmeden devam ettirebilme yetkinliği" olarak tanımlanabilir (Kaypak, 2010; Sarıkaya ve Zişan, 2007). Tarımda, mimarlıkta, işletmecilikte, ekonomide birçok farklı disiplinlerde sıkça kullanılan kavramlardan biri olan sürdürülebilirliğin temel özelliği, insanın geleceğini konu alması ve hangi alanda kullanılıyorsa o alandaki kaynakların korunmasına dayanmasıdır (Sarıkaya ve Zişan, 2007). Özellikle kitle turizmi ve kontrolsüz turizm faaliyetlerinin, doğal, kültürel ve ekonomik kaynaklar üzerindeki olumsuz etkileri ve bu etkilerin turizmin kendi geleceğini tehlikeye attığının görülmeye başlanmasıyla birlikte sürdürülebilirlik anlayışının turizme de yansması sonucu, sürdürülebilir turizm ortaya çıkmıştır (Leung ve ark., 2008; Tosun, 2001; Youell ve Wornell, 2005; Hoşcan, 2017).

Evrensel olarak kesinleşmiş bir tanımı bulunmayan sürdürülebilir turizm, Birleşmiş Milletler Dünya Turizm

Örgütü (BMDTÖ) tarafından, “hem şimdiki hem de gelecekteki sosyal, ekonomik ve çevresel etkilerini bütünüyle göz önüne alan ve aynı zamanda ziyaretçilerin, sektörün, çevrenin ve ev sahibi toplumun gereksinimlerini dikkate alan turizm” olarak tanımlanmaktadır (UNEP ve UNWTO, 2005). Sürdürülebilir turizm, gelecek için farklı bakış açılarının geliştirildiği, destinasyonların ve turistlerin güncel gereksinimlerinin karşılandığı; yaşayan canlı sistemlerin, biyoçeşitliliğin, temel ekolojik süreçlerin ve kültürel bütünlüğün göz önüne alınarak, estetik, sosyal, ekonomik gereksinimlerle birlikte tüm kaynakların yönetiminin bütünlüştürüldüğü bir turizm anlayışıdır (Hoşcan, 2017; Bramwell ve Lane, 1993; Kuntay, 2004; Liu, 2003). Birçok araştırmacı ve yazar tarafından yapılan sürdürülebilir turizm tanımlarında yer alan ortak öğeler; doğal ve insan yapımı kaynakların optimum biçimde uzun süreli kullanımı, insan yerleşimi dokusunun korunması, ekonomik ve sosyal eşitlik, kalite kavramına bağlılık ve sosyal katılımdır (Demir ve Çevirgen, 2006).

BMDTÖ'ye göre sürdürülebilir turizm, turizmin yeni ve farklı bir türü değildir, daha çok turizmin tüm türleri için geçerli olması gereken temel bir koruma yaklaşımıdır (UNEP ve UNWTO, 2005). Turizm türleri arasında destinasyonun doğal yapısına çok aşırı müdahaleye gerek duyulmadan yapılabilecek göl turizmi, kırsal turizm, yayla turizmi, orman turizmi gibi turizm türleri kavramsal olarak bambaşka bir içeriğe işaret etmektedir (Çakılcıoğlu, 2005). Sürdürülebilir turizm kavramı, hangi turizm türü olursa olsun turizm sektörüne yönelik alansal varlıkların ve diğer turizm arz bileşenlerinin optimum koşullarda uzun yıllar boyunca devamlılığının sağlanmasının gerekliliğini ifade etmektedir.

Etnobotanik bir yandan, çeşitli bitkilerin potansiyel ekonomik değerini belirleyen diğer yandan da sürdürülebilir turizm pazarı ile bağı olan hızla büyüyen bir bilimsel alandır. Geleneksel besin, ilaç, insan eliyle yapılmış eşya, kurutulmuş bitkiler veya diğer konseptlerde kullanılan etnobotanik bilgisi, turizm ve özellikle botanik turlarına dayalı ekoturizm faaliyetleri, yerel topluluklar için itici güç olabilmektedir (Luković ve Kostić, 2017).

1.2. Turizm Ekinoksu ve Sürdürülebilirlik

Sürdürülebilirlik kavramının turizmde uygulanmasına ilişkin olarak “Turizm Ekinoksu” kavramı tanımlanmaya başlamıştır. Turizmin destinasyondaki var olan yerel, geleneksel, yerleşik öğeleri bozmadan ve bunların yerine geçmeden yapılandırılması, turizm ekinoksu olarak nitelendirilmektedir. Gündüz ve gecenin eşit olduğu yılın iki gününe verilen ad olan ekinoks, turizm ve tarım gibi eşit odaklı yapılanmaları ifade etmektedir (Arıkan ve Ünsever, 2014). Temel ekonomik dinamiği tarım olan, ekolojik yönden hassas ve biyoçeşitliliği olan birçok destinasyon, turizmden pay alma yönünde ilerlemektedir. Tarım ve turizmin eşitliği ve dengesi üzerine kurulacak sistemde, bir sinerji oluşacaktır. Böylece alansal varlıkların korunması sağlanarak turizm ve tarım destekli sürdürülebilir ekonomik gelişme

gerçekleştirilebilecektir.

Turizm sonucu, ekolojik çevrenin zarar görmesi ve yerel halkın temel ekonomik dinamiğinin değişmesiyle genelde tarım ve hayvancılık faaliyetleriyle ortaya çıkan yöresel gastronomik ürünler gittikçe yok olmaktadır. Destinasyonların kendi ekosisteminin bir parçası olan yenilebilir yabani bitkilerle ilgili sahip olunan kültürel mirasın yani etnobotanik bilgisinin korunması ve gelecek kuşaklara aktarılabilmesi için, biyoçeşitliliğin belgelenmesi oldukça önemli bir konu olarak karşımıza çıkmaktadır.

Somut olmayan kültürel miras arasında yer alan etnobotanik bilgisi, hayvan gözleme ve inceleme, mağaracılık, gibi uzmanlık gerektiren sürdürülebilir bir turizm kaynağıdır ve yan hizmetlerle desteklenmesi gerekmektedir. Temel ekonomik dinamiği tarım olan ve ekolojik yönden hassas birçok destinasyonda, etnobotanik bilgisinin ve uygulamalarının korunması sağlanarak turizm destekli sürdürülebilir ekonomik gelişme gerçekleştirilebilir.

Son zamanlarda, etnobotanikle, gastronomik arz ve sürdürülebilir turizm gelişimi arasındaki ilişki hakkında bilgi toplamak adına farklı çalışmalar yapılmıştır ancak görece yeterli değildir. Etnobotanik, Mudurnu gibi kültürel miras alanı yönetimi yapılanması olan yerlerde özellikle ele alınması gereken bir konudur. Bu çalışmada, Mudurnu genelinde, etnobotanik kültürünün bir parçası olan doğal olarak yetişen, yemek ve medikal amaçlarla kullanılan bitkilerin belirlenmesi amaçlanmıştır. Böylece alanda yetişen doğal bitkilerin, tanınması, korunması, değerlendirilmesi, yerel halk dışında da kullanımının yaygınlaşması ve sürdürülebilir turizm kaynağı olarak değerlendirilebilmesi sağlanabilecektir.

2. Materyal ve Yöntem

2.1. Çalışma Alanı

Bu çalışmada Mudurnu'da sürdürülebilir turizm gelişiminde çekim ögesi olarak Etnobotanಿಗೆ odaklanılmıştır.

2.1.1. Coğrafi konum

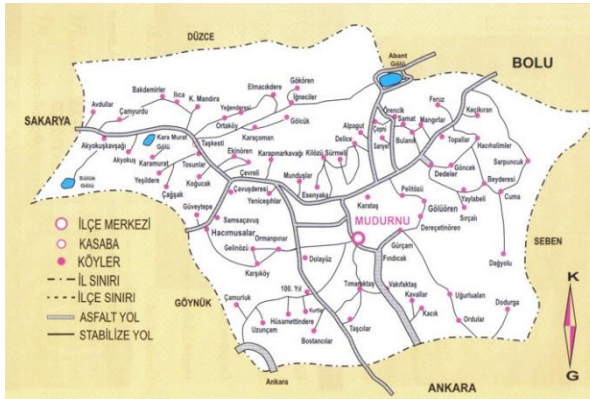
Mudurnu, Türkiye'nin kuzeyindeki Karadeniz Bölgesi'nin Batı Karadeniz Bölümü'nde bulunan Bolu ilinin bir ilçesi olmasına karşın batı ucunda küçük bir alan Marmara bölgesi sınırları içine girmektedir (Şekil 1).



Şekil 1. Mudurnu haritası ve Türkiye'deki konumu.

Mudurnu, 40° 19' - 40° 39' kuzey enlemleri ile 30° 50' - 31° 30' doğu boylamları arasında yer alır. Şekil 1'de görüldüğü gibi Mudurnu, Bolu'nun ilçelerinden doğuda Merkez ve Göynük; batıda Seben ilçeleriyle birlikte; güneyde Ankara'nın Nallıhan; kuzeyde Düzce'nin Gölyaka ve Kaynaşlı; kuzeybatıda da Sakarya'nın Hendek ve Akyazı ilçeleriyle çevrelenmiştir. Mudurnu, Bolu'nun coğrafik yapısı içerisinde dördüncü geniş alanı olan ilçesidir.

Köyleri genellikle dağınık bir görünüm arz eden, 11 orman içi köyü 62 orman kenarı köyü olan Mudurnu (Şekil 2), Bolu'dan yaklaşık olarak 50 km güneybatıda Kulaklı ve Hisar tepelerinin arasında bulunan vadi içerisinde kurulmuş olup, toplam yüzölçümü 1349 km²'dir.



Şekil 2. Mudurnu köyleri haritası.

Abant Gölü, Bolu İl Özel İdaresi mülkiyetinde olmasına karşın Mudurnu ilçe mülki sınırları içindedir. Abant Gölü ve çevresinde 1262,23 ha'lık alan, 21.10.1988 tarihinde tabiat parkı olarak ilan edilmiştir. Konumu, coğrafik özelliklerinden dolayı Abant Gölü Tabiat Parkı'nda biyoçeşitlilik oldukça zengindir ve bu özellikleriyle uluslararası birçok doğa araştırmacısının ilgi odağı olmuştur.

2.1.2. Abiyotik özellikler

İklim

Geniş bir ormanlık arazisi, mera ve yaylaları, akarsu ve gölleri, tarım alanları, dağ ve yüksek tepeleri olan Mudurnu'da coğrafik konumdan dolayı Karadeniz ardı iklimi olarak tanımlanan; deniz iklimi ile İç Anadolu'nun tam karasal iklimi arasında bir geçiş iklimi görülür (Döker, 2013; Tunçkol, 2017). Yıllık ortalama hava sıcaklığı 10 °C olan Mudurnu'da, yazlar kurak ve sıcak, kışlar da soğuk ve sert geçer. Yıllık yağış miktarı ise ortalama 559 mm gibi düşük bir düzeydedir. Kuzeybatıdan ve kuzeyden esen rüzgârlar Mudurnu'ya yağış getirmektedir. Maksimum yağışlar kış ve ilkbahardadır. Ortalama yaz günü sayısı 88 olup, çevresindeki merkezlere göre daha kısadır (T.C. Tarım ve Orman Bakanlığı Meteoroloji Genel Müdürlüğü, 2019).

Jeolojik yapı ve yüzey şekilleri

Bir baştan diğer başa Kuzey Anadolu Dağları ya da Karadeniz Dağları adı verilen sıra dağlar tarafından işgal edilen Batı Karadeniz'in en belirgin özelliği dağlık

olmasıdır. Kıyı çizgisine genel olarak paralel sıralar halinde uzanan bu dağlar, iç kısımlarda aralarına sokulmuş oluk biçimindeki depresyonlarla (çukur alanlarla) birbirlerinden ayrılmaktadırlar (Özçağlar, 2015). Mudurnu'nun üzerinde bulunduğu depresyonlar;

- Mudurnu Suyu ile Soğanlı Çayı kavisi arasında uzanan dağlar: Elmacık Dağı, Seben Dağı, Gerece Dağları, Kocadağ,
- Mudurnu Suyu Vadisi - Kapıorman Dağları - Göynük Çayı Vadisi,
- Köroğlu Dağları'dır.

Kuzey ve kuzey batı Mudurnu'da mezozoik tortul kayalar görülmektedir. Mudurnu'nun jeolojik yapısını, temel olarak fliş formasyonu oluşturmaktadır. Kıvrıntılı kırık ve çöküntülü özelliği olan fliş, aralarında yer yer kireç taşı tabakaları da bulunan killi, çimentolu çoğunlukla pek kalın olmayan tabakalar halinde çeşitli kumtaşı ve çakıltaşı (konglomera) tabaka değişimlerinden oluşur. Bu tabakalar genellikle killi toprak verdiklerinden heyelanlarla çok karşılaşılır (Bolu Tarım Master Planı, 2003).

Depremsellik, Mudurnu için son derece önemli bir konudur çünkü Mudurnu'da, toprak henüz yerleşmemiş olduğundan tarih boyunca birçok deprem olmuştur. Türkiye Deprem Bölgeleri Haritasında birinci derece deprem kuşağındadır (Bolu Tarım Master Planı, 2003).

Mudurnu'da dağ sıraları doğu-batı doğrultusunda alçalarak uzanır. Belirgin olan üç dağ sırası;

1. En kuzeyden geçen Düzce İli sınırı boyunca uzanan Abant Dağları, doğuda daha yükselerek Erenler Tepesi (1794 m) adını alır.
2. Akyazı-Göynük sınırından başlayarak doğuya sırasıyla Akkaya Tepesi (1628 m), Kuzgunkaya Tepesi (1651 m) ve Alaçam Tepesi (1689 m) ile Bolu'nun Aladağ ve Köroğlu sıra dağları ile birleşir.
3. Göynük ilçesini ortalayıp yükselerek güneybatıdan birleşip Mudurnu sınırlarına girer ve bir miktar doğuya uzandıktan sonra kuzeye yönelip ikinci dağ sırası ile birleşir. Mudurnu kent merkezinin doğusunda bulunan Şehriman Tepesi (1115 m) ve Hisar Dağı (1384 m) dikkat çeken önemli iki yükseltidir (Bolu Tarım Master Planı, 2003).

Engibeli bir arazi yapısı olduğundan Mudurnu ve çevresinde ovalar azdır. Abant Dağları'nın güneybatısında Avu, Hoca ve Ardıç Dağları ile çevrili küçük bir alanda ortalama yükseltisi 800 m olan Mudurnu ovası yer alır (Bolu Tarım Master Planı, 2003). Mudurnu kent merkezinin 12 km kadar kuzeyinde ve Abant Gölü'nün güneyinde yoğunlaşan Mudurnu yaylaları, doğuya doğru alçalarak ve bir yarım daire biçiminde Mudurnu'nun kuzey sınırını çizmektedir.

Güneydeki yaylaların ortalama yükseklikleri 1.030 m, kuzeydekilerin 1060 m'dir. Grup içindeki yaylaların en yüksek olanları Dedeler (1600 m), Alpogut (1550 m), Dodurga (1550 m) ve Dağyolu (1550 m) yaylarıdır (Bolu Tarım Master Planı, 2003).

Toprak özellikleri ve arazi kullanımı

Mudurnu'nun toprağı, organik ve inorganik maddeler

bakımından çok zengin durumda değildir. Bolu ilinde yalnızca Mudurnu'da Alüvyal sahil bataklıkları bulunmaktadır ve toplama alanı 37 ha'dır (Keçeli, 2004). Mudurnu'nun deniz seviyesinden yüksekliği 800-850 m olup, arazilerin taban suyu düşük, yağmur sularını bünyesinde tutmayan daha çok II. ve III. sınıf arazi topluluğuna girmektedir. Bu toprak yapısına karşın oldukça geniş ürün desenini:

- Endüstri bitkileri üretiminde; patates, buğday, mısır, şekerpancarı, yulaf, arpa,
- Sebze üretiminde; fasulye ve domates ilk sırada yer almakta olup ıspanak, pırasa, soğan, biber, salatalık, balkabağı, marul, lahana,
- Meyve üretiminde; elma ve armut ağırlıklı olmak üzere ceviz, erik, kiraz, ayva, dut ve üzüm oluşturmaktadır.

Toplam alanı 134900 hektar (ha) olan Mudurnu'nun tarım alanı 21338 ha, orman ve fundalık alanı 87558 ha, çayır ve mera alanı 20717 ha ve tarım dışı alanı 5287 ha'dır (Tablo 1).

Tablo1. Mudurnu'da arazilerin kullanım biçimlerine göre dağılımı

Kullanım biçimi	Alan	%
Tarım alanı	21338 ha	17,69
Orman ve Fundalık	87558 ha	55,75
Çayır ve Mera	20717 ha	14,71
Tarım dışı	5287 ha	11,84
Toplam alan	134900 ha	100,00

Erozyon, Abant Dağları'nın daha çok güney yamaçlarında ve Mudurnu-Seben arasında genişçe alanlarda görülmektedir.

Su varlığı

Tek başına nehir debisine ulaşmayan Abant Dağları'nın güney yamaçlarından doğan Mudurnu Çayı ve Seymen Deresi; Abant Dağları üzerinde oluşmuş bir birikinti ve krater gölü olan Abant (125 ha), Taşkesti Beldesinde yer alan krater gölü olan Karamurat (5 ha) ve heyelan sonucu oluşmuş Sülüklü Göl (6 ha), Mudurnu'nun başlıca yüzey suyu kaynaklarıdır.

2.1.3. Biyotik özellikler

Flora

Barındırdığı bitki türlerinin endemikliği yönünden de zengin olan Türkiye'nin Anadolu kesimindeki endemik bitki türleri daha yaygındır. Endemik bitki türleri özellikle dağlık kısımlarda yoğunlaşır (Karagöz ve ark., 2010). Mudurnu, en fazla endemik bitki türü barındıran yörelerden olan Kuzey Batı Geçit Bölgesinde yer alan Eskişehir, Kütahya, Kastamonu, Bolu ve Bilecik illeri arasında konumlanır.

Mudurnu'da görülen Karadeniz ardı iklimi, depresyonlarda oluşan lokal iklim özellikleri, değişken yükselti ve bakı faktörü, zengin bir floristik çeşitliliğin ortaya çıkmasında etkili olmuştur. Mudurnu'da, Türkiye'de etkisi olan farklı flora alanlarının neredeyse tümünün özellikleriyle karşılaşmaktadır. Mudurnu'nun

kuzeyinde Avrupa-Sibirya Flora Alanı etkileri, güneyinde ise İran-Turan Flora Alanı etkileri görülmektedir. Bunun yanı sıra akarsu vadilerinin batıya ve güneye açılan kesimlerinde kısmen Akdeniz Flora Alanı etkileri görülmektedir. Özellikle dağların kuzeye bakan yamaçları daha çok yağış ve daha az radyasyon aldığı için genellikle nemli ve çok nemli ortamları oluşturur (Tunçkol, 2017; Karagöz ve ark., 2010; Döker, 2013).

Mudurnu yüzölçümünün yaklaşık %65'lik kısmı orman ve fundalıktır. Yer yer ekili tarım alanları arasında bile orman öbekleriyle karşılaşılır. Mudurnu'nun genel olarak doğal bitki örtüsünü oluşturan ormanlarda geniş yapraklı ağaçlarla birlikte iğne yapraklı ağaçlar da bulunur. İğne yapraklı ağaçlar dağlara doğru yükseldikçe geniş yapraklı ağaçların yerini alır. En çok görülen ağaç türleri; Çınar, Gürgen, İhlamur, Karaçam, Kayın, Kestane, Kızılağaç, Kızılcım, Köknar, Meşe, Orman Kavağı ve Sarıçamdır. Meralarda ise her türlü meyvesiz ve meyveli ağaçlar bulunur. Abant Gölü Tabiat Parkı florasına dâhil 84 familyaya ait 332 cins, 66 tür, 147 alt tür ve 69 varyete, toplam 672 takson ve 150 liken türü kaydı vardır. Bu taksonlardan 51'i endemiktir (Külköylüoğlu ve İkizce, 2006).

Fauna

Mudurnu'nun iklimi, topografyası, florası ve su kaynakları, yaban hayatı için elverişli koşullar sunmaktadır. Mudurnu ve çevresinde yer alan göl ve meralarda domuz, kurt, tilki, tavşan, geyik, karaca, çulluk, bıldırcın, keklik, yaban ördeği, üveyik gibi yabani hayvanlar bulunmaktadır.

Abant Gölü Tabiat Parkı'nda hayvan çeşitliliği, en az bitkiler kadar vardır. Abant Gölü'nde bulunan yedi balık türünden biri olan Abant Alası (*Salmo trutta abanticus*) endemik bir alabalık türüdür. Bununla beraber dünyada şimdiye kadar herhangi bir yerde kaydı bulunmayan Küçük Taraklı Semender (*Triturus vulgaris kosswigi*) ve Abant Fındık Faresi (*Muscardinus avellanarius abanticus*) endemik türlerdendir (Külköylüoğlu ve İkizce, 2006).

2.2. Yöntem

Çalışmada ikincil kaynakların yanı sıra, kalitatif araştırma yöntemlerinden de yararlanıldı. Bu bağlamda öncelikle Bolu İl Kültür ve Turizm Müdürlüğü Halkbilim Araştırmalarından bitkilerin kullanımı hakkında detaylı bilgiler derlendi, bununla birlikte Mudurnu köylerinde yaşayanlardan 13 kişi ile yarı yapılandırılmış görüşme yapıldı. Mudurnu'da yetişen yabani bitkilerin yöresel adları, bu bitkilerin nasıl kullanıldığı (yemek yapımında kullanılan bitkiler, tedavi amacıyla kullanılan bitkiler) ile ilgili bilgiler toplanmıştır.

3. Bulgular ve Tartışma

Her mutfak kültüründe olduğu gibi Karadeniz Mutfağını da biçimlendiren en önemli etken coğrafik özellikleridir. Dağların kıyıya paralel ve yakın olması Karadeniz'in iç kesimi ile dış kesiminin yemek kültürünü ikiye bölmüştür. Karadeniz'de tarım alanlarının az olması, hayvancılığın yaylalarla sınırlı kalması ve çok yağış alması yemek kültürünü doğrudan etkilemiştir. Çok yağış

alması, çeşitli yabancı otların yaygınlaşmasına sebze üretiminin ve bahçe sebzeçiliğinin de gelişmesine neden olmuştur (Aker ve Terzihan, 2013). Coğrafik olarak Batı, Orta ve Doğu Karadeniz olarak bölümlere ayrılmış olan Karadeniz gibi büyük bir bölgede, mutfak kültürü, yerel olarak biçimlenmiş benzer yemekleri olan üç ayrı grup oluşturmuştur (Halıcı, 2001);

- Bartın, Bolu, Düzce, Karabük, Kastamonu, Sinop, Zonguldak batı grubunu;
- Artvin, Giresun, Ordu, Samsun, Rize, Trabzon doğu kıyı grubunu;
- Amasya, Bayburt, Çorum, Gümüşhane, Tokat doğu iç grubunu oluşturmaktadır. Artvin kenti bir kıyı kenti olmamasına karşın yemeklerinin benzerliği nedeni ile kıyı grubu içerisinde yer almaktadır.

Batı Karadeniz, Karadeniz bölgesinin diğer iki bölümüne göre topografya, sosyo-ekonomik refah ve etnik köken bakımından farklı özellikler göstermektedir. Batı Karadeniz mutfağı, geleneksel olarak hamur işi ve sebze yemeklerinden oluşmaktadır, özel günlere ve düğünlere özgü yemekleri de vardır. Yemeklerde, yerel tarımsal ürünlerin karakteristiğinden kaynaklanan değişik bir lezzet vardır. Dikkati çeken başka bir özellik ise aynı türe ait yiyeceklerin pek çok hazırlanış biçimi olmasıdır. Deniz kıyısında olan yerleşimlerde deniz ürünleri kültürü vardır. Batı Karadeniz kıyı yerleşimlerinde Doğu

Karadeniz yemek kültürü öğelerine rastlanmasını, göç olgusu üzerinden ayrıca değerlendirmek gerekecektir.

Ot sözcüğü, tarımsal faaliyet sonucu özellikle yetiştirilmeyen, doğada kendiliğinden yetişen bitkiler anlamında kullanılmaktadır. Bunlar, yabancı ya da yabancı otlar olarak nitelendirilmektedir (Baysal, 2003). Karadeniz Bölgesi'nde Ege Bölgesi'nden daha zengin bir ot kültürü vardır. Bu ot kültürünün temelinde bölgenin çok yağış almadan kaynaklanan floristik çeşitlilik bulunmaktadır (Aker ve Terzihan, 2013; Halıcı, 2001). Bu araştırmada da Karadeniz Bölgesinde yer alan Mudurnu'da ot kültürünün zenginliğini gösteren sonuçlara ulaşılmıştır (Tablo 2).

Yenilebilir otlardan çorbalar, sebze yemekleri, sarma ve salatalar yapıldığı gibi; özel aromaya sahip olanlar taze ya da kurutularak tat verici olarak mutfaklarda tüketilmekte, medikal amaçlarla değişik formlarda kullanılmakta veya çay olarak içilmektedir. Yemek olarak kullanılan otlar, bir anlamda sebzelerin kendiliğinden yetişenleridir. Bazı otların ise daha önce kültür bitkisi oldukları, zamanla yabancı ota dönüştükleri düşünülmektedir (Baysal, 2003). Türkiye'nin her bölgesinde olduğu gibi Mudurnu ekosisteminde de yabancı olarak yetişen birçok ot, gerek yemek gerekse medikal amaçlarla kullanılmaktadır (Tablo 2 ve Tablo 3).

Tablo 2. Mudurnu'da yemek yapımında kullanılan yabancı bitkiler*

• Acı Kavun	• Dağ Nanesi	• Kızılacak
• Acı Kevirdek	• Damar Otu	• Kuş Ekmeği
• Ahlat	• Demirdikeni	• Kuşburnu
• Alıç	• Ebegümeci	• Kuzukulağı
• Altınotu	• Efelek	• Labada
• Aşotu	• Gelincik	• Madımak
• Aynısefa	• Hardal	• Menengiç
• Ayrık	• Hayit- Ishal Otu	• Papatya
• Ballibaba	• Isırgan	• Salep
• Balotu	• Karabaş	• Salkım Söğüt
• Beka Otu	• Karakavuk	• Sarı Kantaron
• Beyaz Civan Perçemi	• Kavgan Dikeni	• Semiz Otu
• Bıtrak Otu	• Kaz Ayağı (Kaz Mancarı)	• Sirken
• Çetlemik	• Kekik	• Toklubaşı
• Çiğdem	• Kırkdamar	• Yemlik
• Çitlenbik	• Kırkkilit Otu	• Yılan Yastığı
• Çoban Çökerten	• Kısamahmut Otu	

*Yöresel adlandırmalar korunmuştur. Yabancı formlarda kullanımı olanlar listeye alınmıştır.

Mantarlar, Mudurnu'nun yemek kültüründe önemli bir parçasıdır çünkü Mudurnu'daki ormanlar mantar açısından da oldukça zengindir. Bolu Abant Tabiat Parkı'nın Makrofungal Çeşitliliği adlı çalışmaya göre (Servi ve ark., 2010); Abant Tabiat parkında 34 familya ve 67 cinse ait 103 makrofungi taksonu belirlenmiştir. Özellikle Abant Gölü Tabiat Parkı çevresinde Kanlıca, Karakız, Meşe Mantarı, Söğüt Mantarı, Cincile, Yeryaran, Dedebört, Telli Mantar, Cücekız, Gelin gibi bilinen mantar türleri yetişmektedir (Külköylüoğlu ve İkizce, 2006; Gökçeel, 2017).

Mudurnu'da otlarla yapılan yemeklerde yaygın olarak kullanılan hazırlama ve pişirme yöntemleri; haşlama,

yumurta ile kavurma, kavurularak sarımsaklı veya sarımsaksız yoğurt ile birlikte (yoğurtlama olarak) tüketme, dolma veya sarma yapma, börek iç harcı olarak kullanmadır. Mantarlar ise Mudurnu mutfağında genelde çorba, börek ve tereyağında kızartma biçiminde kullanılmaktadır. İçeceklerde yaygın olarak kullanılan hazırlama yöntemi demleyip çay olarak tüketmedir (Tablo 3).

Tablo 3'teki uygulamalar, bitkilerin medikal amaçlarla kullanıldığını göstermektedir; diğer yandan yenilebilir otlar aromatik tatlarından dolayı yemek, çay ve baharat olarak da tüketilmektedir. Medikal amaçla kullanılan bitkilerin, insan sağlığını ciddi bir biçimde tehdit

etmeyeceği algısı yaygın olsa da olası zararlı etkileri göz ardı edilmemelidir, dolayısıyla bu bitkilerin medikal amaçlarla kullanımı bilimsel denetimle gerçekleştirilmelidir.

Tablo 3. Mudurnu'da çay ve baharat olarak da tüketilen daha çok medikal amaçlarla kullanılan yabani bitkiler

Yerel Adı	Kullanılan Kısmı	Hastalık	Kullanım Biçimi
Ardıç	Reçine - Katran	Kırık, çıkık ve bazı ortopedik rahatsızlıklar	Üzerine sürülür
Ardıç Purçu (Ökse Otu)	Çiçek - Yaprak	Öksürük	Kaynatılıp balla tatlandırılarak içilir
Arpa	Tohum	Soğuk algınlığı	Lapa haline getirmek için tohumlar kaynatılır ve lapa hala sıcakken hastanın vücudu bununla kaplanır ve çarşafa sarılarak terlemeye yatırılır.
Asma Yapağı	Yaprak	Baş ağrısı	Başa sarılır
Ayva	Yaprak	Grip, soğuk algınlığı	Kaynatılıp içilir
Buğday	Tohum	Korku şok durumu	Niştastası sulandırılıp içilir
Ceviz	Yaprak	Baş ağrısı	Başa sarılır
Ceviz	Yaprak	Yüksek ateş	Başa sarılır
Ceviz	Yaprak	Soğuk algınlığı	Vücuda kaplandıktan sonra üzerine çarşaf sarılarak terlemeye yatırılır.
Çam	Reçine - Sakız	Ülser	Açkarına yenilir
Çam	Soymuk - Floem	Nefes darlığı, kanser, ülser	Sütle kaynatılarak sabahları içilir
Ferik Elması	Meyve	Kulak ağrısı	Meyvenin suyu kulağa damlatılır
Ihlamur	Çiçek -Yaprak	Grip, soğuk algınlığı	Kaynatılıp içilir
Isırgan Otu	Yaprak	Baş ağrısı	Başa konulur
Kalabak	Yaprak	Güneş geçmesi	Vücuda sarılır
Karanfil	Tomurcuk	Diş ağrısı	Ağrıyan dişin üzerine konur
Keçieriği	Meyve	Baş ağrısı	Pestili Başa sarılır
Kekik	Yaprak	Ülser ve mide rahatsızlıkları	Ekstrakt suyu içilir
Kelem (Lahana)	Yaprak	Baş ağrısı	Başa sarılır
Kipkesen - Tetere	Yaprak	Mide rahatsızlıkları	Kaynatılarak içilir
Kiren -Kızılıcık	Meyve	İshal ve sindirim hastalıkları	Su ısıtılır ve şeker ile birlikte kızılıcık marmeladına ilave edilir, karıştırılarak şerbet kıvamına getirildikten sonra soğuk olarak içilir
Kökнар	Reçine - Sakız	Çıban, apse	Derinin üzerine uygulanır
Kuşburnu	Tomurcuk	Mayasıl	Kaynatılıp bir gece bekletilerek içilir
Maydanoz	Kök	Mide rahatsızlıkları	Kaynatılarak içilir
Nane	Yaprak	İştahsızlık ve mide rahatsızlıkları	Kaynatılarak içilir
Pamukçu Çalısı	Yaprak + Çiçek	Ülser	Kaynatılarak içilir
Patates	Yumru	Baş ağrısı	Dilimlenerek başa sarılır
Soğan	Yumru	Çıban, apse	Pişmiş olarak sarılır
Zeytin	Meyve	Doku travması	Ezilerek ağırlı bölgeye sarılır
Zeytin	Yağ	Doku travması	Ağırlı bölge ovulur
Zeytin	Yağ	Ülser ve karın ağrısı	İçilir

4. Sonuç ve Öneriler

Ait oldukları kültürün önemli bir mirası olan yöresel yemekler, geleneksel besinler, bitkisel ilaçlar, insan eliyle yapılmış bitkisel ürünler, kurutulmuş bitkiler veya diğer konseptlerde kullanılan etnobotanik bilgisi turizm ve özellikle botanik turlarına dayalı sürdürülebilir turizm faaliyetleri, destinasyonlardaki yerel topluluklar için itici

güç olabilmektedir.

Zengin bir gastronomi kültürü Mudurnu'da, doğal olarak yetişen ve yerel halk tarafından genelde çay yapılarak tüketilen ve medikal amaçlarla kullanılan bitkiler olmasına karşın, bu bitkiler yemek yapımında da kullanılmaktadır. Besin olarak tüketilen bu bitkiler özellikle Nisan ayında ortaya çıktığı için yoğun olarak

bahar ve yaz aylarında tüketilmektedirler. Yenilebilir mantarlar, ilkbahar ve daha çok sonbaharda yağmurlardan sonra çimenlik ve çam ormanlarının atlarından toplanılmaktadır. Medikal amaçlı olarak kullanılan bitkiler de özelliğine göre bahar ve yaz aylarında toplanılarak değerlendirilmektedirler.

Etnobotanik öğelerinin korunması ve sürdürülebilir turizmde kaynak olarak değerlendirilebilmesi için Mudurnu ve yakın çevresindeki potansiyeli geliştirmek gerekmektedir. Mudurnu ve yakın çevresindeki etnobotanik potansiyeli geliştirmek için;

- Biyoçeşitliliğin korunması ve sürdürülebilir kullanımını sağlayacak geniş kapsamlı envanter ve izleme çalışması yapılmalı,
- Kapsamlı envanter ve izleme çalışması sonucunda ticarileşebilecek etnobotanik ürünleri belirlenmeli, bazı bitkiler kültüre alınmalı ve bölgeye özgü önemli bitkiler için coğrafi işaret alınmalı,
- Koruma alanları belirlenmeli ve etnobotanik parkı kurulmalı,
- Mudurnu'ya özgü etnobotanik bilgisinin belirli bir eğitim programı içinde aktarıldığı kısa dönemli, herkese açık atölye çalışmaları tasarlanmalı,
- Gastronomik değeri olan etnobotanik ürünlerine, Mudurnu'daki yiyecek içecek işletmelerinin menülerinde yer verilmelidir.

Böylece, yörenin gastronomik mirasının ve yenilenebilir doğal kaynakların korunması ile ilgili halkın bilinci ve farkındalığı artırılabilecek; Mudurnu'nun yerel kültürünün ve kimliğinin tanınması sağlanacak ve ekogastronomik faaliyetlerin sürdürülebilir olmasına olanak sağlanacaktır.

