



VOLUME: 2 (2019)

ISSUE:1

ISSN: 2687-3052



## Journal of Investigation on Engineering & Technology

<http://dergipark.gov.tr/jiet>



**PUBLISHER:**

*Yayıncı*

Dr. Canan AKSOY

**JIET FOUNDING COMMITTEE:**

*JIET Kurucu Kurul*

Dr. Canan AKSOY  
Dr. Emin TUĞCU  
Dr. Erol İSKENDER  
Dr. Hamdi Tolga KAHRAMAN  
Dr. Hasan Tahsin ÖZTÜRK

**EDITOR IN CHIEF:**

*Baş Editör*

Dr. Canan AKSOY

**EDITORS:**

*Editörler*

Dr. Canan AKSOY  
Dr. Emin TUĞCU  
Dr. Erol İSKENDER  
Dr. Hamdi Tolga KAHRAMAN  
Dr. Hasan Tahsin ÖZTÜRK

**EDITORIAL BOARD:**

*Yayın Kurulu*

Dr. Adem DOĞANGÜN (Uludağ University)  
Dr. Umut TOPAL (Karadeniz Technical University)  
Dr. Zeki KARACA (Ondokuz Mayıs University)  
Dr. Tayfun DEDE (Karadeniz Technical University)  
Dr. Egemen ARAS (Bursa Technical University)  
Dr. Nurcan ÖZTÜRK (Karadeniz Technical University)  
Dr. Ümit ÇALIK (Karadeniz Technical University)  
Dr. Erdem TÜRKELİ (Ordu University)  
Dr. Erdoğan DOĞDU (Çankaya University)  
Dr. Şeref SAĞIROĞLU (Gazi University)  
Dr. Güngör BAL (Gazi University)  
Dr. M. Ali AKCAYOL (Gazi University)  
Dr. Recep DEMİRCİ (Gazi University)  
Dr. Tuncay YİĞİT (Süleyman Demirel University)  
Dr. Ercan Nurcan YILMAZ (Gazi University)  
Dr. Cemal YILMAZ (Gazi University)  
Dr. Uğur GÜVENÇ (Düzce University)  
Dr. Yusuf SÖNMEZ (Gazi University)  
Dr. Eyüp GEDİKLİ (Karadeniz Technical University)  
Dr. Mehmet ŞİMŞEK (Düzce University)

Dr. İbrahim Alper DOĐRU (Gazi University)  
Dr. Atakan AKSOY (Karadeniz Technical University)  
Dr. Cemaleddin ŐİMŐEK (Karadeniz Technical  
Univetsity)  
Dr. Tuncay BAYRAM (Karadeniz Technical Univetsity)  
Dr. Burcu SAVAŐKAN (Karadeniz Technical  
Univetsity)  
Dr. Zeynep HASIRCI (Karadeniz Technical Univetsity)  
Dr. Emel AŐICI (Karadeniz Technical Univetsity)  
Dr. Bakiye AKIR (Artvin oruh Univetsity)  
Dr. Ezgi Taylan KOPARAN (B¼lent Ecevit Univetsity)  
Dr. Cenk ALBAYRAK (Karadeniz Technical Univetsity)  
Dr. ¼m¼r AKYAZI (Karadeniz Technical Univetsity)  
Dr. Esmat ATEŐ (Karadeniz Technical Univetsity)

## Product Information

*Ürün Bilgisi*

**Volume** **2** **Issue** **1** **August/2019**  
*Cilt* *Sayı* *Ağustos/2019*

**Publisher** Dr. Canan AKSOY  
*Yayıncı*

**Web Page** <http://dergipark.gov.tr/jiet>  
*Web Sayfası*

**Date of Publication** August /2019  
*Basım Tarihi* *Ağustos/2019*

**Language** English/Turkish  
*Yayın Dili* *İngilizce/Türkçe*

**Frequency** Published twice in a year  
*Yayın Aralığı* *Yılda iki kez yayınlanır*

**Type of Publication** Periodical  
*Yayın Türü* *Sürekli yayın*

**ISSN Number** 2687-3052  
*ISSN Numarası*

### Yazışma Adresi

Karadeniz Teknik Üniversitesi  
Teknoloji Fakültesi  
İnşaat Mühendisliği Bölümü  
Çamlı M. Hacı Mehmet Baheddin Ulusoy  
Cad. No:144 61830 Of/ TRABZON  
Telefon: +90 462 377 84 07  
E-posta: [jiet@ktu.edu.tr](mailto:jiet@ktu.edu.tr)

Makale gönderimi dergipark üzerinden yapılmaktadır. Tüm yayınlanan makalelere <http://dergipark.gov.tr/jiet> adresinden ulaşılabilir.

### Correspondence Address

Karadeniz Technical University  
Faculty of Technology  
Department of Civil Engineering  
Çamlı M. Hacı Mehmet Baheddin Ulusoy  
St. No:144 61830 Of/ TRABZON  
Phone: +90 462 377 84 07  
E-mail: [jiet@ktu.edu.tr](mailto:jiet@ktu.edu.tr)

Paper submission is done via dergipark. All published papers are available at <http://dergipark.gov.tr/jiet>.