Çıkar İlişkisi

Yazarlar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Teşekkür ve Bilgilendirme

Bu çalışma, özeti 21-22 Eylül 2018 tarihlerinde Bolu'da düzenlenen Uluslararası Mudurnu Araştırmaları, Ahilik ve Halk Kültürü Sempozyumu'nda "Mudurnu'da Sürdürülebilir Turizm Gelişiminde Etnobotanik" adıyla çağrılı bildiri olarak sunulmuş ve bildiri kitapçığında yer almış metnin geliştirilmiş halidir.

Kaynaklar

- Aker G, Terzihan B. 2013. Karadeniz mutfak kültürü. Sabancı Üniv, Murat Belge ile röportaj. Sabancı Üniversitesi, İstanbul.
- Ankan İ, Ünsever İ. 2014. Turizm ekinoksu-tarım ekonomisi odaklı yörelerde turizmin gelişim modeli. *Eko-Gastronomi Derg*, 1(1): 17-25.
- Baysal A. 2003. Beslenme kültürümüz (3 b.). Kültür ve Turizm Bakanlığı Yayınları, Ankara.
- Bolu Tarım Master Planı. 2003. Bolu Tarım Master Planı. Bolu: Bolu Valiliği Tarım İl Müdürlüğü - İl Tarım ve Kırsal Kalkınma Master Planlarının Hazırlanmasına Destek Projesi a, Bolu.
- Bramwell B, Lane B. 1993. Sustainable tourism: An evolving global approach. *J Sustainable Tourism*, 1-5.
- Çakılcıoğlu M. 2005. Turizm odaklı sürdürülebilir kalkınma için bir yöntem önerisi. *Tasarım Kuram Derg*, 9(16): 27-42.

- Demir C, Çevirgen A. 2006. Turizm ve çevre yönetimi: sürdürülebilir gelişme yaklaşımı. Nobel Yayın Dağıtım, Ankara.
- Diker O, Türker N, Kaya FB. 2017. Eskişehir'in etnobotanik çeşitliliğinin eko-gastronomi faaliyetleri çerçevesinde incelenmesi. *Karabük Üniv Sosl Bil Derg*, (3): 51-58.
- Döker MF. 2013. Karadeniz bölgesi. *Türkiyenin bölgeleri dersi 2012-2013 Notu*. Sakarya Üniversitesi Fen Edebiyat Fakültesi Coğrafya Bölümü, Sakarya.
- Gökçeel E. 2017. Mudurnu. N. Doğan (Dü.) içinde, Bolu ve mutfak kültürü (s. 495-559). Abant İzzet Baysal Üniv Bolu Halk Kültürünü Araştırma ve Uygulama Merkezi, Bolu.
- Halcı N. 2001. Karadeniz bölgesi yemekleri. Güray Ofset Matbaacılık, Ankara.
- Hoşcan N. 2017. Seyahat acentelerinde sürdürülebilir pazarlama yaklaşımının kullanımı: Batı Karadeniz destinasyonu örneği. Ankara Üniversitesi Sosyal Bilimler Enstitüsü Sosyal Çevre Bilimleri Anabilim Dalı Yayımlanmamış Doktora Tezi, Ankara.
- Hoşcan N. 2018. Küreselleşme sürecinde yöresel gastronomik ürünlerin üretiminde ve pazarlanmasında kooperatifçilik üzerine bir model önerisi. *Güncel Turizm Araş Derg*, 2(1):390-413.
- Karaca OB, Yıldırım O, Çakıcı C. 2015. Gastronomi turizmde otlar, ot yemekleri ve sağlıkla ilişkisi üzerine bir araştırma. *J Tourism Gastron Stud*, 3(3): 27-42.
- Karagöz A, Zencirci N, Tan A, Takın T, Köksel H, Sürek M, Özbek K. 2010. Bitki genetik kaynaklarının korunması ve kullanımı. Türkiye Ziraat Mühendisliği VII. Teknik Kongresi Bildiriler Kitapçığı, 155-177, Ankara.
- Kayode J, Akinyele O, Ayeni MJ. 2017. Ethnobotany of ecological-based tourist centres in the western zone of Ekiti State, Nigeria. *Ecol Evol Biol*, 2(5): 78-86.
- Kaypak Ş. 2010. Ekolojik turizmin sürdürülebilirliği. *Uluslararası Alanya İşletme Fak Derg*, (2): 93-114.
- Keçeli T. 2004. Batı Karadeniz Bölgesi (Bolu-Zonguldak-Bartın-Kastamonu) ciğerotları (Hepaticae) florası. Ankara Üniversitesi Fen Bilimleri Enstitüsü Doktora Tezi, Ankara.
- Kendir G, Güvenç A. 2010. Etnobotanik ve Türkiye'de yapılmış etnobotanik çalışmalara genel bir bakış. *Hacettepe Üniv Ecz Fak Derg*, 30(1): 49-80.
- Kuntay O. 2004. Sürdürülebilir turizm planlaması. Alp Yayınevi, Ankara.
- Külköylüoğlu O, İkizce N. 2006. Mudurnu ilçesinin biyolojik zenginlikleri. *Mudurnulular Derneği Bülteni*, Kasım 2006: 30-32.
- Leung YF, Marion JL, Farrell TA. 2008. The role of recreation ecology in sustainable tourism and ecotourism. S. F. McCool, & R. N. Moisey içinde, *Tourism, recreation and sustainability: Linking culture and the environment*. CABI, Oxfordshire.
- Liu Z. 2003. Sustainable tourism development: a critique. *J Sustainable Tourism*, 11(6): 459-475.
- Luković M, Kostić M. 2017. Traditional ethnobotanical knowledge in 21th century as an important element of sustainable tourism. The Second International Scientific Conference Tourism In Function Of Development Of The Republic Of Serbia Tourism product as a factor of competitiveness of the Serbian economy and experiences of other countries Thematic Proceedings, 273-289, Banja, University Of Kragujevac Faculty Of Hotel Management And Tourism In Vrnjačka Banja.
- Özçağlar A. 2015. Batı Karadeniz Bölgesi. *Türkiyenin coğrafi bölgeleri II Ders Notları*. Ankara Üniversitesi, DTCF, Ankara.
- Sarıkaya M, Zişan KF. 2007. Sürdürülebilir kalkınmada işletmenin rolü: kurumsal vatandaşlık. *Yönetim ve Ekon Derg*, 14(2): 221-233.
- Servi H, Akata I, Çetin B. 2010. Macrofungal diversity of Bolu Abant Nature Park (Turkey). *African J Biotech*, 9(24): 3622-

- 3628.
- T.C. Tarım ve Orman Bakanlığı Bakanlığı. 2019. Türkiye'nin biyolojik çeşitliliği. T.C. Tarım ve Orman Bakanlığı Bakanlığı Doğa Koruma ve Milli Parklar Genel Müdürlüğü, Ankara.
- Tosun C. 2001. Challenges of sustainable tourism development in the developing world: The case of Turkey. *Tourism Manag*, 22(3): 289-303.
- Tunçkol B. 2017. Bolu ili florasına katkılar. *Mudurnulular Dern Bült*, 68-73.
- UNEP, UNWTO. 2005. making tourism more sustainable - A guide for policy makers. UNEP, Paris.
- Youell R, Wornell R. 2005. Quality as a key driver in sustainable rural tourism business. D. Hall, I. Kirkpatrick, & M. Mitchell içinde, *Rural tourism and sustainable business*. Channel View Publications, Toronto.
- Yücel E, Şengün İY, Çoban Z. 2012. Afyonkarahisar çevresinde gıda olarak tüketilen yabani otlar ve tüketim biçimleri. *Biyol Çeş Kor*, 5(2): 95-105.



Derleme (Review)

Cilt 3 - Sayı 4: 160-164 / Ekim 2020
(Volume 3 - Issue 4: 160-164 / October 2020)

TÜRKİYE’NİN NESLİ TEHLİKE ALTINA OLAN GÖL SOĞANI (*LEUCOJUM AESTIVUM*) ÜZERİNE ARAŞTIRMA VE KORUMA ÇALIŞMALARI

Fisun Gürsel ÇELİKEL¹, Sevim DEMİR^{1*}

¹Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Bahçe Bitkileri Bölümü, 55200, Atakum, Samsun, Türkiye

Gönderi: 19 Şubat 2020; **Kabul:** 24 Mart 2020; **Yayınlanma:** 01 Ekim 2020

(Received: February 19, 2020; **Accepted:** March 24, 2020; **Published:** October 01, 2020)

Özet

Türkiye bitki genetik çeşitliliği ve özellikle de geofitler açısından özel bir konuma sahiptir. Ancak göl soğanı olarak bilinen *Leucojum aestivum* L. da dâhil olmak üzere birçok geofit türü genetik erozyona uğrayarak yok olma tehlikesi ile karşı karşıya kalmaktadır. Ülkemizde doğal olarak yetiştiği bilinen *Leucojum* cinsine ait tek tür olan *L. aestivum* Bolu, İstanbul, Bursa, Erzurum, Kocaeli, Konya ve Samsun illerinde yayılış göstermektedir. Nesli tehlike altında olan bu tür IUCN (International Union for the Conservation of Natural Resources) kırmızı liste kategorilerine göre Düşük Riskli (LC) sınıfında yer almaktadır. Aşırı otlatma ve bilinçsiz arazi yönetimi türün ana tehditleri arasında yer almaktadır. Beyaz çiçekleri ile estetik değeri olan *Leucojum aestivum* süs bitkisi olarak kullanılmasının yanında değerli bir tıbbi bitkidir. Soğanlarında bulunan çeşitli alkaloidler tıpta birçok hastalığın tedavisinde değerlendirilmektedir. Bu alkaloidler arasında galanthamine Alzheimer hastalığının tedavisinde kullanılan bir acetylcholinesterase enzim inhibitörü olması nedeniyle çok önemlidir. Bu nedenle *L. aestivum* habitatlarını korumak ve sürdürülebilirliğini sağlamak önem arz etmektedir. Ülkemizde *L. aestivum* ile ilgili vejetatif ve generatif çoğaltma çalışmalarının yanında in-vitro çoğaltma, içeriğindeki alkaloidlerin belirlenmesi ve bazı karyolojik araştırma çalışmaları yürütülmüştür. Türkiye’de *L. aestivum*’un ‘Doğal Çiçek Soğanlarının Doğadan Toplanması, Üretimi ve İhracatına İlişkin Yönetmelik’e göre ihracatı kotaya tabi tutulmaktadır. Ancak bunun dışında bilinen başka bir koruma çalışması bulunmamaktadır. Bu derleme, nesli tehlike altında olan ve Türkiye florasında doğal olarak bulunan, *L. aestivum* ile ilgili genel bilgilerin yanında türün koruma ve araştırma çalışmaları hakkında bilgi vermek ve bu türün önemine dikkat çekmek amacıyla hazırlanmıştır.

Anahtar kelimeler: IUCN Kırmızı Liste, Koruma, *Leucojum aestivum*, Tehditler, Tehlikede

Research and Conservation Studies on Endangered Snowflake (*Leucojum aestivum*) in Turkey


Abstract: Turkey has a special place in genetic diversity of geophytes. However, many genetic resources of geophytes including *Leucojum aestivum* L. (Amaryllidaceae) known as snowflake are under genetic erosion and face with the danger of extinction. *L. aestivum* is the single *Leucojum* species in Turkey and naturally grows in Bolu, İstanbul, Bursa, Erzurum, Kocaeli, Konya, Samsun. According to IUCN (International Union for the Conservation of Natural Resources) Red List Category of this species is Least Concern. Major threats of *L. aestivum* such as grazing and unfavorable land


management are deteriorating its natural habitats. *Leucojum aestivum* is a seasonal ornamental plant because of its aesthetic white flowers and a valuable medicinal plant. It has various alkaloids used in medicine. Especially galanthamine among them is very important as an acetylcholinesterase inhibitor used in Alzheimer's treatment. Therefore it is important to protect the population of this species and to ensure its sustainability. Numerous studies have been carried out in *L. aestivum* on vegetative, generative and in-vitro propagation, determination of alkaloids and karyological research studies in Turkey. *L. aestivum* has quota for trade according to 'Regulation on collection, production and marketing of natural flower bulbs' in Turkey, but there is no further known conservation action. This review has been prepared to give some information about the endangered *L. aestivum* and its possible conservation actions in Turkey's flora in order to take attention on this important natural species of Turkey.

Keywords: IUCN Red List, Conservation, *Leucojum aestivum*, Threats, Endangered

*Corresponding author: Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Bahçe Bitkileri Bölümü, 55200, Atakum, Samsun, Türkiye

E mail: dmrsevm@gmail.com (S. DEMİR)

Fisun Gürsel ÇELİKEL  <https://orcid.org/0000-0002-4722-2693>

Sevim DEMİR  <https://orcid.org/0000-0001-8523-6175>

Cite as: Çelikel FG, Demir S. 2020. Research and conservation studies on endangered snowflake (*Leucojum aestivum*) in Turkey. BSJ Eng Sci, 3(4): 160-164.

1. Giriş

Dünyanın başlıca gen merkezlerinin arasında yer alan Türkiye bitki genetik çeşitliliği açısından oldukça özel bir konumda bulunmaktadır (Karagöz ve ark., 2010). Ülkemiz Avrupa ve Asya Anakaralarında 78 milyon ha'lık alanda 4080'i endemik olmak üzere toplam 12476 takson barındırmaktadır (Karagöz ve ark., 2010). Bitkisel çeşitliliğimizin zenginliği temel olarak, ülkemiz topografyası, jeolojik yapısı ve buna bağlı olarak değişik karakterdeki toprak yapısı ve farklı iklim koşullarından kaynaklanmaktadır (Dilaver, 2013). Ancak geofitlerin de içerisinde yer aldığı birçok bitki genetik kaynağımız çevresel ve diğer baskılarla gen erozyonuna uğramakta ve yok olma tehlikesi ile karşı karşıya kalmaktadır.

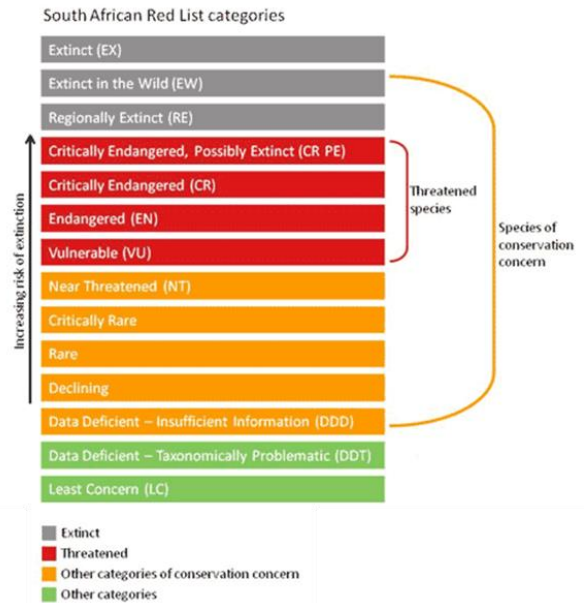
Leucojum aestivum L., ülkemizde doğal olarak yetişen, *Leucojum* cinsine ait tek türdür. Geofit grubuna giren, değişikliğe uğramış toprak altı gövdesi (soğan) bulunduran *Leucojum aestivum* L. ülkemizde 'Göl soğanı' olarak bilinmektedir. Ancak *L. aestivum* ülkemizde ve Dünya'da nesli tehlike altında olan geofitler arasındadır. Bu tür IUCN (International Union for the Conservation of Natural Resources) kırmızı liste kategorilerine göre Düşük Riskli (LC) sınıfında yer almaktadır (Lansdown, 2014). Süs bitkisi olarak değerlendirilen *L. aestivum*, aynı zamanda çok önemli bir tıbbi bitkidir. Başta galanthamine olmak üzere soğanlarında bulunan çeşitli alkaloidler tıpta kullanılmaktadır. Ayrıca *L. aestivum* ticari değeri olan, ülkemizden ihracatı yapılan doğal çiçek soğanları arasında yer almaktadır (Çiçek ve ark., 2013). Bu derleme, nesli tehlike altında olan ve Türkiye florasında doğal olarak bulunan, *L. aestivum* ile ilgili genel bilgilerin yanında türün koruma ve araştırma çalışmaları hakkında bilgi vermek ve türün önemine dikkat çekmek amacıyla hazırlanmıştır.

2. Doğa ve Doğal Kaynakların Korunması için Uluslararası Birlik (IUCN)

IUCN (International Union for Conservation of Nature and Natural Resources) Kırmızı Liste Sınıfları ve Ölçütleri,

farklı türleri tükenme risklerine göre sınıflandırmak, tükenme riski yüksek olan türlere dikkat çekmek, koruma önlemleri arasındaki öncelikleri saptamak amacıyla tasarlanmıştır (IUCN, 2001).

IUCN kapsamında oluşturulan Kırmızı Liste Sınıfları; Tükenmiş, EX; Doğada Tükenmiş, EW; Kritik, CR; Tehlikede, EN; Duyarlı, VU; Tehdite Yakın, NT; Düşük Riskli, LC; Yetersiz Veri, DD; Değerlendirilmemiş, NE olarak belirlenmiştir (Allen ve ark., 2014). Kırmızı liste sınıfları Şekil 1'de gösterilmiştir.



Şekil 1. IUCN, kırmızı liste sınıfları (redlist.sanbi.org).

3. Türkiye'nin Nesli Tehlike Altında Olan *Leucojum* Türü

3.1. *Leucojum aestivum*

Türkiye'de doğal olarak yetiştiği bilinen Amaryllidaceae familyasından *Leucojum* cinsine ait tek tür *Leucojum aestivum*'dur. Göl soğanı olarak bilinen *L. aestivum* çok yıllık, otsu yapıda, soğanlı bitkilerdir. Bataklık, göl

kenarları gibi sulak alanlarda doğal olarak yetişmektedir. *L. aestivum* türünde çiçeklenme Şubat-Nisan aylarında gerçekleşmektedir (Lansdown, 2014). *L. aestivum* çiçekleri Şekil 2'de gösterilmiştir.



Şekil 2. *L. aestivum* çiçekleri (<https://en.wikipedia.org/wiki/Leucojum>).

Tehdit sınıfı: LC, Düşük Riskli (Least Concern) sınıfında yer almaktadır.

Ana tehditler: Bilinçsiz arazi yönetimi ve aşırı otlatma ana tehditleridir.

Endemik: Endemik değildir.

Yayılım alanı: Arnavutluk, Avusturya, Belçika, Bulgaristan, Kanada, Hırvatistan, Çek Cumhuriyeti, Fransa, Gürcistan, Almanya, Yunanistan, Macaristan, İran, İtalya, Montenegro, Hollanda, Romanya, Sırbistan, Slovenya, İsviçre, Türkiye, Ukrayna ve Birleşik Krallık *L. aestivum*'un yayılış gösterdiği ülkelerdir (Lansdown, 2014). Türkiye'de Bolu, İstanbul, Bursa, Erzurum, Kocaeli, Konya ve Samsun illerinde yayılış göstermektedir (Şekil 3).



Şekil 3. *L. aestivum*'un Türkiye'deki dağılım alanları (www.tehditaltindabitkiler.org).

4. Koruma Çalışmaları

Başta aşırı ve zamansız söküm olmak üzere yol, bina, baraj yapımı, sanayileşme, hızlı şehirleşme, turizm faaliyetleri, taş ocakçılığı, rekreasyonel faaliyetler, bilinçsiz otlatma gibi sebepler ile birçok geofitin nesli tehlike altına girmiştir (Balkaya ve ark., 2015; Çelikel,

2015). Tarım ve Köy İşleri Bakanlığı 1989 yılında doğal çiçek soğanlarının ihracatını kontrol altına almak için, ülkemiz florasının korunmasına yönelik 'Doğal Çiçek Soğanlarının Sökümü, Üretimi ve İhracatına ait Yönetmelik'i çıkarmıştır (Çelikel, 2014; Demir ve Çelikel, 2017a, 201b). Bu yönetmelik kapsamında birçok doğal çiçek soğanının ihracatı yasaklanmıştır. İhracata izin verilenlerde doğadan toplama ve üretim kontenjanları getirilmiş, ihraç edilecek soğan büyüklükleri belirlenmiştir. Ayrıca ülkemizde 1996 yılı itibarıyla CITES (Nesli Tehlike Altında olan Yabani Hayvan ve Bitki Türlerinin Uluslararası Ticaretine İlişkin Sözleşme) kuralları uygulanmaktadır (Çelikel, 2014; Çelikel, 2015). Yönetmelik kapsamında ülkemizde doğal olarak yetişen ve nesli tehlike altında olan *L. aestivum* ihracatı kotaya tabi olan türler arasında yer almaktadır (Resmigazete, 2020).

5. Araştırma Çalışmaları

5.1. Dünyada Yapılan Çalışmalar

Stanilova ve ark. (1994), nesli tehlike altında olduğu bildirilen *L. aestivum* türünde *in vitro* çoğaltma çalışması yapmışlardır. En yüksek rejenerasyon yaprak ekstraktlarından elde edilmiştir. Ortam olarak Murashige Skoog (MS) + 1 mg/L kinetin + 1 mg/L 6-benzylaminopurine (BAP) ve Linsmaier Skoog (LS) + 0,1 mg/L kinetin + 0,5 mg/L 1-naphthaleneacetic acid (NAA) önerilmiştir.

Hudson ve ark. (2000), Türkiye'de bulunan *Galanthus elwesii* ve *Leucojum aestivum*'un çok güçlü anti-SINV (Sindbis virus) potansiyelinin olduğunu saptamışlardır. Türkiye'ye ait bu tıbbi bitkilerin antiviral aktivite kaynağı olabileceğini belirtmişlerdir.

Pavlov ve ark. (2007), *in vitro* yöntemlerle *L. aestivum* sürgünlerinden galanthamine üretimi gerçekleştirmişlerdir.

Berkov ve ark. (2008), *L. aestivum* yapraklarından N-(14-Methylallyl) norgalanthamine, N-allylnorgalanthamine, galanthamine, narwedine, epinorgalanthamine ve lycorine alkaloidlerini izole etmişlerdir. Bu alkaloidlerden N-allylnorgalanthamine ve N-(14-methylallyl) norgalanthamine AChE (acetylcholinesterase) inhibe etme konusunda galanthamine göre daha etkili olduğu belirlenmiştir.

Berkov ve ark. (2011), *Narcissus confuses*, *Narcissus pseudonarcissus* ve *Leucojum aestivum* yaprak ve soğanlarından GC-MS ile galanthamine izole etmişlerdir.

Ivanov ve ark. (2012, 2013), *Leucojum aestivum*'un içerdiği alkaloidler (özellikle galanthamine) sebebiyle farmakolojik aktivitelere sahip olduğunu saptamışlardır.

5.2. Türkiye'de Yapılan Çalışmalar

Şener ve ark. (1999) *Leucojum aestivum* soğanlarından galanthamine ve epihydroxybulbispermine alkaloidlerini izole etmişlerdir. Ayrıca *L. aestivum* ekstraktlarının *Enterobacter cloacea*, *Corynebacterium hoffmanni*, *Pseudomonas pseudomolli*, *Vibrio cholerae*, *Corynebacterium diphtheriae* ve *Salmonella typhi*'e karşı

anti bakteriyel, *Curvularia lunata*, *Nigrospora oryzae*, *Microsporium canis*, *Pleuralus ostreatus*, ve *Trichophyton longifusus*'a karşı anti fungal aktiviteye sahip olduğu da çalışma sonucunda ortaya konmuştur. *L. aestivum*'dan elde edilen lycorine, tazettine ve galanthamine alkaloidlerinin antimaral aktiviteye sahip olduğu saptanmış ve *L. aestivum* ekstraktlarının insektisit özellik gösterdiği de belirtilmiştir.

Aksu ve ark. (2001) *Leucojum aestivum* soğanlarının 'chipping' (dilimlere ayırma) ve 'twin-scaling' (ikiz pullara ayırma) yöntemiyle çoğaltılması üzerine bir araştırma yapmışlardır. Çalışmada çevre uzunlukları 6-7 cm'den 12-13 cm'e kadar olan soğanlar kullanılmıştır. Dilimleme yönteminde soğanlar büyüklüklerine göre dörde ve sekize dilimlenmiştir. İnkübasyon periyodunun sonunda uygulamalardan elde edilen soğancık oluşumunun %94-97 arasında olduğu saptanmıştır. 20 °C'e 12 hafta süren inkübasyon periyodundan sonra dikimleri yapılan soğan parçalarından iki yıllık bir dönemin sonunda elde edilen soğan sayısı açısından en iyi sonuçlar, ikiz pullara ayırma uygulamalarından, dilimleme yönteminde ise sekize dilimleme uygulamalarından elde edilmiştir (Aksu ve ark., 2001).

Şenel ve ark. (2002) Amaryllidaceae familyasından Türkiye'nin Karadeniz Bölgesi'nde doğal olarak yayılış gösteren *Pancreatum maritimum*, *Leucojum aestivum* ve *Galanthus rizehensis* türleri üzerine karyolojik çalışma yapmışlardır. Çalışma sonucunda *P. maritimum* ve *L. aestivum*'un kromozom sayılarının 2n:22, *G. rizehensis*'in kromozom sayısının ise 2n:26 olduğu saptanmıştır.

Orhan ve Şener (2003), Türkiye'de doğal olarak yetişen, *Galanthus ikariae*, *Galanthus elwesii*, *Leucojum aestivum*, *Narcissus tazetta* subsp. *tazetta* ve *Pancreatum maritimum* türlerinin anticholinesterase aktiviteye sahip olduğunu saptamışlardır. Acetylcholinesterase enzim inhibitörleri Alzheimer hastalığının tedavisinde kullanılmaktadır.

Şener ve ark. (2003), Türkiye'de bulunan *Leucojum aestivum*, *Pancreatum maritimum* ve *Narcissus tazetta* ssp. *tazetta* üzerine yapmış oldukları çalışmada, bu bitkilerde bulunan lycorine, crinine, tazettine ve galanthamine alkaloidlerinin antimalaryal aktiviteye sahip olduklarını saptamış ve *Plasmodium falciparum*'u inhibie etme potansiyellerinin bulunduğu belirtilmiştir.

Ayan ve ark. (2004), gölgeleme ve farklı büyüme düzenleyicilerin *L. aestivum*'da soğan verimi üzerine olan etkisi incelenmiştir. Bu çalışmanın sonuçlarına göre gölgeleme ile birlikte GA3 ve NAA uygulamalarının soğan verimini artırdığı saptanmıştır.

Çırak ve ark. (2005), *L. aestivum*'larda, yaprak sayısı ve soğan verimi arasında pozitif bir korelasyon olduğunu bildirmişler ve *L. aestivum*'larda yaprak alanı tahminleme modeli geliştirmişlerdir.

Zencirkıran (2005), *Leucojum aestivum*'un vazo ömrü üzerine yaptığı çalışmada, 1 mM STS (gümüş tio sülfat) + %10 sakkarozun vazo ömrünü uzattığını belirtmiştir.

Çiçek ve ark. (2007), stratifikasyonun *Leucojum aestivum* L. tohum çimlenmesi üzerine etkilerini incelemişlerdir. Soğukta katlama (4°C) tohumlarda dormansiyi ortadan kaldırmaz iken, en yüksek çimlenme yüzdesi 6 hafta 22 °C'de stratifikasyona tabi tutulan tohumlarda saptanmıştır. 7 hafta 22 °C'de katlamaya tabi tutulan tohumlarda ise katlama sırasında çimlenmenin gerçekleştiği belirtilmiştir.

Karaoğlu (2010) bazı geofitlerin tohum bağlayamaması, bazı geofitlerin ise çiçek açabilecek büyüklüğe gelebilmeleri için 4-5 yıla ihtiyaç duymaları sebebiyle *in vitro* hızlı çoğaltımın *Leucojum* gibi bitkilerin üretiminde alternatif bir yöntem olabileceğini belirtmiştir. Ayrıca Karaoğlu (2010) *in vitro* hızlı çoğaltımın kullanılmasıyla hem bu bitkilere olan taleplerin karşılanmış olacağını, hem de bu bitkilerin doğadan aşırı sökümleri neticesinde nesillerinin tehlike altına girmesinin engellenebileceğini vurgulamıştır.

Ekici ve Dane (2012), *L. aestivum*'un bazı sitolojik ve embriyolojik karakterleri üzerine çalışmışlardır. Erkek gametlerin gelişiminde bazı düzensizlikler olduğu, fakat bu düzensizliklerin döllenmeyi olumsuz etkilemediği belirtilmiştir. Ayrıca yapılan çalışmada *Leucojum* ların tapetumlarının (anter çeperinin en iç tabakası), aynı familyadan olan *Galanthus* cinsine ait türlerden farklı olduğu saptanmıştır.

Çiçek ve ark. (2013), kurutma, çimlendirme sıcaklığı ve muhafaza süresinin *L. aestivum* tohumlarının çimlenmesine etkisini incelemişlerdir. Yaş ve kuru tohumlarda en yüksek çimlenme 20 °C sıcaklıkta gerçekleşmiştir. 20 °C'de muhafaza edilen tohumlarda 4 ve -5 °C'de muhafaza edilen tohumlara göre çimlenme oranının daha yüksek olduğu saptanmıştır. Ayrıca tohum muhafaza süresi uzadıkça çimlenme oranının düştüğü belirtilmiştir.

Demir (2014), yapmış olduğu çalışmada, *L. aestivum*'un Alzheimer hastalığının tedavisinde kullanılan galanthamine alkaloidini içermesinden dolayı tıbbi ve ekonomik değerinin olduğunu bildirmiştir.

6. Sonuç

Ülkemizde *L. aestivum* ile ilgili vejetatif ve generatif çoğaltma çalışmalarının yanında *in-vitro* çoğaltma çalışmaları, karyolojik çalışmalar ve içeriğindeki alkaloidlerin belirlenmesi üzerine çalışmalar yürütülmüştür. Türkiye'de *L. aestivum* 'Doğal Çiçek Soğanlarının Doğadan Toplanması, Üretimi ve İhracatına İlişkin Yönetmelik'e göre ihracatı kotaya tabi tutulmaktadır. Ancak bunun dışında bilinen başka bir koruma çalışması bulunmamaktadır.

Beyaz çiçekleri ile estetik değeri olan *Leucojum aestivum* süs bitkisi olarak kullanılmasının yanında değerli bir tıbbi bitkidir. Soğanlarında bulunan çeşitli alkaloidler tıpta birçok hastalığın tedavisinde değerlendirilmektedir. Özellikle soğanlarında bulunan galanthamine AChE (acetylcholinesterase) inhibitörü olmasıyla Alzheimer hastalığının tedavisinde kullanılmaktadır. Bunun dışında

L. aestivum'un anti bakteriyel, anti fungal, antimalaryal ve insektisit aktiveye sahip olması bu türün önemini artıran faktörler arasındadır. Bu nedenle *L. aestivum* habitatlarını korumak ve sürdürülebilirliğini sağlamak önem arz etmektedir.

Sonuç olarak, ülkemizde doğal olarak yetişen *L. aestivum*'un hem süs bitkisi olarak değerlendirilmesi hem de sahip olduğu farmatik maddelerin devamlılığı açısından olduğu gibi, çok önemli bir gen kaynağı olan bu türün korunması biyoçeşitlilik açısından da oldukça önemlidir.

Çıkar İlişkisi

Yazar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Teşekkür ve Bilgilendirme

Bu çalışma 28-30 Kasım 2018 tarihlerinde Samsun'da düzenlenen "The 2nd International UNIDOKAP Black Sea Symposium on BIODIVERSITY" isimli sempozyumda sunulmuştur.

Kaynaklar

- Aksu E, Görür G, Çelikel FG. 2001. Göl Soğanının (*Leucojum aestivum*) vegetatif yöntemlerle çoğaltma imkânlarının araştırılması. Atatürk Bahçe Kültürleri Merkez Araştırma Enstitüsü, Bilimsel Araştırmalar ve İncelemeler Yayın No: 150, 31 s.
- Allen D, Bilz M, Leaman DJ, Miller RM, Timoshyna A, Window J. 2014. European red list of medicinal plants, <http://www.iucnredlist.org/initiatives/europe/publications>, (erişim tarihi: 1 Mayıs 2018).
- Ayan AK, Kurtar ES, Çırak C, Kevseroğlu K. 2004. Bulb yield on some plant character of summer snowflake (*Leucojum aestivum* L.) under shading as affected by GA3 and NAA at different concentrations. J Agro, 3(4): 296-300.
- Balkaya A, Duman İ, Engiz M, Ermiş S, Onus N, Özcan M, Çelikel F, Demir İ, Kandemir D, Özer M. 2015. bahçe bitkileri tohumluğu üretimi ve kullanımında değişimler ve yeni arayışlar. Türkiye Ziraat Mühendisliği VIII. Teknik Kongresi, 12-16 Ocak 2015, Ankara, TMMOB Ziraat Mühendisleri Odası, 2: 985-1110.
- Berkov S, Codina C, Viladomat F, Bastida J. 2008. N-Alkylated galanthamine derivatives: Potent Acetylcholinesterase inhibitors from *Leucojum aestivum*. Bioorganic Medic Chem Lett, 18: 2263-2266.
- Berkov S, Bastida J, Viladomat F, Codina C. 2011. Development and validation of a GC-MS Method for rapid determination of galanthamine in *Leucojum aestivum* and *Narcissus* ssp.: A metabolomic approach. Talanta, 83:1455-1465.
- Çelikel FG. 2014. Doğal çiçek soğanları ve süs bitkileri ders notları, Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Bahçe Bitkileri Bölümü, Samsun.
- Çelikel FG. 2015. Süs bitkilerinde tohumluk (tohum, fide, fidan, soğan) üretimi ve kullanımı. SÜSBİR Derg, 3: 32-33.
- Çırak C, Odabaş MS, Ayan AK. 2005. Leaf area prediction model for summer snowflake (*Leucojum aestivum* L.). Int J Botany, 1(1): 12-14.
- Çiçek E, Aslan M, Tilki F. 2007. Effect of stratification on germination of *Leucojum aestivum* L. seeds, a valuable ornamental and medicinal plant. Res J Agri and Biol Sci, 3(4): 242-244.

- Çiçek E, Çetin B, Özbayram AK, Türkyılmaz H. 2013. Kurutma, çimlendirme sıcaklığı ve saklamanın göl soğanı (*Leucojum aestivum* L.) tohumlarının çimlenmesine etkisi. Artvin Çoruh Üniv Orman Fak Derg, 14(2): 245-252.
- Demir A. 2014. Medical resource value appraisal for *Leucojum aestivum* in Turkey. American J Alzheimer's Dis & Other Dem, 29(5): 448-451.
- Demir S, Çelikel FG. 2017a. Türkiye'de doğal olarak yetişen kum zambağının (*Pancreatum maritimum*) farklı özellikleri ve önemi. TURJAF, 5(13): 1726-1731.
- Demir S, Çelikel FG. 2017b. Türkiye'nin nesli tehlike altında olan *Lilium* sp. türleri. TURJAF, 5(13): 1796-1801.
- Dilaver Z. 2013. Conservation of natural plants and their use in landscape architecture, advances in landscape architecture (Chapter 35), InTech, pp. 885-904.
- Ekici N, Dane F. 2012. Microsporogenesis, pollen mitosis and *in vitro* pollen tube growth in *Leucojum aestivum* (Amaryllidaceae), Botany, InTech, Available from: <http://www.intechopen.com/books/botany/microsporogenesis-pollen-mitosis-and-in-vitro-pollentube-growth-in-leucojum-aestivum-amaryllidaceae>.
- Hudson JB, Lee MK, Sener B, Erdemoglu N. 2000. Antiviral activities in extracts of Turkish medicinal plants. Pharma Biol, 38(3): 171-175.
- IUCN 2001. IUCN kırmızı liste sınıfları ve ölçütleri, <http://life.bio.sunysb.edu/ee/akcakayalab/kurallar5.pdf>.
- Ivanova I, Georgieva V, Berkov S, Pavlov A. 2012. Alkaloid patterns in *Leucojum aestivum* shoot culture cultivated at temporary immersion conditions. J Plant Phys, 169: 206- 211.
- Ivanova I, Georgieva V, Pavlov A. 2013. Elicitation of galanthamine biosynthesis by *Leucojum aestivum* liquid shoot cultures. J Plant Phys, 170: 1122-1129.
- Karagöz A, Zencirci N, Tan A, Taşkın T, Köksel H, Sürek M, Tokar C, Özbek K. 2010. Bitki genetik kaynaklarının korunması ve kullanımı, Türkiye Ziraat Mühendisliği VII. Teknik Kongresi, 11-15 Ocak 2010, Ankara, 155-177.
- Karaoğlu C. 2010. Soğanlı bitkiler ve *in vitro* hızlı çoğaltım. Tarla Bit Merk Araş Enst Derg, 19(1-2): 24-29.
- Lansdown RV. 2014. *Leucojum aestivum*. The IUCN red list of threatened species 2014: e.T164488A45461549. <http://dx.doi.org/10.2305/IUCN.UK.2014.1.RLTS.T164488A45461549.en>. Downloaded on 12 November 2018.
- Orhan I, Şener B. 2003. Bioactivity-directed fractionation of alkaloids from some Amaryllidaceae plants and their anticholinesterase activity. Chem of Natural Comp, 39(4): 383-386.
- Pavlov A, Berkov S, Courot E, Gocheva T, Tuneva D, Pandova B, Georgiev M, Georgiev V, Yanev S, Burrus M, Ilieva M. 2007. Galanthamine production by *Leucojum aestivum in vitro* systems. Process Biochem, 42: 734-739.
- Resmigazete 2020. Doğal çiçek soğanlarının ihracat listesi, www.resmigazete.gov.tr (erişim tarihi, 10.02.2020).
- Stanilova MI, Ilcheva VP, Zagorska NA. 1994. Morphogenetic potential and *in vitro* micropropagation of endangered plant species *Leucojum aestivum* L. and *Lilium rhodopaeum* Delip. Plant Cell Rep, 13: 451-453.
- Şener B, Koyuncu M, Bingöl F, Muhtar F. 1999. Production of bioactive alkaloids from Turkish geophytes. IUPAC, <http://www.iupac.org/symposia/proceedings/phuket97/sener.html>.
- Şenel G, Ozkan M, Kandemir N. 2002. A Karyological investigation on some rare and endangered species of Amaryllidaceae of Turkey. Pak J Bot, 34(3): 229-235.
- Şener B, Orhan I, Satayavivad J. 2003. Antimalarial activity screening of some alkaloids and the plant extracts from Amaryllidaceae. Phytother Res, 17: 1220-1223.
- Zencirkıran M. 2005. Effects of sucrose and silver thiosulphate pulsing on stem-base cracking and vase-life in *Leucojum aestivum* flowers. J Hort Sci & Biotech, 80(3): 332-334.



Review

Volume 3 - Issue 4: 165-172 / October 2020

ONCOGENIC VIRUSES AND THEIR MOLECULAR DIAGNOSIS IN POULTRY

Berkant Ismail YILDIZ^{1*}, Mustafa ERSAL¹, Kemal KARABAG²

¹Akdeniz University, Graduate School of Natural and Applied Sciences, 07058, Antalya, Turkey

²Akdeniz University, Faculty of Agriculture, Department of Agricultural Biotechnology, 07058 Antalya, Turkey

Received: March 13, 2020; **Accepted:** March 31, 2020; **Published:** October 01, 2020


Abstract


Infectious animal diseases caused by pathogenic microorganisms such as bacteria and viruses threaten the health and well-being of wildlife, livestock and human populations, limit productivity and causes significant economic losses for each. Avian oncogenic viruses are one of the most dangerous pathogenic microorganisms that threaten the poultry industry and cause damage of over billions dollars annually worldwide. These viruses include a highly contagious herpesvirus Marek's disease virus (MDV), as well as retroviruses such as avian leukosis virus (ALV) and reticuloendotheliosis virus (REV). Each group is distinguished by its nucleic acid type, antigenicity, epidemiology, host range and molecular characteristics. These viruses are cells associated with tumors and are in all organs except in the feather follicle where enveloped infectious virions egress from the body. It is difficult to detect diseases caused by these viruses and at the same time, vaccines that can provide sterile immunity against these diseases and prevent infection are incomplete. In order to be able to fight oncogenic viruses more effectively in the future, it is vital to learn more about the host immunity-oncovirus interaction and to determine powerful diagnostic techniques. In this review, oncogenic viruses and effective diagnostic techniques for these viruses are emphasized in poultry.


Keywords: Avian oncogenic viruses, Molecular diagnosis, Marek's disease virus, Avian leukosis virus, Reticuloendotheliosis virus

***Corresponding author:** Akdeniz University, Graduate School of Natural and Applied Sciences, 07058, Antalya, Turkey

E mail: berkantildiz@gmail.com (B.I. YILDIZ)

Berkant Ismail YILDIZ  <https://orcid.org/0000-0001-8965-6361>

Mustafa ERSAL  <https://orcid.org/0000-0002-2401-1201>

Kemal KARABAG  <https://orcid.org/0000-0002-4516-6480>

Cite as: Yildiz BI, Ersal M, Karabag K. 2020. Oncogenic viruses and their molecular diagnosis in poultry. BSJ Eng Sci, 3(4): 165-172.

1. Introduction

The poultry industry has great importance in providing the protein needed for human nutrition, and is the most consumed meat-poultry meat in the world, especially in OECD countries (OECD., 2014). The global population is expected to reach 9 billion by 2050 and the increased

income growth among the poor population will result in an unprecedented increase in animal protein demand (King et al., 2018). In this context, attention should be given to combating infectious diseases that cause various losses in the poultry industry, which play an important role in ensuring sustainable food supply. Among these

diseases, neoplastic diseases caused by viruses are a major economic problem facing the poultry industry worldwide (Mitra et al., 2012).

The oncogenic viruses causing neoplastic infections in chickens are herpesviruses comprising of Marek's disease virus (MDV), retroviruses comprising of reticuloendotheliosis virus (REV) and avian leukosis virus (ALV) (Davidson and Silva, 2008). The others, retrovirus and lymphoproliferative disease virus have caused significant losses from lymphomas in turkeys in the United Kingdom and Israel, but now appears to be rare (Payne and Venugopal, 2000) and it is not mentioned further in this paper.

Avian oncogenic viral lesions are often similar and it is difficult to make a differential diagnosis based on histopathology (Wang et al., 2014). In particular, the lesions caused by ALV, MDV and REV are very similar and there is no specific symptom in any of these three diseases (Gimeno et al., 2005). The laboratory assays for diagnosis are based on virus isolation, demonstration of specific antibodies and histopathological examination of tumor tissues (Mitra et al., 2012). In addition to their oncogenic property, these retroviruses are also immunosuppressive and may contaminate poultry vaccines (Biswas et al., 2011; Fadly and Witter, 1997). These contaminations have resulted in the large number of deadly infection (Fadly and Garcia, 2006; Li et al., 2013; Wei et al., 2012; Fadly et al., 1996).

2. Avian Leukosis Virus (ALV)

Grooming Avian leukosis virus (ALV) infection of chickens is widespread and known to be of significant economic importance; economic losses due to ALV induced tumors and reduced productivity are estimated to be in millions of U.S. dollars each year (Fadly, 2000). The term leukosis embraces several different leukaemia-like proliferative diseases of the haemopoietic system caused by ALV; the term leukosis is used because a leukaemic blood picture is not always present (De Boer, 1987; Payne and Fadly, 1997).

Based on properties of viral envelope glycoproteins, ALV from chickens, a member of the leukosis/sarcoma (L/S) group of avian retroviruses are classified into six subgroups: A, B, C, D, E and J (Coffin, 1992; Payne and Fadly, 1997) and are classified as exogenous ALVs (A, B, C, D, J) and endogenous ALV (E) based on the viral envelope glycoprotein properties (Fadly, 2000). Particularly, the viral envelope glycoprotein is responsible for attachment and receptor specificity as well as the production of neutralizing antibodies (Coffin, 1992; Payne et al., 1992). Of the viral subgroups so far identified, subgroups A, B, and J are considered most prevalent and more economically important (Dai et al., 2015). Exogenous ALVs are capable of inducing lymphoid leukosis and subgroup A (ALV-A) is more commonly isolated than any other subgroup (Payne and Fadly, 1997). While exogenous ALVs have been shown to cause several neoplastic diseases in infected chickens (Cooper

et al., 1968; Crittenden et al., 1980), and nonneoplastic diseases such as myocarditis (Gilka and Spencer, 1990), and osteopetrosis (Smith, 1982), ALV-E is not known to be pathogenic to chickens (Crittenden et al., 1979; Linial and Neiman, 1976). Subgroup J was first isolated in meat-type chicken in the United Kingdom in 1989 but it is currently causing devastation in the poultry industry worldwide (Payne et al., 1991) and subgroup J associated primarily with myeloid leukosis in meat-type chickens (Fadly and Smith, 1999). Recent laboratory observations provided evidence for recombination between subgroup A and J ALV (ALVA/J), a recombinant ALV with the envelope of subgroup A and long terminal repeat (LTR) of subgroup J; this recombinant ALV resulted from passing ALV- J in cells expressing subgroup A envelope (Lupiani et al., 2003). As a potential contaminant of live-virus vaccines of poultry, ALV can also cause significant losses if contaminated vaccines were used in susceptible flocks (Fadly, 2016). Most recently, a recombinant subgroup A ALV containing envelope of ALV-A and LTR of ALV-E was isolated from commercial Marek's disease vaccines (Fadly et al., 2006; Silva et al., 2007). To date, because no commercial vaccines are available for control of ALV infection, eradication of virus infection at the primary breeder level remains to be the principal method for controlling ALV infection in chickens (Nair and Fadly, 2013).

3. Marek's Disease Virus (MDV)

Marek's disease virus (MDV), or Gallid herpesvirus 2 (GaHV-2) is the etiological agent responsible for Marek's disease (MD) in the chicken, a multifaceted disease most widely recognized by the induction of a rapid and extensive malignant T-cell lymphoma (Couteaudier and Denesvre, 2014), and MD results in substantial economic losses estimated at more than 1 billion per year (Morrow and Fehler, 2004).

Although MD was described in 1907 by Joseph Marek, the virus (MDV) was only isolated in 1967 in the United Kingdom (Churchill and Biggs, 1967) and the United States (Nazerian et al., 1968) independently. MDV is cell associated in body organs and tumors, it replicates and exists as enveloped free form in the feather follicles making feathers particularly dander, dust and litter materials loaded with MD virus, thus facilitating virus transmission by air borne route (Calnek and Witter, 1997; Adene and Akpavie, 2004; Frank, 2001). Susceptible chickens infected with the pathogenic MDV suffer cytolysis of the lymphoid organs and a concomitant immunosuppression (Gordon, 1979; Frank, 2001). Such birds mainly die as a result of tumour development in the visceral organs and peripheral nerves (Frank, 2001). Small RNA profiling studies have revealed that some viruses, particularly large DNA viruses such as Marek's disease virus (MDV) encode their own set of miRNAs (Hicks and Liu, 2013), and thus it is thought to silence the tumor suppressor genes of MDV.

The ability of MDV to replicate in the host is related to its

pathogenicity and the MDV genome load in infected chickens contributes to our understanding of the pathogenesis of MDV infection (Baigent et al., 2005; Islam et al., 2006). In addition, several MDV encoded genes, including meq (MDV EcoRI-Q-encoded protein) (Brown et al., 2009; Lupiani et al., 2004), pp38 (MDV phosphoprotein 38) (Cui et al., 1991; Reddy et al., 2002), vIL-8 (MDV-encoded CXC chemokine viral interleukin 8) (Cui et al., 2004; Parcels et al., 2001), and ICP4 (MDV infected-cell peptide 4) (Strasheim et al., 2012; Gennart et al., 2015), play important roles in MDV pathogenesis. Vaccination with MD vaccines is the primary approach used to protect chickens against MD (Sun et al., 2017). Although the immune-protection mechanisms induced by MD vaccines are not fully understood currently, it is recognized that effective immunity to MD requires the involvement and coordinated activation of innate and adaptive immune responses (Haq et al., 2013). Nevertheless, MD outbreaks have continued around the world in recent years, which is likely due to MDV evolution and co-infection with other viruses (Walkden-Brown et al., 2013b; Zhang et al., 2016).

4. Reticuloendotheliosis Virus (REV)

Reticuloendotheliosis virus (REV) is an oncogenic and immunosuppressive retrovirus that causes reticuloendotheliosis (RE) (Niewiadomska and Gifford, 2013), an avian disease mainly characterized by immunosuppression, runting-stunting syndrome and chronic lymphomas (Walker et al., 1983, Purchase et al., 1973). REV has extensive avian hosts including chickens, turkeys, ducks, mallards, geese, peafowl, pheasants, pigeons, Hungarian partridges, Chinese partridges, Attwater's prairie chickens and many other wild birds (Bohls et al., 2006; Jiang et al., 2013). REV infection of susceptible hosts such as chicks usually causes atrophy of the thymus and bursa of Fabricius, impairing the development and immune system functions of infected hosts, resulting in the suppression of host immune responses to some avian vaccines (Bulow, 1977; Yang et al., 2016).

The genomic structure of REV consists of a group-specific antigen (gag), protease (pro), polymerase (pol) and envelope (env) regions flanked by long-terminal repeats (LTRs) (Witter and Fadly, 2003). The gag gene encodes five structural proteins p10, p12, pp18, pp20 and p30. The p30 (30 kDa) protein is the major REV group-specific antigen (Tsai et al., 1985). The env gene encodes two envelope glycoproteins gp90 and gp120 (Tsai et al., 1986). The pol gene encodes a reverse transcriptase similar to those of mammalian type retroviruses and differs from the Avian Leucosis- Sarcoma Virus (Moelling et al., 1975; Bauer and Termin, 1980).

The wide range of host species and potential for contaminations with REV contribute to viral transmission (Sun et al. 2017). REV can be present as a contaminant in a variety of poultry biologics and vaccines (Fadly and Garcia, 2006; Li et al., 2015) and usually REV

can integrate into the genome of large DNA viruses including Marek's Disease and fowlpox (Isfort et al., 1992). Previous serological surveys have revealed that the positive rate of REV was approximately 2.3–23.5% among commercial chicken and turkey flocks in the United States (Witter et al., 1982). Usually, REV infection in chicken flocks is mainly due to REV contamination in poultry vaccines (Fadly and Garcia, 2006; Li et al., 2013; Wei et al., 2012; Fadly et al., 1996).

5. Molecular Diagnosis of Oncogenic Viruses

The frequent overlap of lesions caused by avian oncogenic virus infections requires specific laboratory diagnosis (Davidson, 2001). Nucleic acid amplification, hybridization technology and immunological tests have been adapted to diagnostic applications and identification of agents responsible for a wide variety of infectious diseases (Hafez and Hess, 1999). Different methods have been established for detecting avian oncogenic viruses, including traditional virus isolation plus an antigen-capture enzyme-linked immunosorbent assay (ELISA) for group-specific antigen of virus, immunofluorescence assay (IFA), loop-mediated isothermal amplification (LAMP) and quantitative reverse transcription PCR (Q-RT-PCR) (Kim and Brown, 2004; Zhang et al., 2010). However, each of these methods has limitations (Dai et al., 2015). For instance, ELISA and IFA are both time-consuming and quantitative data can't be acquired by the current LAMP method (Dai et al., 2015).

The PCR has become a routine technique in many research and diagnostic laboratories and real-time PCR has since its introduction in the mid-1990s removed many limitations of the standard PCR, which is significantly limited by its sensitivity (Edwards et al., 2005). In addition to enhanced sensitivity, the benefits of real-time PCR assays over conventional endpoint detection methods include their large dynamic range, a reduced risk of crosscontamination, an ability to be scaled up for highthroughput applications and the potential for accurate target quantification (Nazarenko et al., 1997; Schweiger et al., 2000; Black et al., 2002).

Zeng et al. (2015) reported the advantages of the GeXP-multiplex PCR assay to include its specificity and its high-throughput ability to immunosuppressive viruses. These advantages stem from the use of chimeric and universal primers in a 3-step PCR procedure with different annealing temperatures: the first step amplifies genespecific sequences within specific regions of the chimeric primers; the second step utilises the entire chimeric primer; and the last step uses universal primers for amplification (Zeng et al., 2015). Abdul-Careem et al. (2006) have shown that realtime PCR is 2.5-10 times more sensitive than conventional PCR techniques used for MDV detection. Moreover, PCR methods allow the detection and quantification of viral DNA in dust

collected and concentrated on filters (Islam et al., 2006; Baigent et al., 2013; Walkden-Brown et al., 2013a).

When real-time PCR is widely used to quantify viral genes, a host gene expressed steadily in a host cells or tissue samples as an internal control becomes one pivot point for calculating the copy number of specific viral genes (Dai et al., 2015). And the sensitivity of real-time PCR assay is at least 100 times higher than that of the routine PCR assay (Dai et al., 2015). Also, it has been possible to combine several assays in a single tube with the development of multicolor real-time PCR cyclers and and "ready-to-use" commercial multiplex real-time PCR kits (Hoffmann et al., 2009). Major advantages of multiplexing include a reduced sample requirement, which is especially important when sample material is scarce (Persson et al., 2005; Belak, 2007), and the ability to combine assays with an internal control system (Hoffmann et al., 2006).

6. Detection formats with Real Time PCR

6.1. Detection Formats without Sequence Confirmation of the PCR Product

Fluorescent dyes such as ethidium bromide (Higuchi et al., 1993; Wittwer et al., 1997a; Le Pecq and Paoletti, 1966) that are specific for double-stranded DNA (dsDNA) were the first systems employed in real-time PCR assays. Other intercalating dyes such as YO-PRO-1 have also been used (Ishiguro et al., 1995; Tseng et al., 1997). SYBR Green I is currently the most frequently used intercalating dye in real-time PCR and it has a 100 times higher binding affinity than ethidium bromide and the fluorescence of bound dye is more than 1000-fold higher than that of free dye (Hoffmann et al., 2006).

These properties make SYBR Green I highly suitable for monitoring product accumulation during PCR (Wittwer et al., 1997b; Morrison et al., 1998). Another detection technology known as 'LUX (Light Upon eXtension)' utilizes a modification to one of the two primers such that it possesses a fluorophore located near the 3' end in a hairpin structure (Nazarenko, 2006; Kusser, 2006). The Amplifluor Quantitative PCR detection system uses a similar approach with labelled and unlabelled primers (Nazarenko et al., 1997; Nuovo et al., 1999; Khripin, 2006).

6.2. Detection Formats with Increased Target Specificity

Fluorophorelabelled oligonucleotide probes are most commonly used for the specific detection of target sequences (Cardullo et al., 1988; Clegg, 1995; Wu and Brand, 1994). In these assays, an increase in fluorescence signal proportional to the accumulation of PCR product arises as a consequence of fluorescence resonance energy transfer (FRET) between separate fluorogenic labels (known as the reporter and quencher) conjugated to the probe (or primers). FRET, also called Förster transfer, is a spectroscopic process by which energy is passed over a maximum distance of 70Å between reporter and acceptor molecules possessing overlapping

emission and absorption spectra (Selvin and Hearst, 1994). The most commonly used fluorogenic quenchers are TAMRA and DABCYL, while Black Hole Quencher (BHQ) is also widely used and disperses energy from the reporter as heat rather than fluorescence (Didenko, 2001).

6.3. Hybridisation Probes

Hybridisation probes, also known as 'HybProbes', use a pair of adjacent, fluorogenic hybridisation oligos and are the only detection format that directly measures FRET (Cardullo et al., 1988). These probes have become the preferred chemistry of the manufacturer for the capillary-based LightCycler system (Wittwer et al., 1997a, b) with special filters for the detection of the acceptor fluorophores Red 640 and Red 705.

6.4. Hydrolysis Probes (50-exonuclease Assay)

Hydrolysis probes (commercially called TaqMan1 probes) are dual-fluorophore-labelled oligonucleotides, with a 50-terminal reporter (e.g. FAM) and a 30-terminal quencher (e.g. TAMRA). (Hoffmann et al., 2009). Once the labels are separated by destroying the TaqMan1 probe based on the 50-exonuclease activity of the DNA polymerase (e.g. Taq polymerase), the increase in reporter fluorescence caused by the removal of the adjacent quencher is monitored by a realtime PCR instrument. (Heid et al., 1996; Livak et al., 1995; Gibson et al., 1996). A modification of this strategy exploits the so-called minor groove binding (MGB) probes. MGB probes form extremely stable duplexes with singlestranded DNA targets mediated via van der Waals forces (Afonina et al., 2002): as a consequence shorter-length probes are required for hybridisation. In comparison with unmodified DNA, MGB probes have higher T_m and are reported to hybridize with greater sequence specificity (Afonina et al., 1996; Kutayavin et al., 2000). These short MGB probes are ideal for allele discrimination studies or for detection of single-nucleotide polymorphisms (SNPs) because they are more significantly destabilised by nucleotide changes within the hybridisation site compared with probes of longer length (De Kok et al., 2002; Belousov et al., 2004; Itabashi et al., 2004).

6.5. Molecular Beacons

Molecular beacons are hairpin-shaped oligoprobes terminally labelled with a reporter and a quencher fluorophore (Tyagi and Kramer, 1996; Tyagi et al., 1998; Vet et al., 2002).

6.6. Scorpion Primer

The Scorpion technology is mainly used in allelic discrimination (Whitcombe et al., 1999; Thelwell et al., 2000) and in SNP genotyping (Roberts, 2000).

6.7. Locked Nucleic Acid (LNA) Probes

Incorporation of LNA residues increases the T_m of the oligonucleotide sequence, allowing the use of markedly shorter probes as allele-specific tools in genotyping assays (Costa et al., 2004; Latorra et al., 2003; Braasch and Corey, 2001).

7. Conclusion

This study is primarily concerned with avian oncogenic viruses, the economically most important viruses such as MDV, ALV and REV. And more focused on some of the molecular techniques most commonly used to improve avian oncogenic virus detection for diagnosis and disease control. We showed some advanced biotechnological approaches that allow early detection of pathogens that affect poultry. Conventional diagnostic techniques are frequently time consuming, labor intensive and require to be performed on sophisticated equipment. Real Time PCR is one of the most important of these biotechnological methods. The high sensitivity and ability to quantify viral targets, the substantial gain in specificity and the reduced risk of cross-contamination are important features of this technology. Also, applications of Real-Time PCR include measurements of viral load, gene expression studies, clinical diagnostics and various pathogen detection.

Since it is suitable for the diagnosis of multiple viral infections, molecular techniques such as PCR and/or Real Time PCR are used for the diagnosis of avian oncogenic viruses (Silva et al., 2007). Nevertheless virus isolation is considered the "gold standard" for diagnosis and is often the starting point for more detailed studies (Nair, 2013). Virus-specific amplifications are confirmed by sequencing of precipitated PCR products. Using molecular methods can be useful for the rapid differential diagnosis of avian oncogenic viruses and for the detection of multiple infections (Gopal et al., 2012). Likewise, multiplex PCR primers can be useful in detecting the presence of ALV-J by modification in ALV primer sequences (Gopal et al., 2012). PCR has emerged as a chosen method for rapid and accurate diagnosis of viruses that appear in poultry. Multiplex PCR primers specific to MDV, ALV, REV and chicken DNA can be designed for rapid differential diagnosis (Gopal et al., 2012). Preferred amplification frequently occurs in multiplex PCR as a result of the various efficiencies of different primer pairs and due to the layout of the primer dimers. This is more likely as the number of primers increases (Elnifro et al., 2000), and the problem of amplification failure disappears when the template DNA contains several copies. This problem can be solved by optimizing multiplex PCR by changing the primer sequences, concentrations and cycle conditions (Frumkin et al., 2008).

Conflict of interest

The authors declare that there is no conflict of interest.

Acknowledgements

This article was presented as an oral presentation at the International Congress on Domestic Animal Breeding Genetics and Husbandry (ICABGEH-2019) held on 11-13 September 2019 in Prague.

References

- Abdul-Careem MF, Hunter BD, Nagy E, Read LR, Sanel B, Spencer IL, Sharif S. 2006. Development of a real-time PCR assay using SYBR green chemistry for monitoring Marek's disease virus genome load in feather tips. *J Virol Methods*, 133: 34-40.
- Adene DF, Akpavie SO. 2004. An overview of Marek's disease and the options in its control, *Poultry Health and Production Principles and Practices*. Sterling Hornden Publishers Nig. LTD. Ibadan, Nigeria. pp. 107-127.
- Afonina IA, Kutyaev I, Lukhtanov E, Meyer RB, Gamper H. 1996. Sequence-specific arrest of primer extension on single-stranded DNA by an oligonucleotide-minor groove binder conjugate. *Proc Natl Acad Sci*, 93(8): 3199-3204.
- Afonina IA, Reed MW, Lusby E, Shishkina IG, Belousov YS. 2002. Minor groove binder-conjugated DNA probes for quantitative DNA detection by hybridization-triggered fluorescence. *Biotechniques* 32(4): 946-949, 940-944.
- Baigent SJ, Kgosana L, Gamawa A, Smith LP, Read AF, Nair V. 2013. Relationship between levels of very virulent MDV in Poultry dust and in feather tips from vaccinated chickens. *Avian Dis*, 57:440-447.
- Baigent SJ, Petherbridge LJ, Howes K, Smith LP, Currie RJ, Nair VK. 2005. Absolute quantitation of Marek's disease virus genome copy number in chicken feather and lymphocyte samples using real-time PCR. *J Virol Methods*, 123: 53-64.
- Bauer G, Temin HM. 1980. Specific antigenic relationships between the RNA-dependent DNA polymerases of avian reticuloendotheliosis viruses and mammalian type C retroviruses. *J Virol*, 34(1): 168-177.
- Belak S, 2007. Molecular diagnosis of viral diseases, present trends and future aspects. A view from the OIE Collaborating Centre for the Application of Polymerase Chain Reaction Methods for Diagnosis of Viral Diseases in Veterinary Medicine. *Vaccine*, 25: 5444-5452.
- Belousov YS, Welch RA, Sanders S, Mills A, Kulchenko A, Dempcy R, Afonina IA, Walburger DK, Glaser CL, Yadavalli S, Vermeulen NM, Mahoney W. 2004. Single nucleotide polymorphism genotyping by two colour melting curve analysis using the MGB Eclipse Probe System in challenging sequence environment. *Hum Genomics*, 1(3): 209-217.
- Biswas S.K, Jana C, Chand K, Rehman W, Mondal B. 2011. Detection of fowl poxvirus integrated with reticuloendotheliosis virus sequences from an outbreak in backyard chickens in India. *Veterinary Italy*, 47: 147-153.
- Black EM, Lowings JP, Smith J, Heaton PR, McElhinney LM. 2002. A rapid RT-PCR method to differentiate six established genotypes of rabies and rabies-related viruses using TaqMan1 technology. *J Virol Methods*, 105(1): 25-35.
- Bohls RL, Linares JA, Gross SL, Ferro PJ, Silvy NJ, Collisson EW. 2006. Phylogenetic analyses indicate little variation among reticuloendotheliosis viruses infecting avian species, including the endangered Attwater's prairie chicken. *Virus Res*, 119: 187-194.
- Braasch DA, Corey DR. 2001. Locked nucleic acid (LNA), fine-tuning the recognition of DNA and RNA. *Chem Biol*, 8(1): 1-7.
- Brown AC, Smith LP, Kgosana L, Baigent SJ, Nair V, Allday MJ. 2009. Homodimerization of the Meq viral oncoprotein is necessary for induction of t-cell lymphoma by Marek's disease virus. *J Virol*, 83: 11142-11151.
- Bulow VV. 1977. Immunological effects of reticuloendotheliosis virus as potential contaminant of Marek's disease Vaccines, *Avian Pathol*, 6(4): 383-393.
- Calnek BW, Witter RL. 1997. Marek's disease. *Diseases of Poultry*, 9th ed. Iowa State University Press, Iowa. pp. 356-430.

- Cardullo RA, Agrawal S, Flores C, Zamecnik PC, Wolf DE. 1988. Detection of nucleic acid hybridization by nonradiative fluorescence resonance energy transfer. *Proc Natl Acad Sci*, 85(23): 8790-8794.
- Churchill AE, Biggs PM. 1967. Agent of Marek's disease in tissue culture. *Nature*, 215: 528-530.
- Clegg RM. 1995. Fluorescence resonance energy transfer. *Curr Opin Biotechnol*, 6(1): 103-110.
- Coffin JM. 1992. Structure and Classification of Retroviruses, In J. A. Levy, (Ed), *The Retroviridae*, vol. 1 (pp. 19- 49). New York: Plenum Press.
- Cooper MD, Payne LN, Dent, PB, Burmester BR, Good RA. 1968. Pathogenesis of avian lymphoid leukosis. I. Histogenesis. *J Natl Cancer I*, 41(2):373-389.
- Costa JM, Ernault P, Olivi M, Gaillon T, Arar K. 2004. Chimeric LNA/ DNA probes as a detection system for real-time PCR. *Clin. Biochem.*, 37: 930-932.
- Couteaudier M, Denesvre C. 2014. Marek's disease virus and skin interactions. *Vet Res*, 45(1): 1-12.
- Crittenden LB, Hayward WS, Hanafusa H, Fadly AM. 1980. Induction of neoplasms by subgroup E recombinant of exogenous and endogenous avian retroviruses (Rous-associated virus type 60). *J Virol*, 33: 915-919.
- Crittenden LB, Witter RL, Fadly AM. 1979. Low incidence of lymphoid tumors in chickens continuously producing endogenous virus. *Avian Dis*, 23: 646-653.
- Cui XP, Lee LF, Reed WM, Kung HJ, Reddy SM. 2004. Marek's disease virus-encoded vIL-8 gene is involved in early cytolytic infection but dispensable for establishment of latency. *J Virol*, 78: 4753-4760.
- Cui ZZ, Lee LF, Liu JL, Kung HJ. 1991. Structural analysis and transcriptional mapping of the Marek's disease virus gene encoding pp38, an antigen associated with transformed cells. *J Virol*, 65: 6509-6515.
- Dai M, Feng M, Liu D, Cao W, Liao M. 2015. Development and application of SYBR Green i real-time PCR assay for the separate detection of subgroup J Avian leukosis virus and multiplex detection of avian leukosis virus subgroups A and B. *Virol J*, 12(1): 52.
- Davidson I. 2001. Differential diagnosis of avian oncogenic viruses. <http://www.poultrymed.com>. Accessed 6 October 2009.
- Davidson I, Silva RF. 2008. Creation of Diversity in Animal Virus World by Inter Species and Intra- Species Recombinations: lessons Learned From Poultry Viruses. *Virus Genes*, 36: 1-9.
- De Boer GF. 1987. Avian leukosis. Martinus Nijhoff Publishing, Boston, 292 pp.
- De Kok JB, Wiegerinck ET, Giesendorf BA, Swinkels DW. 2002. Rapid genotyping of single nucleotide polymorphisms using novel minor groove binding DNA oligonucleotides (MGB probes). *Hum Mutat*, 19(5): 554-559.
- Didenko VV. 2001. DNA probes using fluorescence resonance energy transfer (FRET): designs and applications. *Biotechniques*, 31: 1106-1121.
- Edwards K, Logan J, Saunders N. 2005. Real-time PCR: an essential guide. *Food Microbiol*, 22: 267-268.
- Elnifro EM, Ashshi AM, Cooper RJ, Klapper PE. 2000. Multiplex PCR: optimization and application in diagnostic virology. *Clin Microbiol Rev*, 13: 559 -570.
- Fadly AM. 2000. Isolation and identification of avian leukosis viruses: a review. *Avian Pathology*, 29(6): 529 - 535.
- Fadly AM. 2016. L53 Diagnosis and control of avian tumor viruses in poultry: a review. *The Proceedings of XXV World's Poultry Congress*, 5-9 September 2016, Beijing/China.
- Fadly AM, Garcia MC. 2006. Detection of reticuloendotheliosis virus in live virus vaccines of poultry. *Dev Biol (Basel)*, 126: 301-305.
- Fadly AM, Silva R, Hunt H, Pandiri A, Davis C. 2006. Isolation and characterization of an adventitious avian leukosis virus isolated from commercial Marek's disease vaccines. *Avian Dis*, 50: 380-385.
- Fadly AM, Smith EJ. 1999. Isolation and some characteristics of an isolate asso-ciated with myeloid leukosis in meat-type chickens in the United States. *Avian Dis*, 43: 391-400.
- Fadly AM, Witter RL. 1997. Comparative evaluation of in vitro and in vivo assays for the detection of reticuloendotheliosis virus as a contaminant in a live vaccine of poultry. *Avian Dis*, 41: 695-701.
- Fadly AM, Witter RL, Smith EJ, Silva RF, Reed WM, Hoerr FJ, Putnam MR. 1996. An outbreak of lymphomas in commercial broiler breeder chickens vaccinated with a fowlpox vaccine contaminated with reticuloendotheliosis virus. *Avian Pathol*, 25: 35-47.
- Frank F. 2001. Marek's disease: History, actual and future Perspectives. *Lohmann Inform*, No. 25, pp. 1-5.
- Frumkin D, Wasserstrom A, Itzkovitz S, Harmelin A, Rechavi G, Shapiro E. 2008. Amplification of multiple genomic loci from single cells isolated by laser micro-dissection of tissues. *BMC Biotechnol*, 8(1): 17.
- Gennart I, Coupeau D, Pejakovic S, Laurent S, Rasschaert D, Muylkens B. 2015. Marek's disease: Genetic regulation of gallid herpesvirus 2 infection and latency. *Vet J*, 205: 339-348.
- Gibson UE, Heid CA, Williams PM. 1996. A novel method for real time quantitative RT-PCR. *Genome Res*, 6(10): 995-1001.
- Gilka F, Spencer JL. 1990. Chronic myocarditis and circulatory syndrome in a White Leghorn strain induced by avian leukosis virus: light and electron microscope study. *Avian Dis*, 34: 174-184.
- Gimeno IM, Witter RL, Fadly AM, Silva RF. 2005. Novel criteria for the diagnosis of Marek's disease virus-induced lymphomas. *Avian Pathology*, 34(4): 332-340.
- Gopal S, Manoharan P, Kathaperumal K, Chidambaram B, Divya KC. 2012. Differential detection of avian oncogenic viruses in poultry layer farms and turkeys by use of multiplex PCR. *J Clin Microbiol*, 50(8): 2668-2673.
- Gordon RF. 1979. Poultry disease. Bulliere Tindall, London. 2:80.
- Hafez HM, Hess M. 1999. Modern techniques in diagnosis of poultry diseases: Review. *Arch Geflugelkd*, 63(6): 237-245.
- Haq K, Schat KA, Sharif S. 2013. Immunity to Marek's disease: Where are we now? *Dev Comp Immunol*, 41: 439-446.
- Heid CA, Stevens J, Livak KJ, Williams PM. 1996. Real time quantitative PCR. *Genome Res*, 6(10): 986-994.
- Hicks JA, Liu HC. 2013. Current State of Marek's Disease Virus MicroRNA Research. *Avian Dis*, 57: 332- 339.
- Higuchi R, Fockler C, Dollinger G, Watson R. 1993. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Nat Biotechnol*, 11(9): 1026-1030.
- Hoffmann B, Beer M, Reid SM, Mertens P, Oura, CAL, Rijn PAV, Slomka MJ, Banks J, Brown IH, Alexander DJ, King DP. 2009. A review of RT-PCR technologies used in veterinary virology and disease control: Sensitive and specific diagnosis of five livestock diseases notifiable to the World Organisation for Animal Health. *Vet Microbiol*, 139: 1-23.
- Hoffmann B, Depner K, Schirrmeyer H, Beer M. 2006. A universal heterologous internal control system for duplex real-time RT-PCR assays used in a detection system for pestiviruses. *J Virol Methods*, 136(1-2): 200-209.
- Isfort R, Jones D, Kost R, Witter R, Kung HJ. 1992. Retrovirus insertion into herpesvirus in vitro and in vivo. *Proc Natl Acad Sci*, 89(3): 991-995.