## CONTENTS

### İçindekiler



#### **Research Papers**

*(Araştırma Makaleleri)*

#### **Pages**

*Sayfalar*

Measurement of Mass Attenuation Coefficients and Effective Atomic Numbers for MgB <sub>2</sub> -xSiC <sub>x</sub> (%5) Superconductor Using X- Ray Energies <b>Tuğba BAYAZIT, Canan AKSOY</b>	1-7
Yapay Zekâ Tabanlı Büyük Veri Yönetim Aracı <i>Yapay Zekâ Tabanlı Büyük Veri Yönetim Aracı</i> <b>Fatih ARSLAN, Hamdi Tolga KAHRAMAN</b>	8-21
Discharge Estimation by Drainage Area Ratio Method at Some Specific Discharges for 2251 Stream Gauging Station in East Black Sea Basin, Turkey <b>Fatih SAKA, Hasan Törehan BABACAN</b>	22-25
Performance of Color Shift Keying Modulation in Visible Light Communication Systems <i>Renk Kaydırmalı Anahtarlama Modülasyonunun Görünür Işık Haberleşme Sistemlerinde Başarımı</i> <b>Emin TUĞCU</b>	26-29

## X-ışını Enerjileri Kullanılarak $MgB_{2-x}SiC_x$ (%5) Süperiletken için Kütle Soğurma Katsayıları ve Etkin Atom Numarası Ölçümü

Tuğba BAYAZIT\*<sup>1</sup>, Canan AKSOY\*<sup>2</sup>

<sup>1</sup> Recep Tayyip Erdogan University, Central Research Laboratory, 53100, Rize, Turkey

<sup>2</sup>Karadeniz Technical University, Electronics and Communication Engineering, Faculty of Technology, 61830, Trabzon, Turkey

(Alınış / Received: 09.08.2019, Kabul / Accepted:30.08.2019, Online Yayınlanma/ Published Online: 31.08.2019)

### Anahtar Kelimeler

X- Işınları  
Kütle Soğurma Katsayısı  
Etkin Atomik Numarası

**Özet:** Süper iletken kütle  $MgB_{2-x}SiC_x$  (% 5) kütle soğurma katsayıları (% 5) 14,16 - 36,37 keV enerji aralığında ölçülerek hesaplanmış ve etkili atom sayıları tahmin edilmiştir. Belirlenen enerjileri, 241-Am'in gama ışını fotonları ile ışınlanmış sekonder hedefler kullanılarak elde edildi. X ışınları 5.9 keV'de 150 eV çözünürlüğe sahip Canberra Ultra-LEGe dedektörü kullanılarak sayıldı. Sonuçlar teorik olarak hesaplanan değerlerle karşılaştırıldı ve ortalama bir deneysel hata ile uyumlu olduğu tespit edildi.

## Measurement of Mass Attenuation Coefficients and Effective Atomic Numbers for $MgB_{2-x}SiC_x$ (%5) Superconductor Using X- Ray Energies

### Keywords

X -Rays  
Mass Attenuation  
Coefficient Effective  
Atomic number

**Abstract:** The mass attenuation coefficients for superconducting bulk  $MgB_{2-x}SiC_x$  (5%) were measured in the energy range 14,16 - 36,37 keV and effective atomic numbers were calculated. The energies were obtained by using secondary targets that were irradiated with gamma-ray photons of 241-Am. The X-rays were counted by using a Canberra Ultra-LEGe detector with a resolution of 150 eV at 5.9 keV. The results were compared with theoretical calculated values and fairly good agreement was found between them within an average experimental error.

\*İlgili yazar: tgdilek@gmail.com

<sup>a</sup>ORCID: 0000-0002-8695-0298

<sup>b</sup>ORCID: 0000-0002-0554-7123

## 1. Introduction

The electrical resistivity of many metals and alloys drops suddenly to zero value when the specimen is cooled down to a critical temperature which the resistant of the material suddenly drop to the zero, the phenomena is called superconductivity, was observed first by Kamerlingh Onnes in Leiden in 1911, three years after he first liquefied helium. After learned to superconductor's structure many superconductor alloys, compounds were studied by researchers. One of this superconductor compound is MgB<sub>2</sub> [1-6].

The boron atoms in MgB<sub>2</sub> form graphite like sheets separated by hexagonal magnesium atom layers. This compound is a mixed bonded solid: the bonding between boron atoms is covalent, between boron and magnesium atoms ionic and between magnesium atoms metallic [7]. Recently MgB<sub>2-x</sub>SiC<sub>x</sub> type superconductor was investigated and found that SiC doped are increased of flux for MgB<sub>2</sub> superconductor [8-10]

The mass attenuation coefficients are a measure of the average number of interactions that occur between incident photons and matter mass per unit area. They are related to the atomic number of element and the photon energy. In composite materials like alloys, for photon interactions a single atomic number cannot represent the atomic number uniquely across the entire energy region because of the effective number being related to the density and atomic number of an element. The partial interaction cross-sections have different elemental number dependence. This number for composite materials is called the effective atomic number and also it varies with energy [11].