- Ishiguro T, Saitoh J, Yawata H, Yamagishi H, Iwasaki S, Mitoma Y. 1995. Homogeneous quantitative assay of hepatitis C virus RNA by polymerase chain reaction in the presence of a fluorescent intercalater. *Anal Biochem*, 229(2): 207–213.
- Islam A, Cheetham BF, Mahony TJ, Young PL, Walkden-Brown SW. 2006. Absolute quantification of Marek's disease virus and herpesvirus of turkeys in chicken lymphocyte, feather tip and dust samples using real-time PCR. *J Virol Methods*, 132: 127–134.
- Itabashi T, Maesawa C, Uchiyama M, Higuchi T, Masuda T. 2004. Quantitative detection of mutant alleles of the K-ras gene with minor groove binder-conjugated fluorogenic DNA probes. *Int J Oncol*, 24(3): 687–696.
- Jiang L, Qi X, Gao Y, Hua Y, Li K, Deng X, Wang Q, Zhang L, Chai H, Chen Y, Yin C, Gao H, Qin L, Wang Y, Qu Y, Chen Q, Fan Z, Wang X. 2013. Molecular characterization and phylogenetic analysis of the reticuloendotheliosis virus isolated from wild birds in Northeast China. *Vet Microbiol*, 166: 68–75.
- Khripin Y. 2006. High-throughput genotyping with energy transfer-labeled primers. *Methods Mol Biol*, 335: 215–240.
- Kim Y, Brown TP. 2004. Development of quantitative competitive-reverse transcriptase-polymerase chain reaction for detection and quantitation of avian leukosis virus subgroup J. *Vet Diagn Invest*, 16: 191–196.
- King T, Osmond-McLeod MJ, Duffy LL. 2018. Nanotechnology in the food sector and potential applications for the poultry industry. *Trends Food Sci Technol*, 72: 62–73.
- Kusser W. 2006. Use of self-quenched, fluorogenic LUX primers for gene expression profiling. *Methods Mol Biol*, 335: 115–133.
- Kutyavin IV, Afonina IA, Mills A, Gorn VV, Lukhtanov EA, Belousov ES, Singer MJ, Walburger DK, Lokhov SG, Gall AA, Dempcy R, Reed MW, Meyer RB, Hedgpeth J. 2000. 30-minor groove binder- DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res*, 28(2): 655–661.
- Latorra D, Arar K, Hurley JM. 2003. Design considerations and effects of LNA in PCR primers. *Mol Cell Probes*, 17(5): 253–259.
- Le Pecq J, Paoletti C. 1966. A new fluorometric method for RNA and DNA determination. *Anal Biochem*, 17: 100–107.
- Li J, Dong X, Yang C, Li Q, Cui Z, Chang S, Zhao P, Yu K, Yang H. 2015. Isolation, identification, and whole genome sequencing of reticuloendotheliosis virus from a vaccine against Marek's disease. *Poult Sci*, 94: 643–649.
- Li J, Yang C, Li Q, Li H, Xia Y, Liu D, Yu K, Yang H. 2013. Complete genome sequence of reticuloendotheliosis virus strain MD-2, isolated from a contaminated turkey herpesvirus vaccine. *Genome Announc*, 1, e00785-13.
- Linial M, Neiman PE. 1976. Infection of chicken cells by subgroup E viruses. *Virology*, 73: 508520.
- Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K. 1995. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl*, 4(6): 357–362.
- Lupiani B, Lee LF, Cui XP, Gimeno I, Anderson A, Morgan RW, Silva RF, Witter RL, Kung HJ, Reddy SM. 2004. Marek's disease virus-encoded meq gene is involved in transformation of lymphocytes but is dispensable for replication. *Proc Natl Acad Sci*, 101: 11815–11820.
- Lupiani B, Williams SM, Silva RF, Hunt HD, Fadly AM. 2003. Pathogenicity of two recombinant avian leukosis viruses. *Avian Dis*, 47: 425–432.
- Mitra N, Verma R, Amarjit S. 2012. Early Detection of Avian Oncogenic Viruses from Blood of Apparently Healthy Chickens. *Proc Natl Acad Sci*, 83(1): 53–58.
- Moelling K, Gelderblom H, Pauli G, Friis R. 1975. A comparative study of the avian reticuloendotheliosis virus: relationship to murine leukemia virus and viruses of the avian sarcoma-leukosis complex. *Virology*, 65(2): 546–57.
- Morrison TB, Weis JJ, Wittwer CT. 1998. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques*, 24(6): 954–958, 960, 962.
- Morrow C, Fehler F. 2004. Marek's disease: a worldwide problem. In *Marek's disease: an evolving problem*, Edited by Davison F, Nair V. London: Elsevier Academic Press; 49–61.
- Nair V. 2013. Latency and tumorigenesis in Marek's disease. *Avian Dis*, 57(2s1): 360–365.
- Nair V, Fadly AM. 2013. Leukosis/Sarcoma Group, In *Diseases of Poultry*, 13th edition, Swayne, David E., Glisson, John R., McDougald, Larry R., Nolan, Lisa K., Suarez, David L. And Nair, Venugopal (editors)., p. 553–592. Wiley-Blackwell, Ames.
- Nazarenko I. 2006. Homogeneous detection of nucleic acids using selfquenched polymerase chain reaction primers labeled with a single fluorophore (LUX primers). *Methods Mol Biol*, 335: 95–114.
- Nazarenko IA, Bhatnagar SK, Hohman RJ. 1997. A closed tube format for amplification and detection of DNA based on energy transfer. *Nucleic Acids Res*, 25: 2516–2521.
- Nazerian K, Solomon JJ, Witter RL, Burmester BR. 1968. Studies on the etiology of Marek's disease. II. Finding of a herpesvirus in cell culture. *Proc Soc Exp Biol Med*, 127: 177–182.
- Niewiadomska AM, Gifford RJ. 2013. The extraordinary evolutionary history of the reticuloendotheliosis viruses. *PLoS Biol*, 11(8).
- Nuovo GJ, Hohman RJ, Nardone GA, Nazarenko IA. 1999. In situ amplification using universal energy transfer-labeled primers. *J Histochem Cytochem*, 47: 273–280.
- OECD. 2014. Meat consumption, Retrieved from <http://www.oecd-ilibrary.org/content/indicator/fa290fd0-en>.
- Parcells MS, Lin SF, Dienglewicz RL, Majerciak V, Robinson DR, Chen HC, Wu Z, Dubyak GR, Brunovskis P, Hunt HD, Lee LF, Kung HJ. 2001. Marek's disease virus (MDV) encodes an interleukin-8 homolog (vIL-8): Characterization of the vIL-8 protein and a vIL-8 deletion mutant MDV. *J Virol*, 75: 5159–5173.
- Payne LN, Brown SR, Bumstead N, Howes K, Frazier JA, Thouless ME. 1991. A novel subgroup of exogenous avian leukosis virus in chickens. *J Gen Virol*, 72(4): 801–807.
- Payne LN, Howes K, Gillespie AM, Smith LM. 1992. Host range of Rous sarcoma virus pseudotype RSV(HPRS-103) in 12 avian species: support for a new avian retrovirus envelope subgroup, designated. *J Gen Virol*, 73(11): 2995–2997.
- Payne LN, Fadly AM. 1997. Leukosis/sarcoma group. *Diseases of Poultry*, 10th Ed., pp. 414–466. Iowa State University Press, Ames, Iowa.
- Payne LN, Venugopal K. 2000. Neoplastic diseases: Marek's disease, avian leukosis and reticuloendotheliosis. *Rev Sci Tech Off Int Epiz*, 19(2): 544–564.
- Persson K, Hamby K, Ugozzoli LA. 2005. Four-color multiplex reverse transcription polymerase chain reaction—overcoming its limitations. *Anal Biochem*, 344: 33–42.
- Purchase HG, Ludford C, Nazerian K, Cox HW. 1973. A new group of oncogenic viruses: Reticuloendotheliosis, chick syncytial, duck infectious anemia, and spleen necrosis viruses. *J Natl Cancer Inst*, 51(2): 489–499.
- Puro KU, Bhattacharjee U, Baruah S, Sen A, Das S, Ghatak S, Doley

- SR. Shakuntala I. 2018. Characterization of Marek's disease virus and phylogenetic analyses of meq gene from an outbreak in poultry in Meghalaya of Northeast India. *VirusDisease*, 29(2): 167-172.
- Reddy SM, Lupiani B, Gimeno IM, Silva RF, Lee LF, Witter RL. 2002. Rescue of a pathogenic marek's disease virus with overlapping cosmid dnas: Use of a pp38 mutant to validate the technology for the study of gene function. *Proc Natl Acad Sci*, 99: 7054-7059.
- Roberts L. 2000. Human genome research—SNP mappers confront reality and find it daunting. *Science*, 287(5460): 1898-1899.
- Schweiger B, Zadow I, Heckler R, Timm H, Pauli G. 2000. Application of a fluorogenic PCR assay for typing and subtyping of influenza viruses in respiratory samples. *J Clin Microbiol*, 38 (4): 1552-1558.
- Selvin PR, Hearst JE. 1994. Luminescence energy transfer using a terbium chelate: improvements on fluorescence energy transfer. *Proc Natl Acad Sci*, 91(21): 10024-10028.
- Silva RF, Fadly AM, Taylor SP. 2007. Development of a polymerase chain reaction to differentiate avian leukosis virus (ALV) subgroups: detection of an ALV contaminant in commercial Marek's disease vaccines. *Avian Dis*, 51(3): 663-667.
- Smith RE. 1982. Avian osteopetrosis. *Curr Top Microbiol Immunol*, 101: 75-94.
- Strassheim S, Stik G, Rasschaert D, Laurent S. 2012. mdv1-mir-m7-5p, located in the newly identified first intron of the latency-associated transcript of Marek's disease virus, targets the immediate-early genes ICP4 and ICP27. *J Gen Virol*, 93: 1731-1742.
- Sun GR, Zhang YP, Zhou LY, Lv HC, Zhang F, Li K, Gao YL, Qi XL, Cui HY, Wang YQ, Gao L, Pan Q, Wang XM, Liu CJ. 2017. Co-Infection with Marek's Disease Virus and Reticuloendotheliosis Virus Increases Illness Severity and Reduces Marek's Disease Vaccine Efficacy. *Viruses*, 9(6):158.
- Thelwell N, Millington S, Solinas A, Booth J, Brown T. 2000. Mode of action and application of Scorpion primers to mutation detection. *Nucleic Acids Res*, 28(19): 3752-3761.
- Tsai WP, Copeland TD, Oroszlan S. 1985. Purification and chemical and immunological characterization of avian reticuloendotheliosis virus gag-gene-encoded structural proteins. *Virology*, 140(2): 289-312.
- Tsai WP, Copeland TD, Oroszlan S. 1986. Biosynthesis and chemical and immunological characterization of avian reticuloendotheliosis virus env gene-encoded proteins. *Virology*, 155(2): 567-83.
- Tseng SY, Macool D, Elliott V, Tice G, Jackson R, Barbour M, Amorese D. 1997. An homogeneous fluorescence polymerase chain reaction assay to identify Salmonella. *Anal Biochem*, 245(2): 207- 212.
- Tyagi S, Bratu DP, Kramer FR. 1998. Multicolor molecular beacons for allele discrimination. *Nat Biotechnol*, 16(1): 49-53.
- Tyagi S, Kramer FR. 1996. Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol*, 14(3): 303-308.
- Vet JA, Van der Rijt BJ, Blom HJ. 2002. Molecular beacons: colorful analysis of nucleic acids. *Expert Rev Mol Diagn*, 2(1): 77-86.
- Walkden-Brown S, Islam A, Groves P, Rubite A, Sharpe SM, Burgess S. 2013a. Development, application, and results of routine monitoring of Marek's disease virus in broiler house dust using real-time quantitative PCR. *Avian Dis*, 57: 544-554.
- Walkden-Brown S, Islam A, Islam AF, Burgess SK, Groves PJ, Cooke J. 2013b. Pathotyping of Australian isolates of Marek's disease virus in commercial broiler chickens vaccinated with herpesvirus of turkeys (HVT) or bivalent (HVT/SB1) vaccine and association with viral load in the spleen and feather dander. *Aust Vet J*, 91: 341-350.
- Walker MH, Rup BJ, Rubin AS, Bose HR. 1983. Specificity in the immunosuppression induced by avian reticuloendotheliosis virus. *Infect Immun*, 40(1): 225-235.
- Wang LC, Huang D, Pu CE, Wang CH. 2014. Avian oncogenic virus differential diagnosis in chickens using oligonucleotide microarray. *J Virol Methods*, 210: 45-50.
- Wei K, Sun Z, Zhu S, Guo W, Sheng P, Wang Z, Zhao C, Zhao Q, Zhu R. 2012. Probable congenital transmission of reticuloendotheliosis virus caused by vaccination with contaminated vaccines. *PLoS ONE*, 7: e43422.
- Whitcombe D, Theaker J, Guy SP, Brown T, Little S. 1999. Detection of PCR products using self-probing amplicons and fluorescence. *Nat Biotechnol*, 17(8): 804-807.
- Witter RL, Fadly AM. 2003. Reticuloendotheliosis, Diseases of poultry. Ames, IA: Iowa State University Press; p. 517-535.
- Witter RL, Peterson IL, Smith EJ, Johnson DC. 1982. Serologic evidence in commercial chicken and turkey flocks of infection with reticuloendotheliosis virus. *Avian Dis*, 26(4): 753-762.
- Wittwer CT, Herrmann MG, Moss AA, Rasmussen RP. 1997a. Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques*, 22(1): 130-131.
- Wittwer CT, Ririe KM, Andrew RV, David DA, Gundry RA, Balis UJ. 1997b. The Light Cycler: a micro volume multi-sample fluorimeter with rapid temperature control, *Biotechniques*, 22 (1), 176-181.
- Wu P, Brand L. 1994. Resonance energy transfer: methods and applications. *Anal Biochem*, 218(1): 1-13.
- Yang L, Shuai C, Cui Z, Shuang C, Peng Z. 2016. Genome analysis and pathogenicity of reticuloendotheliosis virus isolated from a contaminated vaccine seed against infectious bursal disease virus, first report in China. *J Gen Virol*, 97: 2809-2815.
- Zeng T, Xie Z, Xie L, Deng X, Xie Z, Luo S, Huang L, Huang J. 2015. Simultaneous detection of eight immunosuppressive chicken viruses using a GeXP analyser-based multiplex PCR assay. *Virol J*, 12: 226.
- Zhang X, Liao M, Jiao P, Luo K, Zhang H, Ren T, Zhang G, Xu C, Xin C, Cao W. 2010. Development of a loop-mediated isothermal amplification assay for rapid detection of subgroup J avian leukosis virus. *J Clin Microbiol*, 48: 2116-2121.
- Zhang YP, Lv HC, Bao KY, Gao YL, Gao HL, le Qi X, Cui HY, Wang YQ, Li K, Gao L, Wang XM, Liu CJ. 2016. Molecular and pathogenicity characterization of gallid herpesvirus 2 newly isolated in China from 2009 to 2013. *Virus Genes*, 52: 51-60.



Derleme (Review)

Cilt 3 - Sayı 4: 173-189 / Ekim 2020
(Volume 3 - Issue 4: 173-189 / October 2020)

KÜMELEME ALGORİTMALARINDA KULLANILAN FARKLI YÖNTEMLERE GENEL BAKIŞ

Tohid YOUSEFİ^{1*}, Mehmet Serhat ODABAŞ¹, Recai OKTAŞ²

¹Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Akıllı Sistemler Mühendisliği Anabilim Dalı, 55139, Samsun, Türkiye

²Ondokuz Mayıs Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 55139, Samsun, Türkiye

Gönderi: 04 Mart 2020; **Kabul:** 21 Nisan 2020; **Yayınlanma:** 01 Ekim 2020

(Received: March 04, 2020; **Accepted:** April 21, 2020; **Published:** October 01, 2020)

Özet

Veri madenciliği, birçok teknik ve algoritmayı kullanarak büyük veri tabanlarından anlamlı bilgileri çıkarma işlemidir. Veri madenciliği genellikle, "verilerde bilgi keşfi" olarak adlandırılan ve bu bilgileri bulmak için kullanılan yöntemlerdir. Veri madenciliğinin temel yöntemlerinden birisi olan kümeleme yöntemidir. Kümeleme yöntemi günümüz dünyasında hızla çoğalan verilerin analizinde kullanılacak en güçlü yöntemlerdendir. Kümeleme bazı benzerlik mesafelerine dayalı olarak verilerdeki doğal gruplamaları veya kümeleri bulma tekniğidir. Kümeleme aslında birçok farklı veri analizlerinde temel bir adımdır. Bundan dolayı bu derlemede kümeleme algoritmalarında kullanılan farklı yöntemler özet bir şekilde anlatılmıştır.

Anahtar kelimeler: Algoritma, Kümeleme, Hiyerarşik kümeleme, Bölümsel Kümeleme, Yoğunluk tabanlı kümeleme


Overview of Different Methods Used in Clustering Algorithms


Abstract: Data mining is the process of extracting meaningful information from large databases using many techniques and algorithms. Data mining is often referred to as "information discovery in data" and many methods are used to find this information. Clustering method, which is one of the basic methods of data mining, is one of the most powerful methods to analyze these data, while data is being produced rapidly in today's world. Clustering is the technique of finding natural groupings or clusters in data based on some similarity distances. Also clustering is essentially a fundamental step in many different data analyzes. Therefore, different methods used in clustering algorithms are briefly described in this review.


Keywords: Algorithm, Clustering, Hierarchical clustering, Partitional clustering, Density-based clustering

***Corresponding author:** Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Akıllı Sistemler Mühendisliği Anabilim Dalı, 55139, Samsun, Türkiye

E mail: tohid.yousefi@hotmail.com (T. YOUSEFİ)

Tohid YOUSEFİ  <https://orcid.org/0000-0003-4288-8194>

Mehmet Serhat ODABAŞ  <https://orcid.org/0000-0002-1863-7566>

Recai OKTAŞ  <https://orcid.org/0000-0003-3282-3549>

Cite as: Yousefi T, Odabaş MS, Oktaş R. 2020. Overview of different methods used in clustering algorithms. BSJ Eng Sci, 3(4): 173-189.

1. Giriş

Bir algoritma genelde belli bir amacı hedefleyen ve o amacı gerçekleştirebilen belirgin, kusursuz, kesin ve mekanik olarak çalıştırabilen bir temel komut dizisidir. Bilgisayar programları algoritmaların somut gösterimleridir. Ancak algoritmalar program değildirler, aksine algoritmalar temel ve ilkel işlemleri destekleyen herhangi bir programlama dilinde uygulanabilen soyut mekanik prosedürlerdir (Erickson 2019; Blass ve Gurevich, 2004). Bu algoritmaların en önemlilerinden olan kümeleme algoritmaları bu derlemede özet biçimde anlatılmıştır.

Kümeleme, etiketlenmemiş bir veri setinin üzerinde benzer nesnelerin gruplara ayırma işlemine denilir. Küme adı verilen her bir grup kendi nesnelere arasında benzerlik sahibiyken diğer grupların nesnelere ile hiçbir benzerlik sahibi değildir (Hartigan, 1996). Kümeleme analizi, keşfedici veri analizi olarak bilinen ve hızla büyüyen bu alanda önemli bir teknik olarak bilinmektedir. Kümeleme biyoloji, tarım, psikoloji, tıp, pazarlama, bilgisayarlı görü ve uzaktan algılama gibi çeşitli mühendislik ve bilimsel alanlarda kullanılmaktadır.

Kümeleme analizi, birçok istatistiksel yöntemde ortak olan varsayımlara ihtiyaç duymaz. Daha doğrusu verilerin yapısını araştırmak için kullanılan bir araçtır. Bundan dolayı buna örüntü tanımada ve yapay zeka literatüründe denetimsiz öğrenme denilmektedir (Jain ve Dubes, 1988).

Bu derlemede algoritma ve kümeleme analizinden bahsedilmiş ve daha sonra hiyerarşik kümeleme yöntemleri, bölümsel kümeleme yöntemleri, yoğunluk tabanlı kümeleme yöntemleri, model tabanlı kümeleme yöntemleri ve ızgara tabanlı kümeleme yöntemleri anlatılmıştır.

2. Algoritma

2.1. Algoritma Kelimesinin Tarihçesi

Algoritma kelimesinin kökeni ünlü İranlı matematikçi, gökbilimci ve coğrafyacı, Abu Jafar Muhammed bin Musa Khwarizmi (al-Khwarizmi) isminden ortaya çıkmıştır. Kharazm aslında bugün Özbekistan ülkesinde bulunan ve Khiyaw olarak adlandırılan "Büyük İran" şehirlerinden birisiydi (Berggren, 2017).

Khwarizmi tarafından yazılan kitap 12. yüzyılda Latince'ye "Algoritmi de numero Indorum" yani "hintli sayılardaki algoritmik" olarak çevrildi. 13. yüzyılda, algoritma kelimesi, hala algoritmanın anlamlarından biri olan "Arapça ondalık sistem" (yani 0'den 9'a kadar artı ondalık kavram) anlamını taşımaktaydı. Bu kelime 19. yüzyılda Fransızca'da "algorithme" kelimesine dönüştürülmesine rağmen anlamı değişmedi ve İngilizceye "algorithm" diye girmesi çok uzun bir zaman almadı (Knuth, 1980).

2.2. Algoritmanın Günümüzde Anlamı ve Kullanımı

Algoritma, bazı görevleri belirli bir sürede gerçekleştirmek için adım adım uygulanan bir

prosedürdür (Goodrich ve Tamassia, 2014) ve bu yüzden algoritmalar bilgisayar biliminde kendine özgü bir rol oynamaktadır. Algoritma öğrenmenin en iyi bilinen ve rahat anlaşılabilen yollardan biri görselleştirmeyi kullanmaktadır. Bu görselleştirmelerin en basit yollarından biri olan akış diagramı algoritmanın kolay anlaşılmasına, takip ve kontrol edilmesine yardımcı olmaktadır (Müldner, 2003).

Algoritma çevresiyle etkileşime girerken girdi ya da girdileri kabul eder ve çıktı yada çıktıları üretir (Goldschlager ve Lister, 1988). Ancak (Knuth, 2014) belirttiğine göre modern anlamda "algoritma" kelimesinin biraz farklı bir şey ifade etmesinin yanı sıra tarif, süreç, yöntem, teknik, prosedür, rutin ve tekerlemeye oldukça benzer bir anlamı olmalıdır. Bir algoritmanın belirli bir problem türünü çözmek için bir işlem sırası veren bir dizi kural olmasının dışında, beş önemli özelliği daha vardır. Bunlar sonu olma, kesinlik, girdi, çıktı ve etkileycilik.

Günümüz dünyasında, algoritmaların optimal çözümler için sağlayabileceği birçok konu vardır. Örneğin Tıp, Genetik, E-ticaret, tarım, sanayi, rota bulma vb.

Bunlara benzer konularda algoritmaların yardımıyla zaman ve maliyetten tasarruf sağlayarak yeni ve optimum çözümler geliştirilebilir (Cormen ve ark., 2009).

3. Kümeleme

3.1. Küme, Kümeleme ve Küme Analizi Nedir?

"Küme" kavramını kesin olarak tanımlamamız mümkün değildir. Bunun sebeplerinden biri birçok kümeleme algoritmasının olmasıdır. Ancak bunların hepsinin ortak noktası bir dizi veri nesnelere olmasıdır (Estivill-Castro, 2002). Genel olarak kümeyi tanımlamak istersek toplanmış veya gruplanmış benzer nesnelere oluşan bir grup olarak söyleyebiliriz (Everitt, 1979) ya da bir başka deyişle bir küme benzer nesnelere dizisi olarak da ifade edilebilir (Hartigan, 1996).

"Kümeleme" kelimesi, sınıflandırma ile neredeyse eş anlamlıdır (Hartigan, 1996) ancak aralarında belirleyici bir fark vardır (Sathya ve Abraham, 2013). Sınıflandırma denetimli öğrenme (bu öğrenme yönteminde (Polycarpou ve ark., 1995) dediğine göre sonucun ne olması için önceden bilgiye sahip olmayı gerektirir aynı zamanda bu yöntemde bir dizi girdiyi yanıtlamasını düzelten bir eğitimciye ihtiyaç vardır (Cord ve Cunningham, 2008). Kümeleme ise denetimsiz öğrenme yönteminde (Hinton ve ark., 1999) belirttiği üzere önceden var olan etiketsiz verilerimizden bilinmeyen örüntüleri ortaya çıkarmakta yardımcı olur. Bu yüzden bu yöntemde eğitmen yoktur ve kümeleme yaklaşımına dayanır (Cord ve Cunningham, 2008).

Kümeleme veri nesnelere ayrı ayrı kümelere ayırma işlemidir. Böylece aynı kümedeki veriler benzer iken, farklı kümelerdeki veriler birbirine benzemezler (Sun ve Qin, 2007). Ayrıca kümeleme, ham verileri makul bir şekilde sınıflandırmanın ve veri kümelerinde bulunabilecek gizli örüntülerin arama yoludur (Huang ve

Discovery, 1998). “Küme Analizi” veride grup bulma sanatıdır (Kaufman ve Rousseeuw, 2009). Örneğin bir dereden bir sürü çakıl taşı toplanıp bunlar boyut, şekil ve renk olarak ayrılması işlemine küme analizi denilir ve ayrılan her bir benzer çakıl taşı yığına ise küme denir (Romesburg, 2004). Bundan dolayı Küme analizinin amacı nesnelere doğal ayrılmalarının keşfetmesidir (Issaq ve Veenstra, 2019).

3.2. Kümeleme Hangi Alanlarda Kullanılır?

Günümüz dünyasında verilerin artırılmasıyla ve verilerden değerli bilgileri öğrenilmesi isteğiyle kümeleme teknikleri birçok alanda hızla kullanılmaya ve uygulanmaya başlamıştır. Örneğin yapay zeka, biyoloji, müşteri ilişkileri yönetimi, veri sıkıştırma, veri madenciliği, bilgi alma, görüntü işleme, makine öğrenmesi, pazarlama, tıp, örüntü tanıma, psikoloji, istatistik ve bunun gibi benzer birçok alanda kümeleme algoritmaları sıkça kullanılmaya başlamıştır (Na ve ark., 2010).

3.3. Kümeler Arasındaki Benzerlik Ölçme

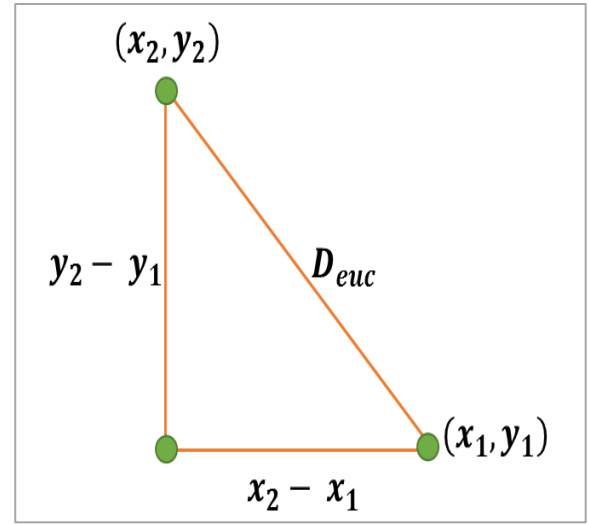
Kümelemenin en önemli adımlarından birisi, iki ögenin nasıl benzerliğini hesaplanacağını belirleyen bir uzaklık ölçüsü seçmektir (Vimal ve ark., 2008). Veriler arasındaki benzerliği ölçmek için mesafe ölçüsü çok önemli bir rol oynamaktadır. Dolayısıyla veri nesnelere arasındaki uzaklığı ve bu nesnelere uzaklık ölçüsü açısından benzer veya farklı olup olmadığını bulmak için kümeleme yöntemleri kullanılır (Wu ve ark., 2008). Dahası değişkenleri kümelemek ve değişkenler arasındaki ilişkileri bulmak için bazı sayısal ölçümleri yapılmalıdır. Dolayısıyla bu gereksinimlere yönelik geleneksel olarak yaklaşım tarzı, değişkenlerin her çift kombinasyonu için bir ilişki ölçüsü hesaplamaktır (Anderberg, 2014).

Bugüne kadar birçok kümeleme algoritması geliştirilmiştir. Bunların birçoğunun temel fikri, kümenin en iyi özelliklerini temsil edebilen küme merkezini bulmaktır ve bu kümelerin merkezine olan uzaklıklarına göre bulduğumuz herhangi bir kümeye veri noktalarını atamaktır (Belli ve ark., 2012). Veri kümelemesi için literatürde birçok uzaklık ölçümü önerilmiştir. Bu bölümde literatürde en çok kullanılan Öklid Uzaklığı, Manhattan Uzaklığı ve Minkowski Uzaklığını özet bir şekilde anlatılmıştır.

3.3.1. Öklid uzaklığı

Öklid uzaklığının hesaplanması (Şekil 1) makine öğrenmenin ve özellikle kümeleme yöntemlerinin en önemli kısımlardan birisidir (Mesquita ve ark., 2017).

Öklid uzaklığı ayrıca L_2 uzaklığı olarak da adlandırılmaktadır (Malkauthekar, 2013). Eğer $u = (x_1, y_1)$ ve $v = (x_2, y_2)$ iki nokta mevcut ise, o zaman u ve v arasındaki Öklid uzaklığı Eşitlik (1)'de gösterildiği gibi hesaplanmaktadır.



Şekil 1. Öklid uzaklığı.

$$EU(u, v) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

Eğer ikiden fazla boyutumuz var ise o zaman Öklid uzaklığının hesaplanması da Eşitlik (2) gibi olacaktır:

$$EU(u, v) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2} \quad (2)$$

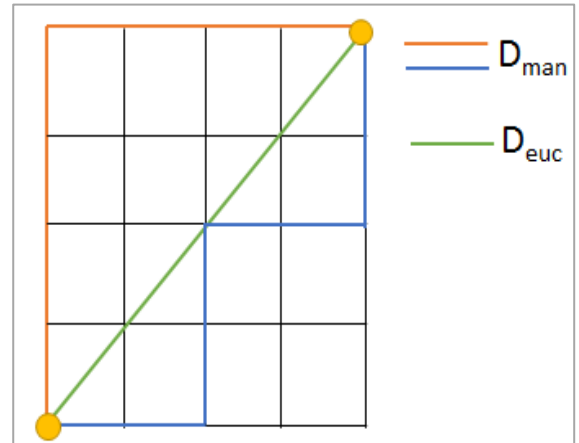
Sonuç olarak Öklid uzaklığı iki farklı şekilde literatürde gösterilmektedir:

$$Deuc = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

$$Deuc = (\sum_{i=1}^n (x_i - y_i)^2)^{\frac{1}{2}} \quad (4)$$

3.3.2. Manhattan uzaklığı

İki vektör arasındaki Manhattan uzaklığı, vektörler arasındaki mesafenin bir normuna eşittir. Manhattan uzaklığı literatürde ayrıca “metric”, “taxicab” ve “city block” olarak adlandırılmaktadır (Szabo, 2015). Manhattan uzaklığı (Xu ve Wunsch, 2005) dediklerine göre, Minkowski uzaklığının (m=1) özel versiyonu gibidir (Şekil 2).



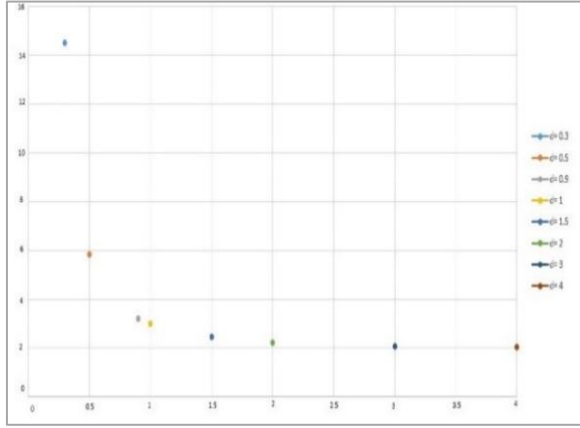
Şekil 2. Manhattan uzaklığı.

Bu uzaklık ölçüsü aykırılıklara karşı hassastır ve kümeleme algoritmalarında bu mesafe kullanıldığında, kümeler dikdörtgen şekli alırlar. Manhattan uzaklığının ölçümü Eşitlik (5)'de verildiği gibi hesaplanmaktadır.

$$D_{man} = \sum_{i=1}^n |x_i - y_i| \quad (5)$$

3.3.3. Minkowski uzaklığı

Minkowski uzaklığı ana uzaklık ölçülerinden birisidir (Şekil 3). Bunun sebebi Hamming Uzaklığı ve Öklid Uzaklığı gibi birçok mesafe ölçülerini genelleştirir (Merigo ve Casanovas, 2011).



Şekil 3. Minkowski uzaklığını d değeri ile karşılaştırma.

Minkowski uzaklığı, Öklid ve Manhattan mesafelerinin bir genellemesi olarak da tanımlanmaktadır (Han ve ark., 2012). Aynı zamanda bu mesafe ölçümü p -norm uzaklığı adıyla da literatürde bilinmektedir (Lu ve ark., 2016). Eğer x ve y iki nokta ve p boyutunda ise bu iki noktanın Minkowski uzaklık ölçümü Eşitlik (6) gibi hesaplanmaktadır (Minkowski değişkeni burada " d " olarak bilinir).

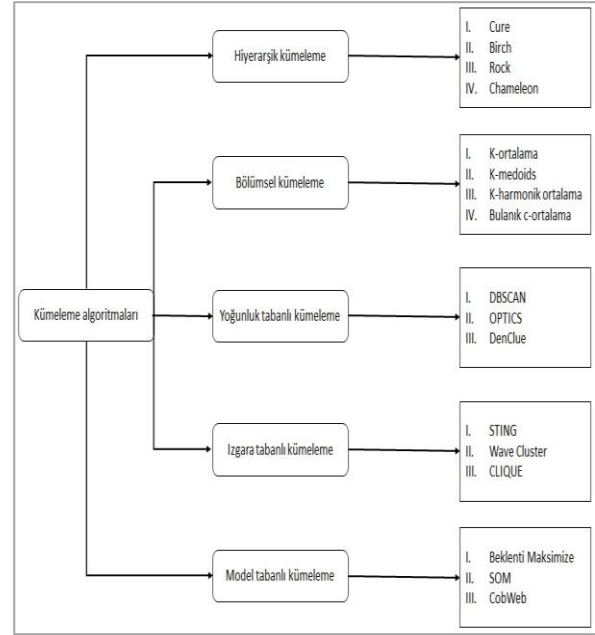
$$D_{mink} = (\sum_{i=1}^p |x_i - y_i|^d)^{\frac{1}{d}} \quad (6)$$

Eğer $d=2$ ise o zaman Minkowski uzaklığı Öklid uzaklığına, eğer $d=1$ ise o zaman Minkowski uzaklığı Manhattan uzaklığına ve son olarak eğer $d \rightarrow \infty$ ise o zaman Minkowski uzaklığı, Chebyshev uzaklığına dönüştürülür (Metcalf ve Casey, 2016).

3.4. Kümeleme Yöntemleri

Daha önceki bölümlerde belirtildiği üzere ve (Estivill-Castro, 2002) dediğine göre birçok kümeleme yöntemine sahip olmamızın temel nedenlerinden birisi "küme" kavramını hala tanımlanmamış olmasıdır. Bundan dolayı birçok kümeleme algoritmasında geliştirilmiştir. Veri noktalarının kümeleme yöntemleri, ölçek alanı kullanarak yapılır ve bu birçok yazar tarafından önerilmiştir (Chakravarthy ve Ghosh, 1996; Kothari ve Pitts, 1999; Leung ve ark., 2000; Nakamura ve Kehtarnavaz, 1998; Roberts, 1997). Ayrıca birçok kümeleme algoritması hiyerarşik kümeleme, bölümsel

kümeleme, yoğunluk tabanlı kümeleme, ızgara tabanlı kümeleme ve model tabanlı kümeleme yöntemlerine ayrılır (Şekil 4).



Şekil 4. Bu çalışmada incelenen kümeleme yöntemlerinin sınıflandırılması.

3.4.1. Hiyerarşik kümeleme yöntemleri

Kümeleme analiz yöntemlerinden birisi hiyerarşik kümeleme yöntemidir (Joshi ve ark., 2013). Bu kısımdaki algoritmalar sezgisel bölme veya birleştirme tekniklerini kullanarak küme ağacı ya da başka bir deyişle dendrogram oluştururlar (Hamerly ve Elkan, 2003). Bir dendrogram bir grup varlık arasındaki benzerlik ilişkilerini temsil eden bir dallanma şemasıdır (Murtagh, 1984). Bir küme ağacı ise, her kümelemede veri kümesinin bir bölümü olan kümeleme dizisini gösteren bir ağaçtır (Leung ve ark., 2000).

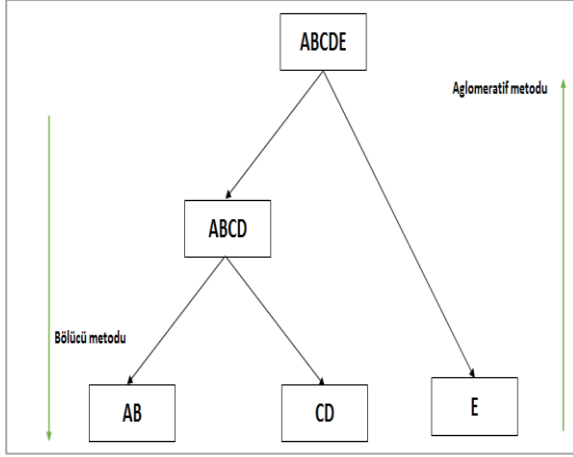
Küme ağacını oluşturmak için bölme yöntemini kullanan algoritmalara bölücü denilmektedir. Aynı zamanda kümeleme ağacını oluşturmak için birleştirme yöntemini kullanan algoritmalara ise aglomeratif denilmektedir (Turi, 2001). Hiyerarşik kümeleme, hiyerarşik ayrışmanın aşağıdan yukarıya (birleşme - aglomeratif) veya yukarıdan aşağıya (bölme - bölünme) bir yaklaşımla oluşturulmasına bağlı olarak Şekil 5'deki gibi ikiye ayrılmaktadır (Rao ve Gudivada, 2018).

a) Aglomeratif hiyerarşik kümeleme

Aglomeratif hiyerarşik kümeleme, yapraklarda depolanan veri elemanlarından başlayarak ikili bir birleştirme ağacının oluşturulmasıyla ve en yakın alt kümelerin köküne ulaşana kadar, ikiye ikiye birleştirilerek devam eder. X 'in tüm öğelerini içeren ağaç $\Delta(X_i, X_j)$ ile X 'in herhangi bir iki alt kümesi arasındaki mesafeye, bağlantı mesafesi olarak adlandırılmaktadır. Bu tekniğe ayrıca aglomeratif hiyerarşik kümeleme denilmektedir (Nielsen, 2016). Aglomeratif Hiyerarşik Kümeleme, tek gözlem kümeleriyle başlar ve giderek

daha fazla gözlem içerir. Aynı zamanda daha az sayıda küme oluşturur ve çift olan kümeleri bir biriyle birleştirir (Myatt ve Johnson, 2009). Aglomeratif kümelemenin oluşturma adımları (Rao ve Gudivada, 2018) dediklerine göre aşağıdaki gibi ilerlemektedir;

1. Her veri noktası tekil bir küme olarak farz edilir.
2. Öklid mesafesi hesaplanın her yinelemesinden sonra, minimum mesafesi olan iki kümeyi birleştirir.
3. Tüm örneklerin tek bir küme haline geldiğinde kümeleme işlemi durdurulur. Aksi takdirde 2. adıma geri dönülür.



Şekil 5. Aglomeratif ve Bölücü kümelemenin örnek akış şekli.

En popüler iki aglomeratif hiyerarşik algoritma, tek bağlantı (Sneath ve Sokal, 1973) ve tam bağlantı (Anderberg, 2014) algoritmalarıdır. Tek bağlantılı algoritmaları, iki kümenin en yakın iki noktası arasındaki mesafeyi kesişim mesafesi olarak kullanır (Turi, 2001).

$$dist(C_i, C_j) = \min\{dist(o_i, o_j) | o_i \in C_i, o_j \in C_j\} \quad (7)$$

öte yandan tam bağlantılı algoritmaları, kesişim mesafesi olarak en uzak iki noktanın mesafesini kullanır (Turi, 2001).

$$dist(C_i, C_j) = \max\{dist(o_i, o_j) | o_i \in C_i, o_j \in C_j\} \quad (8)$$

Genel olarak tek bağlantılı algoritmaları özet kümeler oluştururken, tam bağlantılı algoritmalar uzun kümeler oluşturur ve bundan dolayı tam bağlantılı algoritmaları, tek bağlantılı algoritmalarından daha kullanışlıdır (Jain ve ark., 1999).

b) Bölücü hiyerarşik kümeleme

Bölücü hiyerarşik kümeleme, aynı kümedeki tüm nesnelere ardından sürekli tekrarlanmanın başlamasıyla başlar. Bir küme, özelliklerine bağlı olarak daha küçük kümelere ayrılır. Bu işlem her bir nesne bir kümeyle dahil olana kadar veya sonlandırma şartı sağlanana kadar devam eder (Reddy ve ark., 2017). Bölücü kümelemenin oluşturma adımları (Rao ve Gudivada, 2018) dediklerine göre aşağıdaki gibi ilerlemektedir;

1. Kümedeki tüm veri noktalarıyla başlar.

2. Her yinelemeden sonra, en az uyumlu olan noktaları kümeden çıkarılır.
3. Her örnek kendi tekli kümesinde olduğunda durdurulur. Aksi takdirde 2. Adıma geri dönülür.

Bölücü hiyerarşik kümelemenin başka bir adıma literatürde DIANA (Divisive Analysis – Bölücü Analizi) olarak geçmektedir. Bu algoritma ilk olarak veri kümesindeki tüm veri noktalarını gözlemlenmesiyle başlar ve daha sonra istenen sayıda küme elde edilene kadar kademeli olarak böler. Özet olarak bu yöntemin iki temel aşaması vardır. Birincisi bölünecek uygun bir küme bulmaktır ve ikincisi ise seçilmiş kümenin iki yeni kümeyle nasıl bölüneceğinin belirlemektir (Thilagavathi ve ark., 2013).

Hiyerarşik kümelemenin avantajları (Berkhin, 2006) sırasıyla; verilerin yoğunluğu konusunda esnek davranabilmesi, tüm uzaklık ve mesafe biçimlerini değerlendirebilmesi ve tüm ölçek türleri ile uygulanabilmesidir.

Hiyerarşik kümelemenin dezavantajları (Berkhin 2006) ise değişken sayısı arttıkça yapılacak işlem miktarının artacağından dolayı büyük veri setlerinde zaman alıcı bir yöntemdir. Bu yöntemde daha önce yapılan herhangi bir işlemi geri alamadığımızdan dolayı gürültülü veriye oldukça duyarlıdır.

Hiyerarşik kümelemenin dezavantajlarını azaltmak ve kalitesini arttırmak için umut verici bir taraf, hiyerarşik kümelemeyi diğer kümeleme teknikleri ile kolayca entegre edilmesidir. Bu kısımda popüler olan geliştirilmiş hiyerarşik kümelemelerden Cure (Guha ve ark., 1998), Birch (Zhang ve ark., 1996), Chameleon (Karypis ve ark., 1999) ve Rock (Guha ve ark., 2000) anlatılmıştır.

Cure

Cure küme merkezleri ve tüm veri noktaları arasında bir denge oluşturan aglomeratif hiyerarşik kümeleme algoritmasıdır. Ayrıca Cure temel olarak veri kümesinin bölünmesini kullanan bir algoritmadır. Bu algoritma büyük veritabanlarının işlenmesi için rastgele örnekleme ve bölümlenme kullanır. Bu yapılan işlemde veri kümesinde rastgele seçilen bir örnek önce bölümlere ayrılır ve sonra her bölüm kısmi kümelendirir. Daha sonra bu kısmi kümeler tekrar istenilen kümeleri elde etmek için ikinci aşamada kümelendirir (Guha ve ark., 1998).

Cure algoritmasında her küme bir kümenin kenarlığını temsil eder. Temsili bir nokta içerdiği bu algoritma veri kümeleri bölümler. İlk önce temsili noktalar dağınık nesnelere seçilir ve daha sonra seçilen bu noktalar bütünlük faktörü olarak tanıyan bir faktör tarafından kümenin merkezine taraf hareket ettirilir. Bu algoritmada Kümenin merkezine doğru kullanılan ilerleme yöntemi, küme şekillerinin düzensizliğini azaltır. Bundan dolayı bu algoritma küresel olmayan şekillerde ve çeşitli büyük boyutlarda olan kümelere daha verimlidir. Dolayısıyla bu algoritma büyük veritabanları için kullanışlıdır (Kaur, 2015) ve aykırı değerlerin ele alınması içinde iyi bir yöntemdir (Dunham, 2006).

Birch

Birch hiyerarşik kümeleme ve yinelemeli bölümlenme gibi diğer kümeleme yöntemlerinin birleştirilmesi ile çok miktarda sayısal verilerin kümelmesi için tasarlanmıştır. Çok adımlı kümeleme yapmak için hiyerarşik bir veri yapısı olan kümeleme özellik ağacı oluşturulur (Zhang ve ark., 1996). Kümeleme özellik ağacı, nesnelerin kümeleri hakkındaki bilgileri özetleyen üç boyutlu bir vektöre deniliyor. Bir kümeleme özellik ağacı, yüksek dengeli bir ağaçtır ve hiyerarşik kümelemeyi gerçekleştirmek için kümeleme özelliklerini depolamaktadır (Reddy ve Ussenaiah, 2012). Ayrıca Birch ana bellek tabanlı bir kümeleme algoritmasıdır. Dolayısıyla kümeleme işlemi bir bellek kısıtlamasıyla gerçekleştirilir. Bu algoritma aslında 2 aşamadan oluşmaktadır (Han ve ark., 2011).

1. Kümeleme özellik ağacının başlangıç belleğini oluşturmak için veri tabanı taranır.
2. Kümeleme özellik ağacının yaprak düğümlerini kümeleme için rastgele bir kümeleme algoritması kullanılır.

Birch kümeleme algoritması ölçeklenebilir bir algoritmadır. Dolayısıyla sadece tek bir tarma ile yeni kümeler bulabilir ve eğer tarama sayısı arttırılırsa küme kalitesi de artacaktır. Ayrıca bu algoritma önceki adıma geri dönme özelliğine sahiptir. Ancak bu kümeleme yöntemi sayısal olmayan verileri işleyemez (Xu ve Wunsch, 2005).

Chameleon

Chameleon kümeleri birleştirmek amacıyla dinamik modelleme kullanan aglomeratif hiyerarşik kümeleme yöntemidir. Bu teknikte kullanıcının bilgi sağlamasına gerek yoktur. Ancak kümeleri kendilerine özgü özelliklerine göre birleştirilmiş bir şekilde ayarlar (Berkhin 2006). Bu yöntem kümelerin aralarında olan benzerliğini bulmak için kümelerin birbirine olan bağlılığını ve yoğunluğunu kullanır (Shah ve Nair, 2015). Chameleon'da küme benzerliği, nesnelerin bir küme içinde birbiriyle ne kadar iyi bağlı olduğuna ve kümelerin yakınlığına bağlıdır. Dolayısıyla aralarındaki bağlantı yüksekse ve birbirlerine yakınlarsa iki küme birleştirilir. Bu nedenle Chameleon statistiğe bağlı değildir ve birleştirilen kümelerin iç özelliklerine otomatik olarak uyum sağlar. Dolayısıyla birleştirme işlemi doğal ve homojen kümelerin bulmasını kolaylaştırır (Han ve ark., 2011).

Chameleon algoritmasının temel özelliklerinden birisi kümelerin aralarındaki benzerliği bulmak için kümelerin birbirine olan bağlılığını ve yoğunluğunu kullanmasıdır. Bundan dolayı bu algoritma rastgele şekiller ve değişen yoğunluk kümeler için uygun bir yöntemdir (Kaur, 2015). Ayrıca bu algoritma seyrek bir grafik oluşturmak için k-en yakın komşu grafik yaklaşımını kullanır (Cherng ve Lo, 2001).

Rock

Rock algoritması 2000 yılında Guha, Rastogi ve Shim tarafından önerildi. Rock algoritmasına aslında bağlantıları kullanan güçlü bir algoritma denilmektedir.

Bu algoritma bir hiyerarşik kümeleme algoritmasıdır. Aynı zamanda kategorik ve boole verileri ile çalışmaktadır. Rock en yakın komşu, yer değiştirme ve hiyerarşik aglomeratif kümeleme yöntemlerini birleştirmektedir (Guha ve ark., 2000). Bu algoritmada rastgele bir örnekleme yöntemi uygulanır, daha sonra kümeler arasındaki mesafeyi bulmak için hiyerarşik kümeleme algoritması kullanılır ve son olarak ta tüm veriler etiketlenir (Kaur, 2015).

Rock algoritmasının kategorik verileri kümelemek için en kullanışlı algoritma olmasının nedeni, iki veri noktası arasındaki benzerliği bulmak için Jaccard veya kosinüs benzerlik katsayılarını kullanabiliyor olması ve ayrıca komşuları belirlemek için bağlantı yöntemini kullanabilmesidir (Tyagi ve ark., 2012). Bu algoritma büyük, gerçek ve sentetik veri setlerini çalıştırabilir. Ayrıca kategorik bir veri setini etkin bir şekilde işleyebilir. Ancak bu algoritma artımlı veri kümelerini desteklemez (Dabhi ve ark., 2016).

3.4.2. Hiyerarşik olmayan(bölümsel) kümeleme yöntemleri

Bölümlenme yöntemi bir kümeleme analizi yöntemidir. Bu yöntem bir dizi n nesnesinin verildiği ve veri kümesinin bir nesnel bölümlenme ölçütünü en aza indirmek için $k \leq n$ şartını sağlayarak ve her kümesinde benzer nesneler içerdiği bir dizi k kümesine bölündüğü bir tekniktir (Şekil 4). Bu yöntemle elde edilen k kümeleri iki şartı sağlaması gerekiyor: Birincisi Her kümede en az bir nesne bulunmalıdır ve ikincisi ise Her nesne kesinlikle bir kümeyle ait olmalıdır (Rezende ve Esmine, 2010).

Bölümsel kümeleme yöntemleri, verilerin tek bir bölümünü yinelemeli olarak oluşturur; buradaki objektif fonksiyonu, n-boyutlu uzayda piksel vektöründen küme prototipine olan mesafelerin toplamı ile tanımlanır (Frigui ve Krishnapuram, 1997). Bu yöntem aynı zamanda hiyerarşik olmayan kümeleme yöntemi de denilir, çünkü yalnızca bir küme kümesi tipik bir kısmı kümeleme algoritması oluşturur (Han ve ark., 2012).

Bölümlenme kümeleme algoritmaları, veri kümesini daha önceden belirtilen küme sayısına böler. Bu algoritmalar belirli kriterleri (örnek olarak Kare hata fonksiyonu) en aza indirmeye çalışır ve bu nedenle optimizasyon problemleri olarak değerlendirilebilir (Leung ve ark., 2000). Bölümlenme kümeleme algoritmaları genel olarak yerel optimizasyona dönüşen yinelemeli algoritmalar. Yinelemeli kümeleme algoritmasının adımları (Hamerly ve Elkan, 2002) dediğine göre şunlardır;

1. K kümesinin merkezlerini rastgele seç
2. Her bir x_i veri noktası için $m(c_j | x_i)$ üyeliğini ve her bir merkez c_j ve ağırlığını $w(x_i)$ larını hesaplayın
3. Her bir c_j merkez için, konumunu tüm veri noktalarından yeniden hesaplayın ve k küme merkezlerini yeniden hesaplamak için aşağıdaki formülü kullanın.

$$c_j = \frac{\sum_{i=1}^n m(c_j | x_i) w(x_i) x_i}{\sum_{i=1}^n m(c_j | x_i) w(x_i)} \quad (9)$$

İkinci ve üçüncü aşamayı durdurma kriteri karşılanıncaya kadar devam ettirin.

Yukarıdaki algoritmada $m(c_j | x_i)$, x_i örneğinin k kümesine üye olup olmadığını ölçen fonksiyondur. Üyelik işlevi $m(c_j | x_i)$ aşağıdaki kısıtlamaları karşılamalıdır:

$$m(c_j | x_i) \geq 0, i = 1, \dots, n \text{ ve } j = 1, \dots, n$$

$$\sum_{i=1}^n m(c_j | x_i) = 1$$

Ağırlık fonksiyonu $w(x_i)$, Eşitlik (9)'da bir sonraki yinelemede merkezlerin yeniden hesaplamasında x_i örneğinin ne kadar etkisi olduğunu tanımlar ($w(x_i) > 0$ olduğunda) (Hamerly ve Elkan, 2002).

Bölümsel kümeleme algoritmalarının en popüler olanlarından k -ortalama algoritması (Forgy, 1965), k -medoids algoritması (Kaufmann, 1987), k -harmonik ortalama algoritması (Zhang ve ark., 1999) ve bulanık c -ortalama algoritması (Ruspini ve Control, 1969) aşağıdaki kısımda anlatılmıştır.

K-ortalama algoritması

Son yıllarda dağıtılmış ve paralel hesaplamaya dayanan birçok kümeleme algoritması önerilmiştir. İlk olarak Mac Queen 1967'de bu algoritmayı önerdi; ancak standart algoritma 1957'de Stuart Lloyd tarafından önerilmiştir. Bazen K -Means algoritması Lloyd-Forgy olarak ta anılır; çünkü 1965'te Forgy aynı yöntemi yayınlamıştır (Singh ve Security, 2015).

En yaygın kullanılan bölümsel kümeleme algoritmalarından biri olan k -ortalama algoritmasının optimize ettiği objektif fonksiyonu aşağıda verilmiştir (Forgy, 1965).

$$J_{k\text{-means}} = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (10)$$

Böylece k -ortalama algoritması küme içinde mesafeyi en aza indirmeye çalışır (Hamerly ve Elkan, 2002). K -ortalama algoritması k merkezleri ile çalışmaya başlar (merkezler için başlangıç değerler önce rastgele seçilir ya da önceden verilmiş bilgilere dayanarak türetilir). Akabinde veri kümesindeki her bir numune en yakın kümeye atanır. Son olarak küme merkezleri verilerin mesafesine göre tekrar hesaplanır ve bu işlem küme merkezlerinin değişmediği sürece kadar tekrarlanır.

k -ortalama algoritması için üyelik ve ağırlık fonksiyonları aşağıdaki şekilde tanımlanır (Hamerly ve Elkan, 2002).

$$m(c_j | x_i) = \begin{cases} 1 & \text{if } \|x_i^{(j)} - c_j\|^2 = \arg \min_k \{\|x_i^{(j)} - c_j\|^2\} \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

$$w(x_i) = 1 \quad (12)$$

k -ortalama algoritmasının performansı, çeşitli mesafe ölçütleri ile birçok veri setinde (örneğin Iris, Wine, Vowel vs) ölçülmüştür ve bu algoritmanın performansı,

kullanılan veri tabanına ve mesafe ölçütlerine bağlı olduğu sonucuna varılmıştır (Thakare ve Bagal, 2015).

K -ortalama algoritmasının aşamaları (Hartigan ve Wong, 1979) dediklerine göre aşağıdaki gibi ilerlemektedir;

1. K küme sayısı belirlenirken, her kümenin ağırlıklı ortalaması olan küme merkezi hesaplanır.
2. Her nesnenin seçilen küme merkezine olan uzaklığı hesaplanır ve bu uzaklıklara göre her nesne yakın olduğu kümeye dahil edilir.
3. Yeni küme merkezlerini hesaplamak için artan ya da azalan nesnelere göre ağırlıklı ortalamaları hesaplanır.
4. Küme merkezleri değişmeyene kadar bu işlem tekrarlanır.

k -ortalama algoritmasının avantajları (Turi, 2001);

1. k -ortalama algoritmasının uygulaması çok kolaydır.
2. Zaman karmaşıklığı $O(Np)$ olduğu için büyük veri kümeleri için bile uygundur.

k -ortalama algoritmasının dezavantajları (Davies, 2004);

1. Algoritma verilere bağlıdır.
2. Başlangıç koşullarına bağlı bir algoritmadır.
3. Kullanıcı küme sayısını önceden belirlemesi gerekiyor.

k -ortalama algoritmasının en büyük dezavantajı, farklı kümeleri üreten başlangıçtaki k merkezlerin seçilmesidir. Ancak algoritmanın son küme kalitesi, başlangıçtaki merkezlerin seçimine bağlıdır ve bu orijinal k değerini bulmamıza bir sınır getirir (Nazeer ve Sebastian, 2009) ayrıca k -ortalama algoritması her bir veri nesnesi ile her bir yinelemedeki tüm küme merkezleri arasındaki mesafeyi hesaplaması gerekmektedir. Bu tekrarlı süreç k -ortalama algoritmasının verimliliğini etkilemektedir (Na ve ark., 2010).

K-medoids algoritması

K -ortalama algoritmasında verilerin aralarındaki değer aşırı derecede farklıysa, elde edilen kümeleme sonuçlarında düzensiz değişiklikler olacaktır. Bu nedenle k -ortalama algoritması gürültülü verilere ve aykırı değerlere çok duyarlıdır. Dolayısıyla algoritmanın performansını kötü etkiler (Swarndeeep Saket ve Pandya, 2016). K -ortalama algoritmasının bu şekil dezavantajlarını ortadan kaldırmak için k -ortalama algoritmasına benzeyen k -medoids algoritmasını kullanılması gerekmektedir (Bhat ve Control, 2014; Han ve ark., 2012; Hartigan ve Wong, 1979).

Medoids etrafında bölünme olarak da adlandırılan bu algoritma 1987'de Kaufman ve Rousseeuw tarafından önerilmiştir (Pujari, 2001). K -medoid, k -ortalama'da olduğu gibi ortalama değerlerin yerine küme merkezindeki gerçek nesnelere kullanılmaktadır (Khatami ve ark., 2015; Jayaswal ve ark., 2017). K -ortalama algoritması gibi çalışan k -medoids kümelemesi, mesafe ölçüsünü bir kez hesaplamakta ve her yinelemeli adımda yeni medoidleri bulmak için kullanır (Park ve ark., 2006). K -medoids algoritmasında ilk olarak her küme nesnesi için bir merkezi nokta olarak temsilci seçilir. Geri kalan nesnelere, merkez noktası yönergelerine göre en yakın kümeye yeniden atanır. Daha sonra her bir temsilci nesne

temsilcisi olmayan nesne ile değiştirerek kümelemenin kalitesini arttırmaya kadar devam ettirilmektedir. K-medoids algoritmasının optimize ettiği objektif fonksiyonu aşağıda verilmiştir (Zhang ve ark., 2006).

$$J_{k-medoids} = \sum_{j=1}^k \sum_{i=1}^n |x_i - c_j| \quad (13)$$

K-medoids algoritmasının aşamaları (Kaufmann, 1987) dediğine göre aşağıdaki gibi ilerlemektedir.

1. Başlangıçta rastgele k adet medoid merkezi seçilir.
2. Her nesnenin seçilen küme merkezlerine olan uzaklığı hesaplandıktan sonra veri kümesindeki nesnelere bu uzaklıklara göre en yakın olan medoide atanır.
3. Medoidlerin yer değiştirmesinin toplam maliyeti hesaplanır
4. Mevcut toplam maliyeti ile bir önceki toplam maliyet arasındaki fark alınarak değiştirme maliyeti hesaplanır.
5. Değiştirme maliyeti sıfırdan büyük veya sıfıra eşit çıkana kadar 2, 3 ve 4 aşamalarını tekrarlayın.

k-medoids algoritmasının avantajları (Uppada, 2014);

1. K-medoids algoritması gürültülü verilere karşı çok sağlamdır.
2. K-medoids algoritması aykırı verilere karşı hassas değildir.

k-medoids algoritmasının dezavantajları (Uppada, 2014);

1. Farklı başlangıç medoidler son kümenin şeklini ve etkisini etkiler.
2. Dışbükey olmayan kümelerin kümelenebilmesinde hassastır.

K-harmonik ortalama algoritması

K-harmonik ortalama algoritması (Zhang ve ark., 2000; Zhang ve ark., 1999) 1999 yılında Zhang, Hsu ve Dayal tarafından sunulmuştur. Daha sonra k-harmonik ortalama algoritması Hamerly ve Elkan tarafından değiştirilmiştir (Hamerly ve Elkan, 2002). Bu algoritma desen tanıma ve uzaktan algılanan veri kümeleme gibi birçok alanda kullanıma sahiptir. Aynı zamanda bu algoritma merkeze dayalı bir kümeleme algoritmasıdır ve her veri noktasından merkezlere olan mesafelerin harmonik ortalamalarını performans işlevinin bileşenleri olarak kullanır (Mahi ve ark., 2015). K-harmonik ortalama algoritmasının optimize ettiği objektif fonksiyonu aşağıda verilmiştir (Hamerly ve Elkan, 2002).

$$H(x) = \sum_{i=1}^n \frac{m}{\sum_{j=1}^k \frac{1}{\|x_i - c_j\|^\alpha}} \quad (14)$$

Burada α kullanıcı tarafından belirlenen bir parametredir ve genellikle $\alpha \geq 2$ olarak belirlenir.

k-harmonik ortalama algoritması için üyelik ve ağırlık fonksiyonları aşağıdaki şekilde tanımlanır (Hamerly ve Elkan, 2002).

$$m(c_j | x_i) = \frac{\|x_i - c_j\|^{-\alpha-2}}{\sum_{j=1}^k \|x_i - c_j\|^{-\alpha-2}} \quad (15)$$

$$w(x_i) = \frac{\sum_{j=1}^k \|x_i - c_j\|^{-\alpha-2}}{(\sum_{j=1}^k \|x_i - c_j\|^{-\alpha})^2} \quad (16)$$

k-harmonik algoritmasının aşamaları (Hamerly ve Elkan, 2002) dediklerine göre aşağıdaki gibi ilerlemektedir.

1. K başlangıç küme merkezlerini $c=(c_1, c_2, \dots, c_k)$ 'den rastgele n nokta $x=(x_1, x_2, \dots, x_n)$ 'den seçilir
2. Amaç fonksiyonu değerini Eşitlik (14)'deki gibi hesaplanır.
3. Her bir xi veri noktası için üyelik fonksiyon işlevini Eşitlik (15) ve her bir ağırlık merkezleri için Eşitlik (16) formülleri ile hesaplanır.
4. Her bir c_j merkez için, tüm veri noktalarının konumu yeniden hesaplanır ve k küme merkezlerini yeniden hesaplamak için Eşitlik (9)'u kullanılır.
5. İkinci, üçüncü ve dördüncü aşamayı durdurma kriteri karşılanıncaya kadar devam ettirilir.

Daha öncede belirtildiği üzere k-ortalama algoritmasının genel olarak, yerel optimaya düşmesi ve küme merkezlerinin başlangıç merkezlerine bağlı kalması algoritmanın temel eksigidir. Ancak k-harmonik ortalama algoritması merkeze dayalı ve yinelemeli bölümlenmiş kümeleme algoritması olduğu için k-ortalama algoritmasının başlangıç küme merkezlerine bağlı kalmasını ortadan kaldırır (Tian ve ark., 2009; Zhang ve ark., 1999). Daha doğrusu k-harmonik ortalama algoritması başlangıç koşullarına daha az hassastır (Alldrin ve ark., 2003). K-ortalama algoritmasının aksine k-harmonik ortalama algoritması her veri noktasının merkeze olan uzaklıklarının harmonik ortalamasını kullanır. K-ortalama algoritması gibi k-harmonik kümeleme algoritması da yerel optimumlarda sıkışabilir. Bu problemi çözmek ve kümeleme kalitesini arttırmak için bazı araştırmacılar k-harmonik ortalama algoritması ile global optimizasyon algoritmalarını birleştiren hibrit yöntemler kullanırlar (Emami ve ark., 2015; Guo ve Peng, 2008).

Bulanık c-ortalama algoritması

Bulanık c-ortalama algoritması 1970'lerin sonunda ortaya çıkmıştır ve 1980'lerin başında algoritmanın bulanık küme teorisi geliştirilmiştir (Bezdek, 1973; Bezdek, 1981; Bezdek ve Pal, 1994; Bezdek ve Intelligence, 1980; Ruspini ve Control, 1969; Ruspini, 1970). Bulanık c-ortalama bazen bulanık k-ortalama da adlandırılır. Dolayısıyla bulanık c-ortalama algoritması, k-ortalama algoritmasının başka bir versiyonudur ve amacı optimal bir bulanık bölüm bularak belirli bir objektif fonksiyonunu en aza indirmektir (Bezdek, 1981; Bezdek ve Intelligence, 1980). Bu algoritma çeşitli giriş veri noktaları arasındaki mesafeye dayalı analiz etmek için kullanılır. Dolayısıyla bu algoritma bir veri kümeleme tekniğidir (Chen ve ark., 2004). Aslında bu algoritma Ruspini tarafından önerilen bulanık c-bölme kavramına dayanmaktadır (Ruspini, 1970). Bulanık bir c-bölme keskin bir bölmenin genellemesidir. Keskin bir bölmede sınıflar arasındaki sınır net bir şekilde bellidir. Dolayısıyla her bir nesne sadece bir sınıfa ait olması gerekmektedir.

Ancak bulanık bir bölmede bir nesne farklı derecelerde farklı sınıflara ait olabilir. Bundan dolayı keskin bölümlerin bulunmasının zor veya imkansız olduğu durumlarda bu algoritma çok yararlıdır (Ruspini, 1970). Bulanık c-ortalama algoritması en küçük kare hata ölçütünün bulanık bir uzantısına dayanmaktadır. Bulanık c-ortalama algoritmasının, k-ortalama algoritmasına göre avantajı, bulanık c-ortalama algoritması her bir nesneye ve kümeyle bir dereceye kadar üyelik atamasıdır. Daha öncede belirtildiği üzere bulanık c-ortalama algoritması kümeler arasındaki çakışmaların olduğu veri kümelerinde daha uygundur (Bezdek, 1980; Bezdek, 1981).

Bulanık c-ortalama algoritmasının optimize ettiği objektif fonksiyonu aşağıda verilmiştir (Bezdek, 1981).

$$F_{c\text{-ortalama}} = \sum_{i=1}^n \sum_{j=1}^k u_{i,j}^r \|x_i - c_j\|^2 \quad (17)$$

Yukarıdaki r bulanık üssüdür ve $r \geq 1$ olmalıdır. Ancak burada r değeri arttıkça algoritma daha da bulanık bir hale gelecektir. Üyelik işlevi aşağıdaki kısıtlamaları karşılamalıdır.

$$u_{i,j} \geq 0, \quad j = 1, \dots, k_j \text{ ve } i = 1, \dots, n$$

$$\sum_{i=1}^n u_{i,j} = 1, \quad j = 1, \dots, k_j \quad (18)$$

Bulanık c-ortalama algoritması için üyelik ve ağırlık fonksiyonları aşağıdaki şekilde tanımlanır (Hamerly ve Elkan, 2002).

$$u(c_j | x_i) = \frac{\|x_i - c_j\|^{-2/(r-1)}}{\sum_{i=1}^n \|x_i - c_j\|^{-2/(r-1)}} \quad (19)$$

$$w(x_i) = 1 \quad (20)$$

Bulanık c-ortalama algoritmasının aşamaları aşağıdaki gibi ilerlemektedir (Rao ve ark., 2010).

1. Rastgele üyelik matrisini Eşitlik (18) kullanarak başlatınız.
2. Her bir c_j merkez için, tüm veri noktalarının konumu hesaplanır ve k küme merkezlerini hesaplamak için Eşitlik (9) kullanılır.
3. Veri noktaları ve merkez kümelerin arasındaki farklılığı Eşitlik (1) kullanarak hesaplanır.
4. Yeni üyelik matrisini güncellemek için Eşitlik (17) kullanılır.
5. Küme merkezleri değişmeye kadar 2, 3 ve 4'üncü aşamaları tekrarlayın.

3.4.3. Yoğunluk tabanlı kümeleme yöntemleri

Daha öncede belirtildiği üzere bölümsel kümeleme yöntemleri nesnelere arasındaki mesafe ölçüsünü kümelemek için kullanılırlar. Genel olarak bu yöntemle üretilen kümeler küresel biçimdedirler. Ancak bu yöntemle rastgele şekilli kümeleri bulmak mümkün değildir. Bundan dolayı böyle kümeleri bulmak için yoğunluğu temel alan yoğunluk tabanlı kümeleme yöntemleri kullanılır (Kokate ve ark., 2018).

Yoğunluk tabanlı kümeleme kavramı 1988 yılında Jain ve Dubes tarafından yoğunluk tahmini ve mod arayışıyla kümeleme olarak adlandırılan konu üzerine araştırılmıştır (Jain ve Dubes, 1988). Yoğunluk tabanlı kümeleme, yüksek yoğunluklu alanlar oluşturarak nesne gruplarının tanımlanması anlamına gelmektedir. Genellikle yüksek yoğunluklu alanların etrafında olan düşük yoğunluklu alanlar gürültülü kısım olarak kabul edilir (Kriegel ve ark., 2011; Rodriguez ve Laio, 2014). Veri kümesinin dağılım yoğunluğunu öğrenmek için önce veriler örtüşmeyen hücrelerden oluşan bir ızgaraya bölünür ve daha sonra bir histogram grafiği yardımıyla aşağı yukarı fikir elde etmek mümkündür (Ester ve ark., 1996). Bu elde edilen histogramda küme merkezleri yüksek veri vektörü frekansına sahip ızgara hücreleridir (Jain ve Dubes, 1988).

Yoğunluk tabanlı kümeleme genellikle literatürde parametrik olmayan bir yöntem olarak tanımlanmaktadır. Dolayısıyla küme sayısının ek bir parametre olarak verilmesi gerekmez. Ayrıca veri kümesinin yoğunluk dağılımı veya kümelerdeki varyans hakkında varsayımların belirlenmesine ihtiyaç duymaz (Jain ve Maheswari, 2012; Kriegel ve ark., 2011; Melnykov ve ark., 2015). Yoğunluk tabanlı kümeleme algoritmaları bir alandaki veri nesnelere yoğunluğuna bağlı olarak kümeler bulur. Yoğunluk tabanlı kümelemenin temel fikri, bir kümenin her bir örneği için belirli bir yarıçapın ve komşusunun en az minimum sayıda örnek içermesidir (Chauhan, 2014). Bu yöntemin amacı kümeleri ve dağılım parametrelerini tanımlamaktır. Ayrıca bu yöntem dışbükey olmayan rastgele şekilli kümeleri keşfetmek için tasarlanmıştır (Rokach ve Maimon, 2005).

Yoğunluk tabanlı kümeleme yönteminin avantajları (Rokach ve Maimon, 2005);

1. Değişik boyutlardaki kümeleri bulabilir.
2. Gürültülü ve aykırı değerlere karşı güçlüdür.

Yoğunluk tabanlı kümeleme yönteminin dezavantajları (Rokach ve Maimon, 2005);

1. Giriş parametrelerin ayarlanmasına karşı hassastır.
2. Yüksek boyutlu veri kümeleri için uygun değildir.
3. Küme tanımlayıcıları çok zayıftır.

Yoğunluk tabanlı kümeleme yönteminin önemli algoritmalarından DBSCAN (Ester ve ark., 1996; Kriegel ve ark., 2011), OPTICS (Ankerst ve ark., 1999) ve DenClue (Hinneburg ve Keim, 1998) aşağıdaki kısımda anlatılmıştır.

DBSCAN

En popüler olan yoğunluk tabanlı kümeleme algoritmalarından biri olan DBSCAN, 1996 yılında Ester ve arkadaşları tarafından önerilmiştir. Bu algoritma büyük veri kümeleri için verimli olduğundan dolayı yaygın bir popülerlik kazanmıştır. Ayrıca hiyerarşik ve bölümsel kümeleme algoritmalarından daha iyi bir performans göstermektedir. Ancak bu performans verilerin boyutlarına da bağlıdır (Kotsiantis ve ark., 2004). Bu algoritma 'S' şekli ve oval olan kümeleri bulmak için tasarlanmıştır (Rokach ve Maimon, 2005). Bu

algoritmanın iki giriş parametresi vardır. Birincisi yarıçapıdır ve r ile gösteriliyor. İkincisi ise k ile gösterilen minimum vektör sayısıdır. Bu iki değer göz önüne alındığında, bir veri vektörünün yerel yoğunluğu, veri vektörünün yarıçapında k ile gösterilen veri vektörlerinin sayısı olarak tanımlanır (Gupta, 2014). Geçtiğimiz yıllarda DBSCAN algoritması birçok araştırma konusu olmuştur. Bundan dolayı literatürde bu algoritmanın birçok uzantısı önerilmiştir (Böhm ve ark., 2009; Brecheisen ve ark., 2004; Brecheisen ve ark., 2006; Ester ve ark., 1996; Kailing ve ark., 2004; Lai ve Nguyen, 2004).

DBSCAN algoritmasının adımları aşağıdaki gibi ilerlemektedir (Han ve ark., 2012).

1. Başlangıçta tüm nesnelere ziyaret edilmiş olarak işaretlenir.
2. Rastgele ziyaret edilmiş bir m nesne seçilir ve bu m nesnesi ziyaret edilmiş olarak işaretlenir.
3. Çevresindeki en az minimum sayıda nokta olup olmadığı kontrol edilir ve ardından m noktası güdültülü veri olarak kabul edilir.
4. Yeni bir C kümesi oluşturulur ve m noktası C kümesine eklenir.
5. m noktasının etrafındaki kalan tüm nesnelere N 'ye eklenir.
6. Tüm nesnelere ziyaret edilene kadar, N 'nin ziyaret edilmiş her bir nesnesi için aynı adımlar tekrarlanır.

DBSCAN algoritmasının avantajları (Joshi ve ark., 2013);

1. Küme sayısının önceden belirlenmesine gerek yoktur.
2. İşlem süresi çok hızlıdır.
3. Küresel olmayan şekilli kümeleri bulabilir.
4. Aykırı değerleri tespit edebilir.

DBSCAN algoritmasının dezavantajları (Joshi ve ark., 2013);

1. Verilerin değişebilen yoğunluğu varsa kümeleri bulmakta zorlanır.
2. Yüksek boyutlu veriler için uygun bir yöntem değildir.
3. Yarıçapın ve komşusunun en az minimum sayısı önceden kullanıcı tarafından belirlenmesi gerekmektedir.

OPTICS

OPTICS algoritması 1999 yılında Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel ve Jörg Sander tarafından önerildi. Bu algoritmanın ana fikri DBSCAN algoritmasıyla çok benzerdir. Ancak DBSCAN algoritmasıyla olan farkı verilerdeki değişen yoğunluğun anlamlı kümeler bulunmasında sorun olmamasıdır (Firdaus ve Uddin, 2015). Başka bir deyişle OPTICS diğer yoğunluk tabanlı kümeleme algoritmalarına göre çok sayıda parametre ayarı kullanır. Bundan dolayı OPTICS, DBSCAN algoritmasının geliştirilmiş versiyonudur. Bu algoritma DBSCAN algoritmasının bir uzantısıdır ve yoğunluk tabanlı küme siparişi kavramı üzerinde uygulanmaktadır (Han ve ark., 2012).

OPTICS algoritması belirli olan bir veri kümesindeki bilinmeyen bir kümelemeyi tanımlayan yoğunluk tabanlı

bir kümeleme yöntemidir. Bu algoritma yüksek yoğunluklu kümeleri düşük yoğunluklu kümelere tercih eder ve böylece veri nesnelere işleme sırasını korur (Reddy ve Ussenaiah, 2012). Ayrıca bu algoritma veri tabanının doğru özellik kümelerine yol açan bir sıralama oluşturur (Han ve ark., 2012; Reddy ve Ussenaiah, 2012). OPTICS algoritmasının adımları aşağıdaki gibi ilerlemektedir (Han ve ark., 2012).

1. Önce veri nesnelere siparişi oluşturulur ve her bir veri nesnesine çekirdek mesafesi ve erişebilirlik mesafe kaydedilir.
2. Daha sonra yoğunluk tabanlı kümeleri bulmak için yukarıdaki kaydedilen bilgileri kullanılır.

OPTICS algoritmasının avantajları (Reddy ve Ussenaiah, 2012);

1. Verilerin değişken yoğunluğu varsa doğru kümeler bulma sorununu çözer (Joshi ve ark., 2013).
2. Nesnelere belirli bir sıraya göre sonuçlandırılır.

OPTICS algoritmasının dezavantajları (Reddy ve Ussenaiah, 2012);

1. Bazı yoğunluk türleri küme sınırlarının bulunmasını düşürür.
2. Hatalı verilere karşı az duyarlıdır.

OPTICS algoritması veri nesnelere işleme sırasını koruyarak kaliteli kümeleme yapabilir. Daha doğrusu bu sıralama işleminde yüksek yoğunluklu kümeleri düşük yoğunluklu kümelere tercih eder. Bu algoritma siparişi vermenin yanı sıra, her veri nesnesinin çekirdek mesafesini ve erişebilirlik mesafesini kaydederek daha kaliteli kümeleri bulur. OPTICS kümeleme algoritması veri nesnelere erişebilirlik ve çekirdek değerleri olan bir dizi sıralamasıyla etkili bir küme sırası sağlar. Ayrıca bu algoritma çok boyutlu verilerde piksel odaklı görselleştirme tekniklerini kullanır (Han ve ark., 2012).

DenClue

DenClue algoritması bir dizi yoğunluk dağılım fonksiyonuna dayanmaktadır. Bu algoritma Hinneburg ve Gabriel tarafından 1998 yılında önerilmiştir (Han ve ark., 2012). Bu algoritma kullanıcı tarafından tanımlanan parametrelere dayalıdır ve farklı parametrelerle aynı veritabanında birden fazla küme oluşturabilir. Kümelerin sayısının başlangıçta belirlenmesine gerek olmamasının nedeni kümelerin sadece yoğunluk bazında üretilmesidir (Murtagh ve Contreras, 2011). Bu algoritma aşağıdaki düşünceler üzerine kuruludur (Han ve ark., 2012).

1. Her bir veri noktasının etkisi, bir matematiksel fonksiyonu kullanarak modellenir ve etki fonksiyonu olarak adlandırılır. Dolayısıyla bir veri noktasının komşuları içindeki etkiyi tarif eder.
2. Veri alanının toplam yoğunluğu, etki fonksiyonunun toplamı olarak tüm veri noktalarına uygulanarak modellenir.
3. Kümeler matematiksel olarak yoğunluk çekicilerini belirleyebilirler.

DenClue algoritmasının aşamaları aşağıdaki gibi ilerlemektedir (Murtagh ve Contreras, 2011).

1. Bir ön küme haritası oluşturulur ve veri noktaları sadece yoğunluk sahibi olan hiper küplere bölünür.

2. Daha sonra kümeleri oluşturmak için son derece yoğun olan küpler ve bu küplere bağlı olan diğer küpler bir araya getirilir.

DenClue algoritmasının avantajları (Reddy ve Ussenaiah, 2012);

1. Hatalı verileri iyi bir şekilde tespit eder.
2. Yüksek boyutlu veri kümelerinde küresel olmayan kümelerin kısa bir tanımını yapabilir.
3. İşlem süresi DBSCAN algoritmasından daha hızlıdır.

DenClue algoritmasının dezavantajları (Reddy ve Ussenaiah, 2012);

1. Birçok parametre gerektiren algoritmadır.
2. Aykırı değerlere karşı az duyarlıdır.

3.4.4. Izgara tabanlı kümeleme yöntemleri

Izgara tabanlı kümeleme algoritmasının nesne alanı, izgara yapısı olarak adlandırılan sınırlı sayıda hücrelere bölünür ve tüm kümeleme işlemleri izgara yapısı üzerinden gerçekleştirilir (Kokate ve ark., 2018). Dolayısıyla bu yöntem kümeleme için tüm işlemlerin gerçekleştirildiği bir izgara yapısını oluşturan, sınırlı sayıda hücre ile nesne alanını oluşturur. Izgara tabanlı kümeleme yaklaşımı, geleneksel kümeleme algoritmalarından farklı olmasının sebebi veri noktalarıyla değil, veri noktalarını çevreleyen değer alanı ile ilgili olmasıdır (Cheng ve ark., 2018). Daha doğrusu izgara tabanlı kümeleme veri noktaları yerine hücreleri dikkate alır ve bunun nedeni izgaraya dayalı kümeleme algoritmaları genellikle hesaplama açısından diğer kümeleme algoritmalarından daha verimlidir (Edla ve Jana, 2012; Parikh ve ark., 2014).

Izgara tabanlı kümeleme tekniği, kümede bulunan tüm nesnelere izgara olarak tanınan bir dizi kare hücreye atar. Bu izgaralar, izgaraya benzer bir yapı oluşturmak için bir araya getirilir ve tüm kümeleme işlemleri bu izgaralara uygulanır. Bu teknikte kümeleme işlemleri veri nesnelere sayısına bağlı olmadığı ve sadece x ve y boyutuna bağlı kaldığından dolayı bu yöntem daha güçlü ve daha etkili sonuçlar üretmektedir (Kaur, 2015).

Genel olarak izgara tabanlı kümeleme algoritması aşağıdaki beş temel adımdan oluşmaktadır (Deepa ve ark., 2014).

1. Izgara yapısının oluşturulması(veri alanının sınırlı sayıda hücreye bölünmesi)
2. Her bir hücre için hücre yoğunluğunu hesaplamak.
3. Hücreleri yoğunluklarına göre sıralamak.
4. Küme merkezlerini belirlemek.
5. Komşu hücrelere geçmek.

Izgara tabanlı kümeleme algoritmasının avantajları (Deepa ve ark., 2014): Çok hızlı çalışabilmesi, küme kalitesinin iyi olması, mesafe hesaplamaması, kümelerin nesnelere üzerinde değil hücreler üzerinde gerçekleştirilmesi, hangi kümelerin komşu olduğunun kolayca belirlenmesi ve şekillerin sadece izgara hücreleri ile sınırlı olmasıdır.

Izgara tabanlı kümeleme algoritmasının dezavantajı (Deepa ve ark., 2014) ise tüm küme sınırları yatay veya dikeydir. Bundan dolayı köşegen sınırlı bir kümeyi bu algoritma belirleyemez.

Izgara tabanlı kümeleme yöntemlerinin önemli algoritmaları STING (Reddy ve Ussenaiah, 2012; Wang ve ark., 1997), WAVE Cluster (Sheikholeslami ve ark., 2000) ve CLIQUE (Agrawal ve ark., 1998; Agrawal ve ark., 1999; Ilango ve ark., 2010) aşağıdaki kısımda anlatılmıştır.

STING

STING algoritması 1997 yılında Wang, Yang ve Muntz tarafından önerildi (Wang ve ark., 1997). Bu algoritmada, uzamsal alan dikdörtgen hücrelere bölünür. Çeşitli çözünürlüğe sahip olan bu tip dikdörtgen hücrelerin birçok farklı seviyeleri vardır ve bu hücreler hiyerarşik bir yapı oluştururlar. Yüksek bir seviyedeki her bir hücre, bir sonraki düşük seviyedeki birkaç hücreyi oluşturmak için bölümlere ayrılır. Her bir hücrenin istatistiksel bilgileri önceden hesaplanarak kaydedilir ve bu bilgiler sorguları yanıtlamak için kullanılır (Han ve ark., 2011; Reddy ve Ussenaiah, 2012; Wang ve ark., 1997).

STING algoritmasının parametreleri aşağıdaki gibi hesaplanmaktadır (Wang ve ark., 1997).

$$n = \sum_i n_i, m = \frac{\sum_i m_i n_i}{n}, s = \sqrt{\frac{\sum_i (s_i^2 + m_i^2) n_i}{n} - m^2}, \min = \min_i (\min_i), \max = \max_i (\max_i) \quad (20)$$

$n, m, s, \min, \max, dist$ geçerli hücrenin parametreleridir. $n_i, m_i, s_i, \min_i, \max_i$ ve $dist_i$ karşılık gelen alt düzey hücrelerin parametreleridir.

STING algoritmasının adımları (Reddy ve Ussenaiah, 2012);

1. Daha düşük düzeyde, her bir hücre alt hücrelere bölünür.
2. i düzeyindeki bir hücre, $i+1$ düzeyindeki çocuklarının birleşmesine karşılık gelir.
3. Her hücrenin(yapraklar hariç) dört çocuğu vardır ve her çocuk ana hücrenin bir çeyreğine karşılık gelir.
4. Her izgara hücrenin özellikleri ile ilgili istatistiksel bilgiler önceden hesaplanarak kaydedilir.
5. Daha yüksek düzeydeki hücrelerin istatistiksel parametreleri, daha düşük düzeydeki hücrelerin parametrelerinden kolayca hesaplanır.
6. Her hücre için, öznitelikten bağımsız parametreler ve öznitelige bağlı parametreler Eşitlik (21) vardır.
7. Dağıtım türleri normal, düzgün, üstel veya hiçbirisi olabilir.
8. Dağılım değeri kullanıcı tarafından belirlenebilir veya X_2 testi gibi hipotez testleri ile hesaplanabilir.
9. Veriler veri tabanına yüklendiği zaman, parametreler doğrudan alt düzey hücrelerin hesaplanmasını sağlar.
10. Alakasız hücreler çıkarılır ve bu işlem alt seviyeye ulaşılan kadar tekrarlanır.

STING algoritmasının avantajları (Han ve ark., 2011); öncelikle sorgudan bağımsız ve paralelleştirmesi kolaydır. Paralel işlemi ve artımlı güncellenmeyi kolaylaştırır. Çok verimli bir işlevi vardır.

STING algoritmasının dezavantajları (Han ve ark., 2011) ise bu kümelemenin kalitesi, ızgara yapısının en düşük düzeyine bağlıdır ve bu yöntemin çok hızlı işlem yapabilmesine rağmen doğru olmayan kümeler oluşturur.

STING küme analizinin çoklu çözüm yaklaşımını kullandığından dolayı, bu kümelemenin kalitesi ızgara yapısının ayrıntı düzeyine bağlıdır. Eğer taneciklik çok iyi ise, işleme maliyeti daha çok artacaktır. Ancak ızgara yapısının alt seviyesi çok kalınsa kümeleme kalitesi azalacaktır. Ayrıca STING bir ana hücrenin oluşunda, çocuklar ve komşu hücrelerin arasındaki uzamsal ilişkiyi dikkate almaz. Sonuç olarak ortaya çıkan kümelerin şekilleri yatay veya dikey olacaktır ve hiçbir köşegen sınır algılanmaz. Bu tekniğin işlem süresinin hızlı olmasına rağmen kümelerin kalitesini ve doğruluğunu kötü etkileyebilir (Han ve ark., 2012).

Wave Cluster

Wave Cluster algoritması 1998 yılında Sheikholeslami ve arkadaşları tarafından geliştirilmiştir. Bu algoritma orijinal özellik alanını dalgacık dönüşümü ile dönüştüren çok çözünürlüklü bir kümeleme yöntemidir ve büyük uzamsal veritabanlarındaki kümeleri bulmak için kullanılır (Sheikholeslami ve ark., 1998). Bu algoritma her hücredeki veri noktalarını, nokta sayısına göre ifade eden iki boyutlu ızgarayı kullanır (Murtagh ve Contreras, 2011). Bu algoritmanın çok hızlı işlem süresi vardır. Bundan dolayı görevleri çok hızlı bir şekilde yerine getirir. Ayrıca bu algoritmanın temel amacı dalgacık dönüşümü uygulanarak orijinal özellik uzayını dönüştürmek ve daha sonra yeni uzaydaki yoğun alanları bulmaktır (Sheikholeslami ve ark., 2000).

Dalgacık dönüşümü, bir sinyali farklı frekans alt bantlarına ayırarak bir sinyal işleme yöntemidir. Dalgacık modeli, tek boyutlu dönüşümün birden fazla uygulanabildiği n boyutlu sinyallere genelleştirilebilir (Sheikholeslami ve ark., 1998). Bu yöntem veri sıkıştırma (Hilton ve ark., 1994) ve dalgalardan özellik elde etmek için kullanılır (Sheikholeslami ve Zhang, 1997; Sheikholeslami ve ark., 1997; Smith ve Chang, 1994).

Wave Cluster algoritmasının adımları (Sheikholeslami ve ark., 2000);

1. Çok boyutlu veri nesnelere özellik vektörleri girdi olarak verilir.
2. Özellik alanı nicelendikten sonra nesnelere hücrelere atanır.
3. Dalgacık dönüşümü nicelenmiş özellik alanına uygulanır.
4. Değişik düzeylerde, dönüştürülmüş özellik alanının alt gruplardaki kümeler bulunur.
5. Hücrelere etiket atanır.
6. Arama tablosu oluşturulur.
7. Kümelerdeki nesnelere çizilir.

Wave Cluster algoritmasının avantajları (Sheikholeslami ve ark., 2000) sırasıyla dalgacık dönüşümü otomatik olarak sonuçlardan aykırı değerleri çıkarır. Dalgacık dönüşümü yüksek olasılıkta, çeşitli doğruluk

düzeylerinde kümelerin belirlenmesinde yardımcı olur. Dalgacık tabanlı kümelemenin işlem süresi çok hızlıdır. Verilerin giriş sırasına önem vermez. 20 boyuta kadar veri işleyebilir. Rasgele şekillere sahip kümeleri bulabilir ve verimli bir şekilde herhangi bir uzamsal veri tabanı işleyebilir.

CLIQUE

CLIQUE algoritması 1998 yılında Agrawal ve arkadaşları tarafından çok boyutlu verilerde kümeleme yapmak için önerildi (Agrawal ve ark., 1998). Bu yöntem çok boyutlu veri alanında kümeleme yapmak için ızgara tabanlı ve yoğun tabanlı bir yaklaşımı temsil eder. Ayrıca her bir boyutuda bir ızgara yapısı gibi bölümler ve bir hücrenin içerdiği nokta sayısına göre yoğun olup olmadığını belirler (Belacel ve ark., 2006).

CLIQUE (QUEST'te kümeleme), statik ızgaralar çizebilen ve aşağıdan yukarıya bir alt alan kümeleme yapan bir algoritmadır. Apriori yöntemini arama alanını azaltmak için kullanır (Kriegel ve Zimek, 2010). Bu algoritma daha öncede belirtildiği üzere yoğunluk ve ızgara tabanlı bir algoritmadır. Daha doğrusu alt uzay kümeleme algoritmasıdır ve girdi eşiği olarak yoğunluk eşiğini ve ızgara sayısını alarak kümeleri ayırır eder. CLIQUE, ilk aşamada tek bir boyutu kullanarak çok boyutlu veriler üzerinde çalışır ve daha sonra çok boyutlu olana doğru hareket eder (Agrawal ve ark., 1998).

CLIQUE çok boyutlu kümelemeyi iki aşamada uygulayabilir (Han ve ark., 2011).

1. CLIQUE, verilen ızgara boyutuna göre n boyutlu veri alanını üst üste gelmeyen dikdörtgen birimlere böler ve daha sonra belirli bir eşik değerine göre yoğun olan bölgeyi bulur. Buradaki yoğunluk, veri noktalarının eşik değerini aşmasıyla ortaya çıkar.
2. Kümeler, apriori yöntemini kullanarak tüm yoğun alt uzaylardan üretilir (Kriegel ve Zimek, 2010). CLIQUE algoritması ilk önce alt alanlardaki maksimum yoğun alanları bulur ve daha sonra o belirlediği alandan her küme için minimum kaplamayı belirleyerek elde edilen kümeler için minimum açıklama üretir. Tüm boyutlar örtülene kadar aynı işlem tekrarlanır.

CLIQUE algoritmasının adımları aşağıdaki kısımda belirtilmiştir (Agrawal ve ark., 1998).

1. Her boyutu aynı sayıda eşit uzunluk mesafelerine ayırır.
2. N boyutlu bir veri alanını örtüşmeyen dikdörtgen birimlerine böler.
3. Toplam veri noktalarının oranı, giriş modeli parametresini aşarsa birim yoğundur.
4. Bir küme bir alt düzey içindeki maksimum bağlı yoğun birimler kümesidir.

CLIQUE algoritmasının avantajları (Han ve ark., 2011); CLIQUE otomatik olarak en yüksek boyutsallığın alt düzeylerinde yüksek yoğunluklu kümeler bulur. Verilerin giriş sırasına önem vermez. Ölçeklenebilir bir algoritmadır ve verilerin boyutu arttıkça iyi bir performans gösterebilir.

CLIQUE algoritmasının dezavantajları (Han ve ark., 2011) ise ızgara boyutunun ve yoğunluk eşiğinin ayarlanması

gerekir. Eşik değeri sabit olursa kümeler çok değişik yoğunluklara sahip olur ve bu kümeleme işlemi kötü etkiler. Yüksek maliyetli bir işlem olabilir. 4. düşük ve yüksek boyutlar için aynı yoğunluk eşiğine sahiptir.

3.4.5. Model tabanlı kümeleme yöntemleri

Kümeleme yöntemleri, verilere önceden tanımlanmış matematiksel modeller uygulayarak bunları optimize eder. Burada temel varsayım, verilerin olasılık dağılımları açısından hibrit olmasıdır. Model tabanlı yöntemler standart istatistiklere dayanarak küme sayısını belirler. Bundan dolayı Standart istatistik hesaplanırken, gürültülü ve aykırı değerleri dikkate alarak sağlam kümeleme yöntemine sahip olunur. Dolayısıyla böyle yöntemler gürültülü ve aykırı değerlere karşı çok sağlam yöntemlerdir (Kokate ve ark., 2018).

Bu yöntemin temel fikri, her kümeye belli bir model seçmek ve bu model için en uygun olanı bulmaktır. Birisi istatistiksel öğrenme yöntemine ve diğeri ise sinir ağ öğrenme yöntemine dayalı olarak, iki tür model tabanlı kümeleme algoritması vardır (Xu ve Tian, 2015). İstatistiksel öğrenme yönteminin popüler olan algoritmalarından beklenti-maksimize algoritması (Dempster ve ark., 1977) ve CobWeb (Fisher, 1987) algoritması, ayrıca sinir ağ öğrenme yönteminden ise SOM (Kohonen, 1982; Kohonen, 1990) algoritması aşağıda özet olarak anlatılmıştır.

3.4.5.1. Beklenti-maksimize algoritması

Beklenti maksimize algoritması Dempster (Dempster ve ark., 1977) tarafından kanıtlanmadan ve beklenti maksimize terimini ortaya koymadan önce birkaç farklı araştırmacı tarafından bağımsız olarak keşfedildi ve kullanıldı. Beklenti maksimize algoritmasının tipik uygulama alanlarından birisi genetik bilimdir (Jiang ve ark., 1994). Başka kullanım alanlarından biriside karışım dağılımlarının parametrelerini tahmin etmektir (Redner ve Walker, 1984). Ayrıca bu algoritma klinik ve sosyolojik çalışmalarında yaygın olarak kullanılmaktadır (Schmee ve Hahn, 1979).

Beklenti maksimize algoritması, kümeleme algoritmalarından biridir (Bishop, 1995; McLachlan ve Krishnan, 2007; Redner ve Walker, 1984). Bu algoritma yinelemeli bir optimizasyon yöntemidir ve bundan dolayı bazı bilinmeyen parametreleri tahmin etmek için kullanılır (Dellaert, 2002; Hamerly ve Elkan, 2003). Dolayısıyla beklenti maksimize algoritması temel olarak istatistiksel modellerin parametrelerini değerlendirmek için maksimum olasılığı arar (Weinstein, 2006) aynı zamanda eksik veya gizli verilerin olması durumunda maksimum olasılık tahminini hesaplamak için etkili bir yinelemeli prosedürdür.

Beklenti maksimize algoritmasının optimize ettiği objektif fonksiyonu aşağıda verilmiştir (Hamerly ve Elkan, 2002).

$$BM_{\text{beklenti-maksimize}} = -\sum_{i=1}^n \log(\sum_{j=1}^k p(x_i|c_j)p(c_j)) \quad (22)$$

Beklenti maksimize algoritması için üyelik ve ağırlık fonksiyonları aşağıdaki şekilde tanımlanır (Hamerly ve Elkan, 2002):

$$m_{BM}(c_j | x_i) = \frac{p(x_i | c_j)p(c_j)}{p(x_i)} \quad (23)$$

$$w_{BM}(x_i) = 1 \quad (24)$$

Beklenti maksimize algoritması ilk olarak parametrelerin tahmini ile başlar. Ardından değerleri belli olan verileri kullanarak, bilinmeyen verilerin değerlerin hesaplayarak bir beklenti adımı uygulanır (Hamerly ve Elkan, 2003). Genel olarak beklenti maksimize algoritması iki temel adımdan oluşmaktadır (Dempster ve ark., 1977).

Beklenti adımı; bu adım, ölçülen verilerdeki her bir kümeye ait her veri noktasının olasılığını tahmin etmektedir.

Maksimizasyon adımı; bu adım, bir sonraki adım için her bir kümenin olasılık dağılımının parametrelerini tahmin etmekten sorumludur. Beklenti maksimize algoritması ile k-ortalama algoritmalarının performansı (Hamerly ve Elkan, 2003; Veenman ve ark., 2002) araştırmalarına göre karşılaştırılabilir. Ayrıca (Alldrin ve ark., 2003) söylediklerine göre bu algoritma çok boyutlu veri setlerinde başarısız olduğu ortaya çıkmıştır. Bunlara ek olarak beklenti maksimize algoritması parametrelerin başlangıç tahminine bağlıdır (Turi, 2001) ve kullanıcının küme sayısını önceden belirtmesi gerekir (Hamerly ve Elkan, 2003).

CobWeb

CobWeb algoritması 1987 yılında Fisher tarafından önerildi (Fisher, 1987). Bu algoritma sınıflar hiyerarşisi üreten artımlı ve denetimsiz bir kümeleme algoritmasıdır. Artımlı yapısı, Önceden yapılmış kümelenmeyi tekrarlamak zorunda kalmadan, yeni verilerin kümelenmesini sağlayabilir (Theodorakis ve ark., 2004).

Kavramsal kümeleme, kümeleme için bir makine öğrenme örneğidir. Her küme için bir kavram tanımlayıcı oluşturması diğer kümeleme algoritmalarından farklıdır. CobWeb en yaygın kullanılan kavramsal kümeleme algoritmalarından biridir. CobWeb algoritmasının oluşturduğu kümeler bir ağaçtan oluşur. Ağacın yaprakları her kavramı, kök düğüm tüm veri kümesini ve dallar ise veri kümesindeki hiyerarşik kümeleri temsil eder. Aynı zamanda toplam küme sayısı, veri kümesinin boyutuna kadar olabilir. Dolayısıyla CobWeb veri yapısı, her bir düğümün belli bir kavramı temsil ettiği bir ağaçtır. Her kavram bir küme, çoklu küme veya çanta nesnesiyle ilişkilidir. Her bir kavramla ilişkili veriler, o konseptte ait nesnelere için tamsayı sayımıdır (Xia ve Xi, 2007).

Daha öncede belirtildiği üzere CobWeb, hiyerarşik kavramsal kümeleme için artımlı bir yöntemdir. CobWeb kademeli olarak gözlemleri bir sınıflandırma ağacında oluşturur. Bir sınıflandırma ağacındaki her düğüm bir sınıfı temsil eder ve düğüm altında sınıflandırılan

nesnelerin öznitelik değeri dağılımlarını özetleyen olasılıklı bir kavram tarafından etiketlenir. Sınıflandırma ağacı, eksik nitelikleri veya yeni bir nesnenin sınıfını tahmin etmek için kullanılabilir (Iba ve Langley, 2011).

SOM

SOM algoritması 1982 yılında Teuvo kohonen tarafından sunulan çok popüler bir algoritmadır (Kohonen, 1982). Denetimsiz bir kümeleme algoritmasına ilaveten güçlü bir görüntüleme aracıdır. Ekonomi, sanayi, yönetim, sosyoloji, coğrafya, metin madenciliği gibi birçok uygulama alanında yaygın olarak kullanılmaktadır (Cottrell ve ark., 2018). SOM aynı zamanda noron hesaplama algoritması olarak da bilinmektedir ve bu algoritmanın amacı benzer özelliklere sahip nesne gruplarını pozitif olarak tanımlamak ve anlamlı boyutsal verileri uzayın iki boyutuna doğru eşleştirmektir (Mallick ve ark., 2019).

Som bir veri kümesindeki kümelerin sayısını otomatik olarak bulmak için kullanılabilir. Bu algoritmanın amacı herhangi bir denetim olamadan veri kümesindeki kümeleri bulmaktır (Pandya ve Macy, 1995). Som tek katmanlı bir denetimsiz yapay sinir ağıdır. Bunun nedeni som algoritması giriş örüntülerinin, bir durdurma şartı sağlana kadar kademeli olarak değiştirilen ağırlıklar yoluyla çıkış düğümlerini ilişkilendirilmesidir (Jain ve ark., 1999). Ayrıca som, rekabetçi öğrenmeyi düğümlerin topolojik yapılandırması ile birleştirir. Dolayısıyla komşu düğümler benzer ağırlık vektörlerine sahip olma eğilimindedir (Mehrotra ve ark., 1997; Pandya ve Macy, 1995).

SOM algoritması aşağıdaki dezavantajlara sahiptir (Jain ve ark., 1999). Bunlar; başlangıç koşullarına bağlı bir algoritmadır. Öğrenme hızı parametresi ve komşuluk fonksiyonu algoritmanın performansını etkiler. Sadece hiper küresel kümelerde iyi çalışabilir. Çıkış düğümde sabit sayıyı kullanır. Veri noktalarının sunulduğu sıraya bağlıdır. Bu sorunun üstesinden gelmek için, her yineleme sırasında veri noktalarının seçimi rastgele seçilebilir (Pandya ve Macy, 1995).

4. Sonuç

Bu derlemede farklı kümeleme yöntemlerine genel bir bakış sunulmuştur. İlk olarak algoritmanın tarihçesi, günümüzdeki anlamı ve nerelerde kullanıldığından bahsedilmiştir. Daha sonra kümeleme analizi nedir ve kümeleme yöntemleri nerelerde kullanıldığı anlatılmıştır. Ayrıca kümeler arasındaki benzerlik mesafe ölçüm yöntemleri incelenmiştir. Son olarak ta literatürde bulunan hiyerarşik kümeleme, bölümsel kümeleme, yoğunluk tabanlı kümeleme, ızgara tabanlı kümeleme ve model tabanlı kümeleme yöntemlerinin birbirinden farklılıkları, avantajları ve dezavantajları tartışılmıştır.

Çıkar İlişkisi

Yazarlar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Kaynaklar

- Agrawal R, Gehrke J, Gunopulos D, Raghavan P. 1998. Automatic subspace clustering of high dimensional data for data mining applications. Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data.
- Agrawal R, Gehrke JE, Gunopulos D, Raghavan P. 1999. Automatic subspace clustering of high dimensional data for data mining applications. In: Google Patents.
- Alldrin N, Smith A, Turnbull DJ. 2003. Clustering with EM and K-means. University of San Diego, California, Tech Report. 261-295.
- Anderberg MR. 2014. Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks, Cilt 19. Academic press.
- Ankerst M, Breunig MM, Kriegel HP, Sander J. 1999. OPTICS: ordering points to identify the clustering structure. ACM SIGMOD Record, 28(2): 49-60.
- Belacel N, Wang Q, Cuperlovic-Culf M. 2006. Clustering methods for microarray gene expression data. OMICS, 104: 507-531.
- Belli F, Beyazit M, Güler N. 2012. Event-Oriented, Model-Based GUI Testing and Reliability Assessment—Approach and Case Study. Advances in Computers, 85: 277-326.
- Berggren JL. 2017. Episodes in the mathematics of medieval Islam: Springer.
- Berkhin P. 2006. A survey of clustering data mining techniques. Grouping multidimensional data s. 25-71: Springer.
- Bezdek JC. 1973. Cluster validity with fuzzy sets.
- Bezdek JC. 1981. Objective function clustering. Pattern recognition with fuzzy objective function algorithms. 43-93: Springer.
- Bezdek JC, Pal SK. 1994. Fuzzy models for pattern recognition.
- Bezdek JC. 1980. A convergence theorem for the fuzzy ISODATA clustering algorithms. IEEE transact, 1: 1-8.
- Bhat A. 2014. K-medoids clustering using partitioning around medoids for performing face recognition. Int J Soft Comput, Math Cont, 3(3): 1-12.
- Bishop CM. 1995. Neural networks for pattern recognition: Oxford university press.
- Blass A, Gurevich Y. 2004. Algorithms: A quest for absolute definitions. Current Trends in Theoretical Computer Science: The Challenge of the New Century Vol 1: Algorithms and Complexity Vol 2: Formal Models and Semantics, Scientific.
- Böhm C, Noll R, Plant C, Wackersreuther B. 2009. Density-based clustering using graphics processors. Proceedings of the 18th ACM conference on Information and knowledge management.
- Brecheisen S, Kriegel HP, Pfeifle M. 2004. Efficient density-based clustering of complex objects. Paper presented at the Fourth IEEE International Conference on Data Mining ICDM'04.
- Brecheisen S, Kriegel HP, Pfeifle M. 2006. Parallel density-based clustering of complex objects. Pacific-Asia Conference on Knowledge Discovery and Data Mining.
- Chakravarthy SV, Ghosh J. 1996. Scale-based clustering using the radial basis function network. IEEE Transact, 7(5): 1250-1261.
- Chauhan R. 2014. Clustering Techniques: a comprehensive study of various clustering techniques. Inter J Advance Res Comput Sci, 5(5).
- Chen S, Zhang D. 2004. Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. IEEE Transact, 34(4): 1907-1916.
- Cheng W, Wang W, Batista S. 2018. Grid-based clustering. Data Clustering s. 128-148: Chapman and Hall/CRC.
- Cherng JS, Lo MJ. 2001. A hypergraph based clustering algorithm for spatial data sets. Proceedings 2001 IEEE International Conference on Data Mining.
- Cord M, Cunningham P. 2008. Machine learning techniques for

- multimedia: case studies on organization and retrieval: Springer Science & Business Media.
- Cormen TH, Leiserson CE, Rivest RL, Stein C. 2009. Introduction to algorithms: MIT press.
- Cottrell M, Olteanu M, Rossi F, Villa Vialaneix N. 2018. Self-organizing maps, theory and applications.
- Dabhi DP, Patel MR. 2016. Extensive survey on hierarchical clustering methods in data mining. *Int Res J Eng and Tech IRJET*, 3: 659-665.
- Davies ER. 2004. Machine vision: theory, algorithms, practicalities: Elsevier.
- Deepa MS, Sujatha N. 2014. Comparative Studies of Various Clustering Techniques and Its Characteristics. *Int J Adv Netw App*, 5(6): 2104.
- Dellaert F. 2002. The expectation maximization algorithm.
- Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J Royal Stat Soc: Series B Meth*, 39(1): 1-22.
- Dunham MH. 2006. Data mining: Introductory and advanced topics: Pearson Education India.
- Edla DR, Jana PK. 2012. A grid clustering algorithm using cluster boundaries. *World Congress on Information and Communication Technologies*.
- Emami H, Dami S, Shirazi H. 2015. K-Harmonic means data clustering with imperialist competitive algorithm. *UPB Sci Bull, Series C*, 77(1): 91-104.
- Erickson J. 2019. Algorithms Jeff Erickson.
- Ester M, Kriegel HP, Sander J, Xu X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Paper presented at the Kdd.
- Estivill Castro V. 2002. Why so many clustering algorithms: a position paper. *ACM SIGKDD*, 4(1): 65-75.
- Everitt BS. 1979. Unresolved problems in cluster analysis. *Biometrics*, 169-181.
- Firdaus S, Uddin MA. 2015. A survey on clustering algorithms and complexity analysis. *Int J Comp Sci Iss IJCSI*, 12(2): 62.
- Fisher DH. 1987. Improving Inference through Conceptual Clustering. *AAAI*, 461-465.
- Fisher DH. 1987. Knowledge acquisition via incremental conceptual clustering. *Mach Learn*, 2(2): 139-172.
- Forgy EW. 1965. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21: 768-769.
- Frigui H, Krishnapuram R. 1997. Clustering by competitive agglomeration. *Pattern Recog*, 30(7): 1109-1119.
- Goldschlager L, Lister A. 1988. Computer science: a modern introduction: Prentice Hall International UK Ltd.
- Goodrich MT, Tamassia R. 2014. Algorithm design and applications: Wiley Publishing.
- Guha S, Rastogi R, Shim K. 1998. CURE: an efficient clustering algorithm for large databases. *ACM Sigmod Rec*, 27(2): 73-84.
- Guha S, Rastogi R, Shim K. 2000. ROCK: A robust clustering algorithm for categorical attributes. *Information Sys*, 25(5): 345-366.
- Guo C, Peng L. 2008. A hybrid clustering algorithm based on dimensional reduction and k-harmonic means. 4th International Conference on Wireless Communications, Networking and Mobile Computing.
- Gupta GK. 2014. Introduction to data mining with case studies: PHI Learning Pvt. Ltd.
- Hamerly G, Elkan C. 2002. Alternatives to the k-means algorithm that find better clusterings. Proceedings of the eleventh international conference on Information and knowledge management.
- Hamerly GJ, Elkan CP. 2003. Learning structure and concepts in data through data clustering. University of California, San Diego,
- Han J, Kamber M, Pei J. 2012. Data mining: concepts and techniques. Elsevier.
- Han J, Pei J, Kamber M. 2011. Data mining: concepts and techniques: Elsevier.
- Hartigan JA. 1996. Clustering algorithms. New York: John Wiley & Sons.
- Hartigan JA, Wong MA. 1979. Algorithm AS 136: a k-means clustering algorithm. *J Royal Stat Soc, Series C*, 28(1): 100-108.
- Hilton ML, Jawerth BD, Sengupta A. 1994. Compressing still and moving images with wavelets. *Multimedia Sys*, 2(5): 218-227.
- Hinneburg A, Keim DA. 1998. An efficient approach to clustering in large multimedia databases with noise. Paper presented at the KDD.
- Hinton GE, Sejnowski TJ, Poggio TA. 1999. Unsupervised learning: foundations of neural computation: MIT press.
- Huang Z. 1998. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining Know Disc*, 2(3): 283-304.
- Iba W, Langley P. 2011. 11 cobweb models of categorization and probabilistic concept formation. *Formal approaches in categorization*, 253.
- Ilango M, Mohan V. 2010. A survey of grid based clustering algorithms. *Int J Eng Sci and Tech*, 2(8): 3441-3446.
- Issaq HJ, Veenstra TD. 2019. Proteomic and metabolomic approaches to biomarker discovery: Academic Press.
- Jain AK, Dubes RC. 1988. Algorithms for clustering data. Englewood Cliffs, New Jersey: Prentice-Hall.
- Jain AK, Maheswari S. 2012. Survey of recent clustering techniques in data mining. *Int J Comput Sci Manag Res*, 3(2): 68-75.
- Jain AK, Murty MN, Flynn PJ. 1999. Data clustering: a review. *ACM Comp Surv CSUR*, 31(3): 264-323.
- Jiang C, Pan X, Gu M. 1994. The use of mixture models to detect effects of major genes on quantitative characters in a plant breeding experiment. *Genetics*, 136(1): 383-394.
- Joshi A, Kaur R. 2013. A review: Comparative study of various clustering techniques in data mining. *Int J Adv Res in Comp Sci and Softw Eng*, 3(3).
- Kailing K, Kriegel HP, Kröger P. 2004. Density-connected subspace clustering for high-dimensional data. Proceedings of the 2004 SIAM International Conference On Data Mining.
- Karypis G, Han EH, Kumar V. 1999. Chameleon: Hierarchical clustering using dynamic modeling. *Comp*, 32(8): 68-75.
- Kaufman L, Rousseeuw PJ. 2009. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons.
- Kaufmann L. 1987. Clustering by Means of medoids. Proc. Statistical Data Analysis Based on the L1 Norm Conference, Neuchatel, 1987, 405-416.
- Kaur PJ. 2015. A survey of clustering techniques and algorithms. Paper presented at the 2015 2nd international conference on computing for sustainable global development (INDIACom).
- Khatami A, Mirghasemi S, Khosravi A, Nahavandi S. 2015. A new color space based on k-medoids clustering for fire detection. IEEE International Conference on Systems, Man, and Cybernetics.
- Knuth DE. 1980. Algorithms in Modern Mathematics and Computer Science.
- Knuth DE. 2014. Art of computer programming, volume 2: Seminumerical algorithms: Addison-Wesley Professional.
- Kohonen T. 1982. Self-organized formation of topologically correct feature maps. *Biol Cyber*, 43(1): 59-69.
- Kohonen T. 1990. The self-organizing map. Proceedings of the IEEE, 78(9): 1464-1480.

- Kokate U, Deshpande A, Mahalle P, Patil P. 2018. Data stream clustering techniques, applications, and models: comparative analysis and discussion. *Big Data Cogn Comput*, 2(4): 32.
- Kothari R, Pitts D. 1999. On finding the number of clusters. *Pattern Recog Lett*, 20(4): 405-416.
- Kotsiantis S, Pintelas P. 2004. Recent advances in clustering: A brief survey. *WSEAS Transactions on Information Sci and App*, 1(1): 73-81.
- Kriegel HP, Zimek A. 2010. Subspace clustering, ensemble clustering, alternative clustering, multiview clustering: what can we learn from each other. Paper presented at the Proceedings of the 1st international workshop on discovering, summarizing and using multiple clusterings MultiClust held in conjunction with KDD.
- Kriegel HP, Kröger P, Sander J, Zimek A. 2011. Density-based clustering. *Wiley Interdis Rev: Data Mining Know Disc*, 1(3): 231-240.
- Lai C, Nguyen NT. 2004. Predicting density-based spatial clusters over time. Fourth IEEE International Conference on Data Mining ICDM'04.
- Leung Y, Zhang JS, Xu ZB. 2000. Clustering by scale-space filtering. *IEEE trans*, 22(12): 1396-1410.
- Lu B, Charlton M, Brunson C, Harris P. 2016. The Minkowski approach for choosing the distance metric in geographically weighted regression. *Inter J Geo Inf Sci*, 30(2): 351-368.
- Mahi H, Farhi N, Labed K. 2015. Remotely sensed data clustering using K-harmonic means algorithm and cluster validity index. *IFIP International Conference on Computer Science and its Applications*.
- Malkauthekar M. 2013. Analysis of Euclidean distance and Manhattan distance measure in Face recognition.
- Mallick P, Ghosh O, Seth P, Ghosh A. 2019. Kohonen's Self-organizing Map Optimizing Prediction of Gene Dependency for Cancer Mediating Biomarkers. *merging Technologies in Data Mining and Information Security*, 863-870: Springer.
- McLachlan GJ, Krishnan T. 2007. The EM algorithm and extensions. John Wiley & Sons.
- Mehrotra K, Mohan CK, Ranka S. 1997. Elements of artificial neural networks: MIT press.
- Melnykov V, Michael S, Melnykov I. 2015. Recent developments in model-based clustering with applications. *Partitioning clustering algorithms* s. 1-39: Springer.
- Merigo JM, Casanovas M. 2011. A new Minkowski distance based on induced aggregation operators. *Int J Comput Intel Sys*, 4(2): 123-133.
- Mesquita DP, Gomes JP, Junior AHS, Nobre JS. 2017. Euclidean distance estimation in incomplete datasets. *Neurocomp*, 248: 11-18.
- Metcalf L, Casey W. 2016. *Cybersecurity and Applied Mathematics*: Syngress.
- MITS M. 2017. Mining of Images by K-Medoid Clustering Using Content Based Descriptors. *Int J Signal Proces, Image Proces and Pattern Recog*, 10(8): 135-144.
- Müldner T. 2003. An algorithm for explaining algorithms. Unpublished manuscript, Acadia University.
- Murtagh F, Contreras P. 2011. Methods of hierarchical clustering. *arXiv preprint arXiv:1105.0121*.
- Murtagh F. 1984. Counting dendrograms: a survey. *Discrete App Math*, 7(2): 191-199.
- Myatt GJ, Johnson WP. 2009. Making sense of data II: A practical guide to data visualization, advanced data mining methods, and applications: Wiley Online Library.
- Na S, Xumin L, Yong G. 2010. Research on k-means clustering algorithm: An improved k-means clustering algorithm. Paper presented at the 2010 Third International Symposium on intelligent information technology and security informatics.
- Nakamura E, Kehtarnavaz N. 1998. Determining number of clusters and prototype locations via multi-scale clustering. *Pattern Recog Lett*, 19(14): 1265-1283.
- Nazeer KA, Sebastian M. 2009. Improving the Accuracy and Efficiency of the k-means Clustering Algorithm. Paper presented at the Proceedings of the world congress on engineering.
- Nielsen F. 2016. Hierarchical clustering. *Introduction to HPC with MPI for Data Science* s. 195-211: Springer.
- Pandya AS, Macy RB. 1995. *Pattern recognition with neural networks in C++*: CRC press.
- Parikh M, Varma T. 2014. Survey on different grid based clustering algorithms. *International Journal of Advance Research in Computer Science and Management Studies*, 2(2).
- Park HS, Lee JS, Jun CH. 2006. A K-means-like Algorithm for K-medoids Clustering and Its Performance. *Proceedings of ICCIE*, 102-117.
- Polycarpou MM, Helmicki AJ. 1995. Automated fault detection and accommodation: a learning systems approach. *IEEE Trans*, 25(11): 1447-1458.
- Pujari AK. 2001. *Data mining techniques*: Universities press.
- Rao C, Gudivada VN. 2018. *Computational Analysis and Understanding of Natural Languages: Principles, Methods and Applications*. Elsevier.
- Rao VS, Vidyavathi DS. 2010. Comparative investigations and performance analysis of FCM and MFPCM algorithms on iris data. *Indian J Comp Sci and Eng*, 1(2): 145-151.
- Reddy BO, Usseinaiah DM. 2012. Literature survey on clustering techniques. *IOSR J Comp Eng*, 3: 01-12.
- Reddy M, Makara V, Satish R. 2017. Divisive Hierarchical Clustering with K-means and Agglomerative Hierarchical Clustering. *Int J Comp Sci Trends Technol*, 5(5): 6-11.
- Redner RA, Walker HF. 1984. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev*, 26(2): 195-239.
- Rezende HR, Esmin AAA. 2010. Proposed application of data mining techniques for clustering software projects. *INFOCOMP J Comp Sci*, 9(6): 43-48.
- Roberts SJ. 1997. Parametric and non-parametric unsupervised cluster analysis. *Pattern Recog*, 30(2): 261-272.
- Rodriguez A, Laio A. 2014. Clustering by fast search and find of density peaks. *Sci*, 344 6191: 1492-1496.
- Rokach L, Maimon O. 2005. Clustering methods. *Data mining and knowledge discovery handbook*. 321-352. Springer.
- Romesburg C. 2004. *Cluster analysis for researchers*: Lulu. com.
- Ruspini EH. 1969. A new approach to clustering. *Inf Cont*, 15(1): 22-32.
- Ruspini EHJIS. 1970. Numerical methods for fuzzy clustering. *Inform Sci*, 2(3): 319-350.
- Sathya R, Abraham A. 2013. Comparison of supervised and unsupervised learning algorithms for pattern classification. *Int J Adv Res in Artif Intel*, 2(2): 34-38.
- Schmee J, Hahn GJ. 1979. A simple method for regression analysis with censored data. *Technometrics*, 21(4): 417-432.
- Shah M, Nair S. 2015. A survey of data mining clustering algorithms. *Int J Comp App*, 128(1): 1-5.
- Sheikholeslami G, Chatterjee S, Zhang A. 1998. Wavecluster: A multi-resolution clustering approach for very large spatial databases. *Proceedings the VLDB*, 98.
- Sheikholeslami G, Chatterjee S, Zhang A. 2000. WaveCluster: a wavelet-based clustering approach for spatial data in very large databases. *The VLDB J*, 8(3-4): 289-304.
- Sheikholeslami G, Zhang A. 1997. Approach to clustering large visual databases using wavelet transform. *Visual Data Exploration and Analysis IV*.

- Sheikholeslami G, Zhang A, Bian L. 1997. Geographical image classification and retrieval. 5th ACM international workshop on Advances in geographic information systems.
- Singh M. 2015. A survey on various k-means algorithms for clustering. *Inter J Comput Sci Network Sec* 15(6): 60.
- Smith JR, Chang SF. 1994. Transform features for texture classification and discrimination in large image databases. *Proceedings of 1st International Conference on Image Processing*.
- Sneath PH, Sokal RR. 1973. Numerical taxonomy. The principles and practice of numerical classification.
- Sun S, Qin K. 2007. Research on Modified k-means Data Cluster Algorithm. *Fine Particles, Thin Films and Exchange Anisotropy. Comp Eng*, 7(6): 200-202.
- Swarndeept Saket J, Pandya S. 2016. Implementation of Extended K-Medoids Algorithms to Increase Efficiency and Scalability using Large Dataset. *Int J Comp App*, 975: 8887.
- Szabo F. 2015. The linear algebra survival guide: illustrated with Mathematica: Academic Press.
- Thakare YBagal SJIJoCA. 2015. Performance evaluation of K-means clustering algorithm with various distance metrics. *IEEE Trans*, 110(11): 12-16.
- Theodorakis M, Vlachos A, Kalamboukis TZ. 2004. Using hierarchical clustering to enhance classification accuracy. Paper presented at the Proceedings of 3rd Hellenic Conference in Artificial Intelligence, Samos.
- Thilagavathi G, Srivaishnavi D, Aparna N. 2013. A survey on efficient hierarchical algorithm used in clustering. *Int J Eng*, 2(9): 165-176.
- Tian Y, Liu D, Qi H. 2009. K-harmonic means data clustering with differential evolution. *International Conference on Future BioMedical Information Engineering FBIE*.
- Turi RH. 2001. Clustering-based colour image segmentation: Monash University PhD thesis.
- Tyagi A, Sharma S. 2012. Implementation Of ROCK Clustering Algorithm For The Optimization Of Query Searching Time. *Int J Comp Sci and Eng*, 4(5): 809.
- Uppada SK. 2014. Centroid based clustering algorithms—A clarion study. *Int J Comp Sci Inf Tech*, 5(6): 7309-7313.
- Veenman CJ, Reinders MJT, Backer E. 2002. A maximum variance cluster algorithm. *IEEE Trans*, 24(9): 1273-1280.
- Vimal A, Valluri SR, Karlapalem K. 2008. An Experiment with Distance Measures for Clustering. *Proceedings COMAD 2008*.
- Wang W, Yang J, Muntz R. 1997. STING: A statistical information grid approach to spatial data mining. *Proceedings VLDB '97*.
- Weinstein E. 2006. Expectation maximization algorithm and applications. *Courant Institute of Mathematical Sciences*.
- Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Philip SY. 2008. Top 10 algorithms in data mining. *Know Inf Sys*, 14(1): 1-37.
- Xia Y, Xi B. 2007. Conceptual clustering categorical data with uncertainty. 19th IEEE International Conference on Tools with Artificial Intelligence ICTAI 2007.
- Xu D, Tian Y. 2015. A comprehensive survey of clustering algorithms. *Annals Data Sci*, 2(2): 165-193.
- Xu R, Wunsch DC. 2005. Survey of clustering algorithms.
- Zhang B, Hsu M, Dayal U. 2000. K-harmonic means-a spatial clustering algorithm with boosting. *International Workshop on Temporal, Spatial, and Spatio-Temporal Data Mining*.
- Zhang B, Hsu M, Dayal U. 1999. K-harmonic means-a data clustering algorithm. *Hewlett-Packard Labs Technical Report HPL 1999-124*, 1999, 55.
- Zhang T, Ramakrishnan R, Livny M. 1996. BIRCH: an efficient data clustering method for very large databases. *ACM Sigmod Rec*, 25(2): 103-114.
- Zhang X, Wang J, Wu F, Fan Z, Li X. 2006. A novel spatial clustering with obstacles constraints based on genetic algorithms and K-medoids. *Sixth International Conference on Intelligent Systems Design and Applications*.



Derleme (Review)

Cilt 3 - Sayı 4: 190-197 / Ekim 2020
(Volume 3 - Issue 4: 190-197 / October 2020)

YAPAY SİNİR AĞLARI İLE DOĞAL GAZ TÜKETİM TAHMİNİNİ ÜZERİNE YAPILMIŞ ÇALIŞMALARIN İRDELENMESİ

Bedia KANT¹, Mehmet Serhat ODABAŞ^{1*}

¹Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Akıllı Sistemler Mühendisliği, Samsun, Türkiye

Gönderi: 21 Mayıs 2020; **Kabul:** 17 Haziran 2020; **Yayınlanma:** 01 Ekim 2020
(Received: May 21, 2020; **Accepted:** June 17, 2020; **Published:** October 01, 2020)

Özet

Talep tahmini tüketicinin gelecek zamanda ne kadar mal ve hizmet talebinde bulunacağını değişkenler yardımıyla tahmin edilmesidir. Bu çalışmanın amacı, literatürde yapay sinir ağları kullanılarak doğal gaz sektöründe talep tahmin çalışmalarının nasıl yapıldığını irdelemektir. Bilindiği üzere doğal gaz temini açısından dışa bağımlı bir politika izlemektedir. Günümüzde doğal gaz enerji kaynakları arasında hem stratejik hem de ticari avantajları nedeniyle önemli bir konuma sahiptir. Doğal gaz sektöründe talep tahminin yapılması hem yatırım planlaması hem de ekonomik gibi stratejik açıdan oldukça fazla önem arz etmektedir. Doğal gaz açısından dış ülkelere bağılı olunması, alınacak gaz miktarının ve fiyatların belirlenmesi açısından önemlidir.

Anahtar kelimeler: Yapay sinir ağları, Talep tahmini, Doğal gaz


Investigation of Studies on Natural Gas Consumption Forecasting by Artificial Neural Networks


Abstract: Demand forecasting is the estimation of how much goods and services the consumer will request in the future. The purpose of this study is to investigate how demand forecasting Works are done in the natural gas sector by using artificial neural Networks in the literature. As it is known, we follow a foreign-dependent policy in terms of natural gas supply. Today, it has an important position among natural gas anergy sources due to its strategic and commercial advantages. Demand forecasting in the natural gas sector is of great strategic importance, both in terms of investment planning and economics. In terms of natural gas, adhering to foreign countries is important in terms of determining the amount of gas to be bought and prices.

Keywords: Artificial neural networks, Demand forecast, Natural gas

*Corresponding author: Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Akıllı Sistemler Mühendisliği, Samsun, Türkiye

E mail: mserhat@omu.edu.tr (M. S. ODABAŞ)

Bedia KANT  <https://orcid.org/0000-0001-6137-9518>

Mehmet Serhat ODABAŞ  <https://orcid.org/0000-0002-1863-7566>

Cite as: Kant B, Odabaş MS. 2020. Investigation of studies on natural gas consumption forecasting by artificial neural networks. BSJ Eng Sci, 3(4): 190-197.

1. Giriş

Yer kabuğunun içinde bulunan doğal gaz karbon temelli bir fosil yakıttır. Günümüzde petrolden sonra ekonominin en değerli ve en stratejik hammaddesinden birisi haline gelmiştir. Doğal gaz arzında herhangi bir aksama yaşandığı zaman ekonominin olumsuz etkilendiğini görmekteyiz. Petrolün çevreye verdiği zarar ve pazarındaki düzensizlikler sebebiyle, doğal gaz petrole alternatif kaynaklar arasında ilk sırada yer almaktadır. Çevreyi kirletmemesi, havadan daha hafif olması ve taşıma kolaylığı ile doğal gaza olan talep her geçen gün artmaktadır. Evlerde ve endüstride karbon ve kükürt oksit gibi havayı kirleten atık maddeleri açığa çıkarmadığı için daha çok tercih edilmektedir. Daha çok tercih edilmesinin bir diğer sebebi de kalorifer yakıtı olarak kullanılmak istendiği zaman herhangi bir ön hazırlık ve depolama gerektirmemesidir (Kuru, 2014).

Doğal gaz Türkiye 'de ilk defa Kırklareli'nde 1970 yılında keşfedilmiştir. Pınarhisar Çimento fabrikasında 1976 yılında tüketilen doğal gaz, 1948'de SSCB ile yapılan doğal gaz ithalatı anlaşması sonucu enerji ihtiyacımızın önemli bir kısmını karşılıyor hale gelmiştir. Doğal gaz piyasası oluşturulması yolunda 2001 yılında 4646 sayılı Doğal Gaz Piyasası Kanunu oluşturulmuş, iletim, depolama, ithalat, ihracat, toptan satış, gibi faaliyetlerin lisansının alınması zorunlu hale getirilmiştir.

Günümüzde ülkemizde az üretilen ve %99'unu dışarıdan ithal ettiğimiz doğal gazın tüketiminin tahmin edilmesi son derece önemli bir noktadır. Talep tahminin doğru yapılması, yurtdışından gaz alımı ile ilgili anlaşmaları, sektöre yönelik yapılacak yatırımları ve buna bağlı olarak sektörün gelişimini etkileyecek etkenlerden birini oluşturur. Bilindiği üzere doğal gaz tüketiminin doğru tahmin edilmesi hem gaz dağıtıcıları hem de tüketicileri açısından önemlidir. Doğal gaz dağıtıcısı açısından incelendiğinde doğru tahmin değerleri sistemde oluşabilecek hataları azaltır ve gaz dağıtım limitlerinin doğru şekilde düzenlenmesini sağlar. Talep tahmini ne kadar doğru yapılırsa, planlama çalışmaları da olumlu sonuç gösterecektir. Böylece gaz sistemleri çok daha tutarlı bir hale gelir. Ayrıca tüketici açısından bakıldığı zaman doğru tahmin değerleri tüketim miktarını azaltır. Talep tahmin çalışmaları birçok farklı yöntem kullanılarak yapılabilir. Bu yöntemler arasında son zamanlarda sıklıkla kullanılan yapay sinir ağları yöntemi yer almaktadır. YSA, insan beyninin karar verme ve işlem yapma özelliklerinden esinlenilerek oluşturulmuştur. Yapay sinir ağları yöntemi ile geçmiş dönemdeki örnekler kullanılarak genel yargılar veya sonuçlar üretilmeye çalışılır. Örnekler veya eldeki veriler kullanılarak ağı öğrenme işlemi tamamlanmış olur. Daha sonraki aşamalarda ağ hiç görmediği örnekler hakkında edindiği bilgileri kullanarak sonuçlar üretebilir.

Doğal gaz tüketiminin tahmini son derece önemlidir. Bu derlemede yapay sinir ağlarının açıklanması, ağ elamanlarından bahsedilerek talep tahmini ile ilişkisinden bahsedilmesi ve doğal gaz tüketiminde YSA'nın nasıl

kullanıldığının irdelenmesi amaçlanmıştır. Yapay sinir ağları yöntemi sıklıkla kullanılan yöntemlerden daha farklıdır. Belirsizlikler altında karar verebilen, hata toleransı yüksek olan YSA yöntemini bu derlemede incelenecektir.

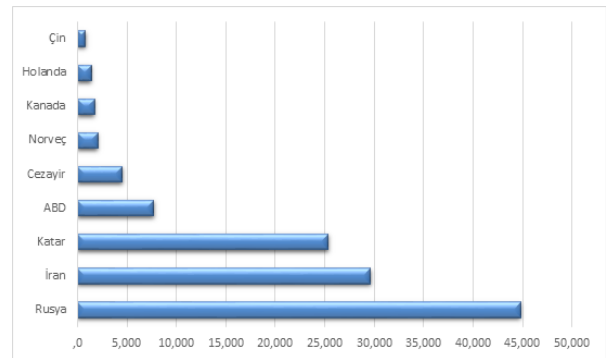
1.1. Doğal Gaz Piyasası

Günümüzde doğal gaz enerji kaynakları arasında hem stratejik hem de ticari avantajları nedeniyle önemli bir konuma sahiptir. Ülkeler bazında doğal gaz piyasası pazarlama yöntemlerini ve kaynak yönetimini kapsayan uluslararası bir piyasadır. Ülkeler arasında doğal gaz arz ve talebini oluşturan sözleşmeler olduğu gibi her ülkenin kendi bünyesinde doğal gaz ticaretini ve dağıtımını belirleyen düzenlemeleri bulunmaktadır (Taşkın, 2018).

Dünya'da birçok ülke ilk petrol krizinin ardından doğal gaz sektöründe yaşanan talep patlaması sebebi ile doğal gaz enerjisi üzerine yoğunlaşmışlardır. Üretim ve tüketim adına birçok plan ve politika arayışına girilmesi doğrultusunda doğal gaz enerjisi bu zamana kadar gelişimini sürdürmüş ve sürdürmeye devam etmektedir (Oruç, 2017).

19. yüzyılın ortalarından itibaren ABD'de enerji kaynağı olarak kullanılan doğal gaz, 40-50 sene önce gereksiz sayılıp dünyada geniş bir kullanımı yoktu. Ancak 1973 yılında gerçekleşen ilk doğal gaz krizi sonrası önem kazanmıştır. Doğal gaz günümüzde 70'ten fazla ülkede üretilmektedir (Mazak, 2004).

Dünyanın en büyük doğal gaz rezervine sahip Rusya 2011 yılı BP istatistiki verilerine göre 44,8 trilyon metreküp kanıtlanmış rezerve sahiptir. Bu oran bütün dünya rezervlerinin %24'üne denk gelmektedir. İran Rusya'dan sonra Dünyanın en büyük ikinci doğal gaz rezervine sahip ülkesidir. 29 trilyon 610 milyar metreküp kanıtlanmış rezerve sahip olan İran'ın doğal gaz ihracatı dünyadaki ihracat oranının %1'ini aşmamaktadır. Cezayir 4 trilyon 502 milyar metreküp ile kanıtlanmış doğal gaz rezervine sahiptir (Şekil 1).



Şekil 1. Dünya doğalgaz rezerv miktarları m³.

Avrupa Birliğinin ihtiyacının %25'ini Cezayir karşılamaktadır. Hollanda 1 trilyon 387 milyar metreküp kanıtlanmış rezerve sahiptir. Hollanda doğal gaz üretiminin büyük bir kısmını kıyı kuyularından elde etmektedir. İn 800 milyar metreküp kanıtlanmış rezerve sahiptir. Çin hükümeti doğal gazın bir kısmını ithal

ederken kaya gazı üretimi için gereken yasal işlemleri tamamlamıştır. Norveç 2 trilyon 39 milyar metreküp kanıtlanmış rezerve sahiptir. Kendi ihtiyacının 16 katı fazla üretim yapan Norveç arta kalan gazın tamamını Avrupa'ya ithal etmektedir (Salih,2012).

Dünya enerji tüketimi 2016 yılında 13258 mtpe iken 2017 yılında 13511 mtpe ulaşmıştır. Bunun %34,2'si petrol, %23,4'ü doğal gaz, %27,6'sı kömür olmak üzere %85,2'si fosil kaynaklardan, %4,4'ü nükleer enerji, %6,8'i hidrolik, % 3,6'sı yenilebilir enerjiden sağlanmaktadır. 85 mtpe ile 2017 yılında 2016 ya göre artıştaki en büyük katkıyı doğal gaz yapmıştır (Wec,2019).

Türkiye'de 1986 yılında doğal gaz taşımacılığının ve ticaretinin ilk adımı BOTAŞ tarafından SSCB şirketi ile 25 yıllık yapılan gaz alımı anlaşmasıyla atılmıştır. İlk doğal gaz ithalatı fiili olarak 1987 yılında gerçekleşmiştir.

Cezayir ile 1988 yılında LNG alım anlaşması imzalanmış olup, Marmara Ereğlisi LNG Terminali 1994 yılında devreye girmiştir. 1995 yılında ikinci LNG alım anlaşması Nijerya ile imzalanmıştır. Ülkeye yeni arz kaynakları kazandırmak adına 1996 yılında İran ile anlaşma imzalanmıştır. Mavi Akım Hattı için 1997 yılında, ilk doğal gaz ithalatı yapılan Batı Hattı için 1998 yılında Rusya ile anlaşma imzalanmıştır. Arz kaynakları genişletilerek 2001 yılında Azerbaycan ile doğal gaz alım anlaşması imzalanmıştır.

4646 sayılı Doğal Gaz Piyasası Kanunu Resmi Gazete 'de 2001 yılında yayımlanıp Türkiye'nin doğal gaz piyasasının liberalleşmesi adına ciddi bir adım atılmıştır. BOTAŞ'ın piyasadaki payının azaltılarak piyasanın serbestleşmesi adına bir diğer adım atılmıştır (Yıkılmaz, 2015).

Türkiye'de 2018 yılı itibariyle 49.328,93 milyon m³ doğal gaz tüketimi gerçekleşmiştir. Bu tüketimin 428,17 milyon m³ lük kısmı ülkemizde bulunan rezervlerden karşılanmıştır.

2. Doğal Gaz Tahminin Literatür İncelenmesi

Doğal gaz sektöründe talep tahminin yapılması hem yatırım planlaması hem de ekonomik gibi stratejik açıdan oldukça fazla önem arz etmektedir. Doğal gaz açısından dış ülkelere bağlı olunması, alınacak gaz miktarının ve fiyatların belirlenmesi açısından önemlidir. Literatürde doğal gazın tahmin edilmesine dair çalışmalar farklı yöntemlerle yer almaktadır.

Khotanzad ve Elragal 2000 yılında "Combination of artificial neural-network forecasters for prediction of natural gas consumption" adlı çalışma yapmışlardır. Bu çalışmanın amacı, gaz tesislerinin ihtiyaç duydukları günlük tüketim miktarının yapay sinir ağları kullanarak öngörülmesidir. YSA ile önceki gaz tüketimi ve hava parametreleri arasındaki karmaşık ilişki gelecekteki tüketim ile modellenilebilir. İlk aşamada çok katmanlı ileri beslemeli YSA ve fonksiyonel bağlantı YSA içeren iki aşamalı bir sistem önermişlerdir. Başlangıçta yapay sinir ağı geri yayılım algoritması ile eğitilir ve tahmin sırasında

ağırlıklarını ayarlayabilmek için strateji kullanırlar. İkinci aşama, birinci aşamada üretilen iki farklı tahminin karşılaştırılmasından oluşmaktadır. En küçük kareler, ortalama, bulanık mantık, ileri beslemeli YSA, fonksiyonel bağlantı YSA, Karmarkar'ın doğrusal programlama algoritması (1984), sıcaklık uzay yaklaşımı ve yerel uzmanların uyarlanabilir karışımından oluşan sekiz farklı kombinasyon algoritmasını incelemişlerdir. Altı farklı gaz tesisinden alınan gerçek veriler üzerinde test etmişlerdir. Yaptıkları analizler sonucunda tek bir YSA modeline bağlı kombinasyon stratejilerinin diğer yaklaşımlara göre daha iyi sonuç verdiğini gözlemlemişlerdir (Khotanzad ve ark., 2000).

Kızılaslan 2008 yılında " Forecasting of short term and mid term İstanbul natural gas consumption values by neural Networks algorithm" adlı yüksek lisans çalışmasını ele almıştır. Bu çalışmanın amacı, yapay sinir ağı algoritmalarını kullanarak saatlik, günlük, haftalık ve aylık doğal gaz tüketim tahmininin bulunmasıdır. Çalışmada ilk olarak doğal gaz tüketimine etki eden faktörleri belirlemek amacıyla analiz yapmıştır. Elde ettiği parametreleri geri beslemeli yapay sinir ağları modeline uygulamıştır. Bu yöntemi kullanmasının amacı YSA algoritmalarının birden fazla tahmini aynı anda yapabilmesi ve doğrusal olmayan verilerin modellenmesinde iyi sonuçlar alabilmesidir. Yedi farklı algoritma kullanarak bu algoritmaların karşılaştırmasını yapmıştır. Tüm modeller için uygunluk ölçütlerini 200 kere tekrar etmiştir (Kızılaslan,2008).

Kaynar ve diğerleri "Yapay Sinir Ağlarıyla Doğal Gaz Tüketim Tahmini" adlı makalelerinde doğal gaz bakımından dışa bağlılığımızdan bahsedip tahmin edilmesinin ne kadar önemli olduğunu vurgulamışlardır. Tahmin yöntemlerinden YSA ve zaman serileri analizi yönteminden ARIMA modelini kullanarak sonuçlar elde etmiş ve tahminin hangisinde daha iyi sonuç verdiğini karşılaştırmışlardır. Çalışmada Ankara iline ait Ocak 2005 Haziran 2006 arasındaki günlük ve haftalık doğal gaz arz verilerini kullanılmıştır. Çalışmada %80'i eğitim %20'si doğrulama verisi olarak kullanılmıştır. YSA modelinde MATLAB programı kullanılarak bir model geliştirilmiştir. MSE performans ölçütü kullanılarak sonuçlar kıyaslandığında yapay sinir ağları yönteminin daha iyi sonuçlar verdiği görülmüştür (Kaynar ve ark., 2011).

Demirceylan 2012 yılında ele aldığı " Erzurum doğal gaz tüketim miktarının yapay sinir ağları ile çoklu doğal regresyon analizi sonuçlarını karşılaştırmıştır. Doğal gaz tüketimini etkileyen verileri kullanarak analizi gerçekleştirmiştir. Mevsimsel etkiyi incelemek için aylar, aylık ortalama sıcaklık, mevsimsellik ve aylık tüketim miktarı girdi olarak ağda kullanılmıştır. Aylık tüketim miktarı ağına çıktı katmanıdır. Çoklu regresyon yönteminden aldığı sonuç ile yapay sinir ağlarından aldığı sonucu karşılaştırmıştır. Yapay sinir ağları hata oranı daha düşük sonuç vermiştir. Çoklu regresyon yönteminde 30 ayın verileri kullanılarak sadece gelecekteki 30 ay için tahmin yapılabilir. Ancak yapay sinir ağları yönteminde ağı eğittikten sonra her değer için tahmin

yapılabilmektedir (Demirceylan, 2012).

Akgül ve diğerleri “ Doğal Gaz Tüketim Tahmini” adlı makalesinde Dünyada ve Türkiye’de doğal gaz kullanımından bahsetmiştir. TUİK’ ten elde ettiği verilerle yapay sinir ağları ve klasik zaman serileri (ARIMA) yöntemlerini kullanmıştır. Eğitim ve test verileri için ek küçük MSE (Mean Squared Error) değerine göre en iyi modeller seçilmiştir. Yapılan analizler onucunda yapay sinir ağları yönteminin daha iyi sonuç verdiği görülmüştür (Akgül, 2013).

Rodger 2014 yılında ele aldığı makalesinde enerji maliyet tasarrufunu sağlamak amacıyla doğal gaz talebini tahmin etmeyi amaçlamıştır. Bir kablosuz sensör sisteminin günlük olarak havalandırma, ısıtma ve klima sistemlerini kontrol etmek amacıyla kullanılan gerekli doğal gaz talebinin tahmin edilmesi için YSA modellemiştir. Sensör sistemine ait 111 günlük tüketim verileri toplayıp, giriş çıkış verilerini ağı eğitmek için kullanmıştır. Dört bileşenden oluşan bulanık en yakın komşu sinir ağı istatistik modelini kullanmıştır. Bununla birlikte doğal gaz talebini optimize etmek amacıyla işletme giderleri, fiyat, yeni kuyu açma maliyeti, gazı açma maliyeti, petrol fiyatı verilerini kullanarak en yakın komşu YSA model maliyet fonksiyonunu kullanmıştır (Rodger, 2014).

Kuru 2014 yılında ele aldığı “Tümevarım Yöntemi Kullanarak Türkiye Geneli Toplam Doğal Gaz Tüketimi Analizi” adlı çalışmada enerji kaynaklarının üretim ve tüketim şekillerine ait yakın dönem verilerini kullanarak yenilebilir enerji ve mevcut fosil kaynakları için gelecekte ne olacağına dair araştırma yapmıştır. Doğal gaz tüketim tahminini tümevarım ve tümdengelim yöntemleri ile yapmıştır. 2013 yılı tüketimi ele alınarak analitik model oluşturmuştur. Yaptığı çalışma sonucunda tümevarıma dayalı olarak geliştirdiği model, tümdengelim yöntemine göre üstünlük sağlamış, 2013 yılı doğal gaz tahmini daha az hata ile bulunmuştur (Kuru, 2014).

Szoplik 2015 yılında ele aldığı “Forecasting of natural gas consumption with artificial neural network” adlı makalesinde polonya’ya ait gerçek doğal gaz tüketim verilerini kullanarak talep tahmin çalışması yapmıştır. MLP modelinin tasarımında küçük sanayi ve bireysel tüketiciler tarafından gaz kullanımını etkileyen hava durumu ve takvim faktörlerini göz önünde bulundurmıştır. Oluşturulan modelin sonuçlarını karşılaştırmıştır. MLP modeli ile yılın herhangi bir günü ve herhangi bir saatindeki gaz tüketiminin başarılı bir şekilde bulunabileceği tahmin edilmiştir (Szoplik,2015).

Taşkıner “Ankara İli Doğal Gaz Tüketiminin Yapay Sinir Ağları İle Öngörüsü” adlı çalışmada Ankara ili için doğal gaz tüketimini yapay sinir ağları yöntemini kullanarak tahmin etmiştir. Ağa girdi olarak ortalama sıcaklık tahmini, hissedilen maksimum sıcaklık tahmini, nem, ısı değer ve bağımsız birim sayısı verilmiştir. Çıktısında tüketim bulunmaktadır. 2014-2017 yılı arası veriler kullanılarak tahminleme yapmıştır (Taşkıner, 2018).

Öztürk ve diğerleri 2018 yılında ele aldığı “ Rüzgâr ve güneş santrallerinde kısa dönem enerji üretim tahmini

için matematiksel modellerin oluşturulması” isimli araştırma makalelerinde meteorolojik veriler ile doğru üretim tahmini yapan matematiksel modellerin bulunmasını amaçlamışlardır. MATLAB programı kullanılarak matematiksel modeller oluşturulmuştur. Rüzgâr ve güneş enerji santrallerinde kısa dönem enerji üretimi tahmini yapılmıştır. Geliştirilen matematiksel modellerin doğruluğu çoklu regresyon analizi kullanılarak incelenmiştir (Öztürk ve ark., 2018).

Meral 2019 yılında ele aldığı “Türkiye’deki Enerji Santrallerinde Doğal Gaz Tüketiminin Destek Vektör Regresyon İle Tahmini” adlı çalışmada doğal gaz tüketimini regresyon yöntemi kullanarak tahmin edilmiştir. Araştırmasında kullanacağı veriler 2013-2018 yılları arasında Enerji İşleri Genel Müdürlüğü ve Türkiye Enerji Piyasası Düzenleme Kurumu’ndan temin edilmiştir. Çalışmada ilk önce Türkiye’ de doğal gazın önemi, üretim, tüketim, ihracat ve ithalat miktarları değerlendirilmiştir. Bu değerlerin birim farklılıklarında dolayı veriler önce standartlaştırılmıştır. Çekirdek fonksiyon olarak Radyal Tabanlı Fonksiyon (RTF) seçilmiştir. Sonrasında destek vektörler, karar sabiti ve ağırlıklar belirlenmiştir. Destek vektörler ve ağırlıklar çarpılarak yan yana ekleyip modeli elde etmiştir (Meral, 2019).

Tuna 2019 yılında ele aldığı “ Doğal gaz talep tahmini: Erzurum ili üzerine bir uygulama” adlı çalışmada daha önce Erzurum iline ilişkin böyle bir çalışma yapılmadığından bahsetmiş, Anadolu’daki iklim değişikliğinin doğal gaz talebine olan etkisini görmek için tahmin yapmayı öngörmüştür. Ocak 2009- Kasım 2018 yıllarına ait konut sektörü aylık doğal gaz tüketim verileri kullanılmıştır. Verilerin mevsimsellik içerdiği tespit edilmiştir. Tahmin edilmesinde tek değişkenli zaman serisi yöntemlerinden olan Box Jenkins metodu ile ARIMA modeli kullanılmıştır. Aralık 2018-Aralık 2020 dönemine ait doğal gaz tüketim değerleri için öngörüle bulunulmuştur (Tuna, 2019).

3. Talep Tahmini

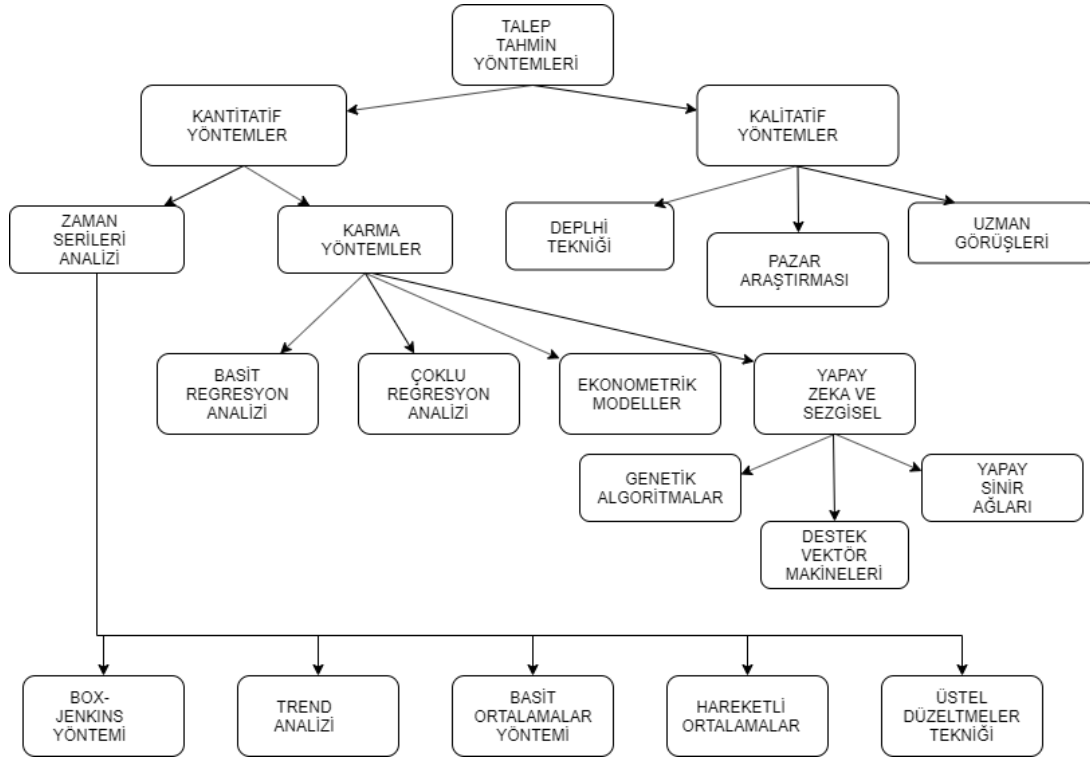
Tüketicilerin bir hizmeti ve ya ürünü belirli bir fiyattan almayı kabul ettikleri miktara talep denir. Gelecekteki bir zaman dilimi için firmanın bir ürünü veya farklı ürünleri için talep düzeyini tespit etmeye talep tahmini denir.

Müşteri talebinin tahmin edilmesini ne kadar ürünün, ne kadar işgücü ve hammaddeyle, ne kadar üretileceğinin planlanması, stok miktarının belirlenmesi açısından önemlidir. Talep tahmini büyük, orta, küçük çaplı işletmelerin, yatırımcıların tümünde kullanılan bir yöntemdir (Soysal, 2010).

Talep tahmini şirket yönetiminin taktik ve stratejik kararlar alınmasında, kısa ve uzun dönemli hedeflere ulaşmak için kullanılması gereken en temel araçlardan biridir. Doğru bir şekilde yapılmış talep tahmini ile şirket optimum kar seviyesine ulaşır. Yapılacak tahmin tipini ve süresini talep yöntemi ve ürünün özellikleri etkilemektedir. Ürüne olan talep miktarı sabit ise tahmin süresi kısaldır. Ancak ürüne ait talepte dalgalanmalar

mevcutsa tahminleme süresi uzayabilir. Bazı ürünler mevsimsel farklılıklar gösterdiği için talep değerleri azalabilir veya çoğalabilir. Bu değişimlerin sebebini

belirleyebilmek için mevsimsel tahmin yöntemleri kullanılmalıdır (Adıyaman, 2007). Talep tahmin yöntemlerine ait bir diyagram Şekil 2’de verilmiştir.



Şekil 2. Talep tahmin yöntemleri (Karaatlı, 2012).

Talep tahmin aşamaları:

- Talebi Etkileyen Etkenlerin Belirlenmesi: Talep tahmini yapılmadan önce şirketin ürettiği ürünler, şimdiki durumu ve gelecekte gelmek istediği durum, hangi çevre ortamında olduğu, fiyat talep ilişkisi, rakiplerinin durumu, iktisadi, endüstriyel, sosyal değişimler, teknolojik gelişim gibi talebi etkileyen etkenler ve buna bağlı olarak ağırlıkları belirlenmelidir.
- Verilerin Toplanması: Talebi etkileyen tüm veriler toplanır. Genellikle geçmiş gerçek veriler, hedef veriler ve istatistiksel olasılıklardan oluşur. Verilerin düzenlenmesinde şirketin yapmış olduğu anketler, yayınlanmış istatistik verileri, iç kaynaklar ve iktisadi veriler kullanılabilir.
- Talep Tahmin Periyodunun Tespiti: yapılacak tahmin ile kullanım süresi arasında yakın bir ilişki bulunmaktadır. Örneğin; aylık yapılan tahminlerin uzun dönemlerde kullanılması doğru sonuçlar vermeyebilir.
- Tahmin Yönteminin Seçimi: Talep tahmin yöntemleri kantitatif ve kalitatif olmak üzere ikiye ayrılır. Kantitatif yöntemler matematiksel modellemeye dayanırken, kalitatif yöntemler tecrübeye dayanmaktadır.
- Tahmin Sonuçlarının Geçerliliğinin Araştırılması: Sonuçların gerçek verilerle karşılaştırılarak doğruluğu tespit edilmelidir.

3.1. Yapay sinir ağları

Yapay sinir ağları insan beyninin en temel fonksiyonunu gerçekleştiren bilgisayar sistemidir. Öğrenme işlemi geçmiş veriler yardımıyla gerçekleştirilir. Bu ağlar birbirine bağlı yapay sinir hücrelerinden oluşmaktadır. Her hücrenin bir ağırlık değerleri saklı olup ağı yayılmıştır.

Yapay sinir ağları, basit işlem hücrelerinden oluşturulup bilgiyi depolama ve kullanma yeteneğine sahip olarak geliştirilen işlemcidir. İki yönden beyin ile benzerlik göstermektedir. Bunlardan birincisi; bilgiyi öğrenme yoluyla çevreden elde etmesi ikincisi ise; öğrenilen bilgiyi depolamak için sinaptik ağırlık olarak da bilinen nöronlar arasında bulunan bağlantıyı kullanmasıdır (Hatipoğlu, 2010).

Yapay sinir ağları diğer hesaplama yöntemlerinden farklılık göstermektedir. Buldukları ortama adapte olabilen, eksik verilerle çalışabilen, hatalara karşı toleranslı olabilen, belirsizlik altında karar verebilen bu ağlar hemen hemen her kullanım alanında başarılı uygulama örneklerine sahiptir. Ağ parametresinin seçiminde belli standardın olmaması, verilerin yalnızca numerik olarak ifade edilmesi, ağın davranışlarının tam olarak bilinmemesine rağmen yapay sinir ağlarına ilgi sürekli artmaktadır. Örüntü tanıma, sınıflandırma, veri sıkıştırma, optimizasyon ve sinyal filtreleme çalışmalarında yapay sinir ağları en sık kullanılan teknikler arasında yer alır. Bunlar haricinde; optimum

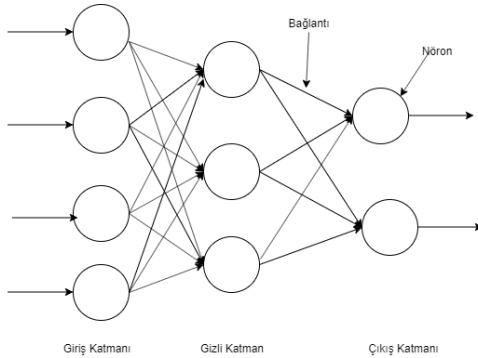
rota belirleme, malzeme analizi, parmak izi tanıma, kalite kontrol, veri madenciliği, iş çizelgeleme gibi konularda da örnekleri bulunmaktadır (Öztemel, 2012).

Yapay sinir ağları makine öğrenmesi alanında geliştirilen uygulamalardan birisidir. Örnek veriler yardımıyla genel yargılar ve sonuçlar üretmeye çalışırlar. Bilgisayar sistemindeki ağa uygulayıcı tarafından örnekler girilir ve ağın bu örnekler yardımıyla sonucu bulması sağlanır. Ağ öğrenimini tamamladıktan sonra hiç görmediği örnekler hakkında bilgi verebilir, sonuç üretebilir (Sönmez Çakır, 2019).

1970 yılının yapay sinir ağlarında bir dönüm noktası olarak ifade edilebilir. Bu tarihten önce birçok araştırma yapılsada 1969 yılında XOR probleminin çözülmemesi ile birlikte çalışmalara ara verildiği görülmektedir. 1970 yılından sonra XOR problemi birkaç araştırmacının çalışmasını sürdürmesi sonucu çözüme ulaşmıştır ve yapay sinir ağları konusu tekrardan ilgi görmeye başlamıştır. 10 yıl içerisinde birbirinden farklı 30 civarlarında model geliştirilmiştir. Bununla birlikte yapay sinir ağları laboratuvar ortamından çıkarak günlük hayatta da kullanılan sistem haline gelmiştir. Bu çalışma hem donanım teknolojisi hem de yapay zekâ ile desteklenerek herkesin kabul ettiği ve kullanmak istediği teknoloji haline gelmiştir (Öztemel, 2012).

3.2. Yapay sinir ağlarının temel elemanları

Yapay sinir hücreleri beynin sinir hücrelerine benzetilmektedir. Biyolojik sinir ağları sinir hücrelerinden oluşurken, yapay sinir ağları da yapay sinir hücrelerinden oluşmaktadır. YSA hücreleri süreç elemanları olarak da adlandırılmaktadır. Bir yapay sinir hücresinin girdiler, ağırlıklar, toplama fonksiyonu, aktivasyon fonksiyonu ve çıktılar olmak üzere 5 tane temel elemanı vardır (Demirceylan, 2012). Yapay sinir hücresinin yapısı Şekil 3'de gösterilmiştir.



Şekil 3. Yapay sinir ağı modeli.

3.2.1. Girdiler

Yapay sinir hücresine gelen bilgidir. Bu bilgiler diğer sinir hücresinden ya da dış ortamdan gelebilir (Demirceylan, 2012).

3.2.2. Ağırlıklar

Yapay sinir hücresine giren bilgilerin hangi ağırlıkta değerlendirilmesi gerektiğini belirleyen değerlerdir. Ağırlıkların negatif veya pozitif olması ağa etkisinin hangi yönde olacağını göstermektedir. Sinir hücresinin ağırlığı

sabit olacağı gibi değişken de olabilir (Yılmaz, 2010).

3.2.3. Toplama fonksiyonu

Birleştirme fonksiyonunun en yaygın kullanılan türü olan toplama fonksiyonu, biyolojik sinir hücresinde dentritlerin yaptığı görevi yapmaktadır. Sinir hücresine giren net girdileri hesaplamaktadır. Girdilerin ilgili ağırlıkla çarpımının toplamıdır. Net girdinin hesaplanma işlemidir. Bu fonksiyon aşağıdaki formülle gösterilmektedir.

$$\text{Net} = \sum G_i \cdot A_i \quad (n = 1, \dots, n)$$

Bu formülde G girdileri, A ağırlıkları, n girdi sayısını gösterir. Bir problemde optimal toplama fonksiyonunun bulunmasını sağlayacak bir formül yoktur. Her girdi aynı toplama fonksiyonuna sahip olabileceği gibi farklı toplama fonksiyonuna da sahip olabilmektedir. Aşağıdaki örnekler de toplama fonksiyonuna örnek olarak verilebilir:

$$\text{Maksimum: Net} = \text{Max} (G_i \cdot A_i)$$

$$\text{Minimum: Net} = \text{Min} (G_i \cdot A_i)$$

$$\text{Çarpım: Net} = \sum G_i \cdot A_i$$

$$\text{Kümülatif Toplam: Net} = \text{NET (eski)} + \sum (G_i \cdot A_i)$$

$$\text{Çoğunluk: Net} = \sum \text{sgn} (G_i \cdot A_i) \quad (\text{Ballı, 2014}).$$

3.2.4. Aktivasyon fonksiyonu

Hücre davranışını belirleyen bir diğer fonksiyonlardan birisi de aktivasyon fonksiyonudur. Toplam fonksiyonunun sonuçları ile birlikte çıktıya dönüştürülür. Sinir hücresine gelen girdinin işlenerek ne kadar çıktı üreteceğini belirlemektedir. Aktivasyon fonksiyonunun amacı, çıktıların büyük değerlere ulaşmasını engellemektir. Çıktı yüksek çıkarsa ağı eğitimi engellenmiş olmaktadır. Bir yapay sinir ağında aktivasyon fonksiyonu nöronun çıkış değerini [0,1] veya [-1,1] arasında sınırlamaktadır. Lineer, Logaritmik Sigmoid ve Hiperbolik Tanjant modelleri en sık kullanılan aktivasyon modelleridir.

3.2.5. Çıktılar

Çıktılar en uç katman olarak bilinmektedir. Ağların sonuçlarıdır ve birden fazla girdiye sahip olabilmektedir. Gizli katmandan veriyi alıp işler ve çıktısını vermektedir. Çıktı katmanındaki sinir hücre sayısı, yapay sinir ağına sunulan her verinin çıkış sayısı kadardır. Aktivasyon fonksiyonu tarafından çıktı oluşturulmaktadır. Her sinir ağının sadece bir çıktı düzeyi bulunmaktadır (Oğuz, 2019).

3.3. Talep tahmini ve yapay sinir ağları

Birçok perakendeci ve üretici gelecek döneme ilişkin satış tutarlarını kesin olarak tahmin etmek ister. Çoğu tedarikçinin, satışçının geçmişe yönelik detaylı satış, tüketim verileri ellerinde bulunmaktadır. Yapay sinir ağları yöntemi geçmişe dayalı büyük veri tabanlarını analiz edip, veriler arasındaki karmaşık ilişkileri değerlendirerek gelecek döneme ilişkin tahmin yapabilen teknolojidir. Yapay sinir ağları yönteminde geleneksel yöntemlerde olduğu gibi geçmiş veriler kullanılmaktadır. Parametre değerlerini belirlemek amacıyla en küçük kare hatasını üretmeye çalışır. En uygun çözüm elde edilene

kadar birçok yapay sinir ağı iterasyondan geçer. Talep tahmininde kullanılan YSA da bir girdi katmanı, bir gizli katman ve çıktı katmanı bulunmaktadır. Girdi katmanı bağımsız değişkenlerden oluşmaktadır. Gizli katmanda bulunan hücre sayısı girdi katmanını geçmemelidir. Doğal gaz tüketimi tahmininde kullanılan yapay sinir ağları yönteminde ağa girdi olarak günlük, haftalık, aylık olmak üzere dönemsel bazda ortalama sıcaklık, nem, basınç, fiyat gibi tüketimi etkileyen veriler öğretilmiştir. Tüketim miktarları ağın çıktı katmanında kullanılmaktadır. Verilerde eksik, yanlış olsa bile yapay sinir ağlarının hata toleransın yüksek olmasından faydalanarak tahmin yapılmaktadır. Farklı algoritmalar eşliğinde yapılan tahminlerde düzeltilmiş belirleme katsayısına bakarak tahminin doğruluğu ölçülür. Katsayı ne kadar yüksekse doğruluk oranı o kadar fazla demektir.

4. Sonuç

Bu derlemede yapay sinir ağları kullanılarak doğal gaz tüketimine ait talep tahmin çalışmalarının nasıl yapıldığı incelenmiştir. Yapay sinir ağları yöntemi anlatılmış ayrıca doğal gaz piyasasından bahsedilip küresel ve Türkiye enerji piyasasındaki gelişmeler değerlendirilip, Türkiye'nin ekonomisi için doğal gaz tüketim tahmininin önemine dikkat çekilmiştir.

Dünya'daki doğal gaz rezervleri incelendiğinde, artan enerji talebinin ve gelişen ekonomilerin enerji ihtiyacını artırdığı gözlemlenmektedir. Doğal gaz talebinin doğru yapılması, yurtdışından gaz alımları ile ilgili anlaşmaları, sektöre yönelik yapılacak yatırımları ve buna bağlı olarak sektörün gelişimini etkileyecek etkenlerden birini oluşturmaktadır.

Türkiye'de doğal gaz kullanımına başladığı günden bu yana gitgide yaygınlaşmış ve ülkemizin enerji arzının en değerli kaynaklarından biri olmuştur. Ancak ülkemizin doğal gaz rezervi oldukça düşüktür. Bun bağlı olarak kendi ihtiyaçlarımızı yurt dışından ithal ederek karşılamaktayız.

Doğal gaz tüketiminin sektörel bazda kullanım miktarlarına bakıldığı zaman 2018 EPDK verilerine göre en fazla dönüşüm sektöründe kullanıldığı görülmüştür. Buna bağlı olarak doğal gaz akışının güvenilir ve kesintisiz bir şekilde devam etmesi gerekmektedir.

İncelenen tezlerden çıkarılan ortak sonuç, doğal gaz tüketimi etkileyen faktörlerin geçmiş verileri dönemsel olarak (günlük, haftalık, aylık...) bağımsız değişkenler şeklinde ağa girdi olarak verilmiştir. Böylece yapay sinir ağlarının problemi öğrenmesi sağlanarak istenilen döneme ilişkin tüketim verilerinin tahmini yapılmıştır. Çoğu tezde görüldüğü üzere karmaşık problemler arasındaki ilişkiyi çözebilen, belirsizlik altında karar verebilen yapay sinir ağları yöntemi geleneksel tahmin yöntemlerine göre daha doğru sonuçlar elde etmektedir.

Çıkar İlişkisi

Yazarlar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Kaynaklar

- Adıyaman F. 2007. Talep tahmininde yapay sinir ağlarının kullanılması. Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı, 109, İstanbul.
- Akgül S, Yıldız Ş. 2013. Doğal gaz tüketim tahmini. Sosyal ve Beşeri Bilimler Derg, 5(1): 440-452.
- Ballı MT. 2014. Yapay sinir ağları ile talep tahmini ve gıda sektöründe uygulanması. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı, 180, İstanbul.
- Demirceylan S. 2012. Erzurum'da doğalgaz tüketim miktarının yapay sinir ağları yöntemi kullanılarak tahmin edilmesi. Yüksek Lisans Tezi, Atatürk Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı, 100, Erzurum.
- Hatipoğlu T. 2010. Galvaniz sektöründe bir yapay sinir ağı uygulaması. Yüksek Lisans Tezi, Sakarya Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı, 117, Sakarya.
- Karaatlı M, Helvacıoğlu ÖC, Ömürbek N, Tokgöz G. 2012. Yapay sinir ağları yöntemi ile otomobil satış tahmini. Uluslararası Yönetim İktisat ve İşletme Derg, 8(17): 88-100.
- Kaynar O, Taştan S, Demirkoparan F. 2011. 10. Ekonometri ve İstatistik Sempozyumu Özel Sayısı [Özel Sayı]. Atatürk Ü. İBBF Derg, 463-474.
- Khotanzad A, Elragal H. 2000. Combination of artificial neural network forecasters for prediction of natural gas consumption. IEEE Transacts on Neural Networks, 11: 464-473.
- Kızılaslan R. 2008. Forecasting of short term and mid term İstanbul natural gas consumption values by neural Networks algorithm. Master Thesis, Fatih University the Graduate Institute of Sciences and Engineering Industrial Engineering, 157, İstanbul.
- Kuru V. 2014. Türkiye doğal gaz tüketiminin sektörel bazda analitik incelenmesi. Yüksek Lisans Tezi, Türk Hava Kurumu Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, 104, İstanbul.
- Mazak M. 2004. Doğal gazın tarihi serüveni. DTK Dergisi. URL: <http://mehmetmazak.net/makale/3/270-istanbulda-ilk-modern-aydinlatilan-mekan-dolmabahce-sarayi-ve-dolmabahce-gazhanesi-tbmm-150-yilinda-dolmabahce-sarayi-uluslararası-sempozyumu-2006#.Xd2M90gzaM9> (erişim tarihi: 03.03.2020).
- Meral G. 2019. Türkiye'de enerji santrallerinde doğal gaz tüketiminin destek vektör regresyon ile tahmini. Yüksek Lisans Tezi, Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, 47, Afyon.
- Oğuz ZD. 2019. Türkiye'de tasarruf açığının yapay sinir ağları modeli ile analizi. Yüksek Lisans Tezi, Van Yüzüncü Yıl Üniversitesi Sosyal Bilimler Enstitüsü İktisat Anabilim Dalı, 105, Van.
- Oruç KO, Eroğlu ŞÇ. 2017. Isparta ili için doğal gaz talep tahmini. Süleyman Demirel Üniv İktisadi ve İdari Bil Fak Derg, 22(1): 31-42.
- Öztemel E. 2012. Yapay sinir ağları (Üçüncü baskı). Papatya Yayıncılık, 44, İstanbul.
- Öztürk A, Alkan Ö, Tosun S. 2018. Rüzgar ve güneş santrallerinde kısa dönem enerji üretim tahmini için matematiksel modellerin oluşturulması. Düzce ÜnivBilim ve Teknoloji Derg. 6: 188-195.
- Rodger JA. 2014. A fuzzy nearest neighbor neural network statistical model for predicting demand for natural gas and energy cost savings in public buildings. Expert Systems with App, 41: 1813-1829.
- Salih M. 2012. Dünyanın en büyük doğal gaz üreticileri. URL: <https://www.bloomberght.com/haberler/haber/1269673->

- dunyanin-en-buyuk-dogalgaz-ureticileri (erişim tarihi: 26.11.2019).
- Sönmez Çakır F. 2019. Matlab kodları ve matlab toolbox çözümleri (İkinci Baskı). Nobel Yayıncılık, 107, Ankara.
- Szoplik J. 2015. Forecasting of natural gas consumption with artificial neural network. Energy J, 85: 208-220.
- Taşkırner B. 2018. Ankara ili doğal gaz tüketiminin yapay sinir ağları ile öngörüsü. Yüksek Lisans Tezi, İstanbul teknik Üniversitesi Enerji Enstitüsü Enerji Bilim ve Teknoloji Anabilim Dalı, 67, İstanbul.
- Tuna Ç. 2019. Doğal gaz talep tahmini: Erzurum ili üzerine bir uygulama. Yüksek Lisans Tezi, Erzurum Teknik Üniversitesi Sosyal Bilimler Enstitüsü İktisat Anabilim Dalı, 167, Erzurum.
- WEC. 2019. Natural Gas Resources Report. World Energy Council.
- Yılmaz Ü. 2010. Altı sigma ve yapay sinir ağlarının tekstil sektöründe karşılaştırmalı bir uygulaması. Yüksek Lisans Tezi, Uludağ Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı, 143, Bursa.