There are many studies like Singh *et al* [12] measured gamma-ray attenuation coefficients in bismuth borate glasses using the narrow transmission method. Akkurt *et al* [13] made a study on Z dependence of partial and total mass attenuation coefficients with the help of XCOM. Materna *et al* [14] performed near K-edge measurement of x-ray attenuation coefficients of

bismuth and uranium using a tuneable hard X-ray source based on the linear electron accelerator. Moreover, there are many studies on the mass attenuation coefficients, ranging from determining the energy and composition dependences of the mass attenuation coefficient of building materials to determining the mass attenuation coefficients of Earth, Moon and Mars samples [15-23].

Effective atomic numbers and electron densities have been calculated for some investigators whose studies about amino acids and sugars [24-27] and polypyrrole [28] using total attenuation coefficients. El-Kateb *et al* [29] have measured the atomic cross sections and effective atomic numbers for some alloys. For some biological compounds containing H, C, N and O, Manjunathaguru and Umes [30, 31] have measured the total cross sections, effective atomic numbers and electron densities in the energy range 6.4-1330 keV. In addition to these studies, many researchers [32-42] performed a number of studies on the electron densities and effective atom and effective atomic numbers of different materials.

In this study the mass attenuation coefficients and effective atomic number for MgB<sub>2</sub> doped SiC (5%) was studied.

## 2. Theoretical information

Mass attenuation coefficients for the superconducting materials and energies are determined by the transmission for collimated monoenergetic beam. If a material of thickness x (cm) is placed in the path of beam of gamma or X-ray radiations, the intensity of beam will be attenuated according to Beer-Lambert's law:

$$I = I_0 e^{-\mu x} \quad (1)$$

where  $\mu$  (cm<sup>-1</sup>) is linear attenuation coefficient of the material and  $I_0$  and  $I$  are the non-attenuated and attenuated photon intensity. A coefficient, accurately characterizing a given material, is the density-independent mass attenuation coefficient  $\mu/\rho$  (cm<sup>2</sup>g<sup>-1</sup>).

$$I = I_0 e^{-(\mu/\rho)\rho x} = I_0 e^{-(\mu/\rho)d} \quad (2)$$

Where d is the mass per unit area (g/cm<sup>2</sup>)

The mass attenuation coefficient,  $\mu/\rho$ , for a compound or a mixture is given by:

$$\frac{\mu}{\rho} = \sum_i w_i \left( \frac{\mu}{\rho} \right)_i \quad (3)$$

where  $\rho$  is density of mass of the target, and,  $w_i$  and  $(\mu/\rho)_i$  are the weight fraction and mass attenuation coefficient, respectively, of the constituent element i. For a chemical compound, the fraction by weight is given by:

$$w_i = \frac{n_i A_i}{\sum_i n_i A_i} \quad (4)$$

where  $A_i$  is the atomic weight of the ith element and  $n_i$  is the number of formula units.

Theoretical values for the mass attenuation coefficients of MgB<sub>2-x</sub>SiC<sub>x</sub> alloy was calculated by program XCOM and photon cross-section database [44] according to the mixture rule:

$$\sigma_{t,m} = 1/N (\mu/\rho)_c \sum_i (n_i A_i) \quad (5)$$

where N ia the Avogadro number.

Total atomic cross-section ( $\sigma_{t,a}$ ) can be easily determined by using equation (5) as

$$\sigma_{t,a} = \frac{\sigma_{t,m}}{\sum_i n_i} = 1/N \sum_i f_i A_i (\mu/\rho)_i \quad (6)$$

where  $f_i = n_i / \sum_i n_i$  is the fractional abundance of element i with regard to number of atoms.

The total electronic cross-section ( $\sigma_{t,el}$ ) for the individual elements is expressed by the following formula:

$$\sigma_{t,el} = 1/N \sum_i f_i A_i / Z_i (\mu/\rho) \quad (7)$$

where  $Z_i$  is the atomic number of the ith element in a molecule.

The total atomic and electronic cross-sections are related to the effective atomic number ( $Z_{eff}$ ) through the following relation [4]:

$$Z_{eff} = \frac{\sigma_{t,a}}{\sigma_{t,el}} \quad (8)$$

The electron density  $N_{el}$  (number of electrons per unit mass) can be derived by using Equations (2) and (7)

### 3. Experimental procedure

According to general formula MgB<sub>2-x</sub>SiC<sub>x</sub> (x=0.05), MgB<sub>2</sub> and SiC-doped bulk superconductors were prepared by the standard ceramic processing. The mixtures of the corresponding powders were sintered at 750 0C for 0.5 h under pressure of 8 bar Argon.

The schematic arrangement of experimental set up was shown that previous paper [43].

The X-rays transmitted from the absorber were detected using a Canberra Ultra-LEGe detector with an active area of 30 mm<sup>2</sup>, a thickness of 5 mm, a Be window thickness of 0.4 $\mu$ m and an energy resolution of 0.15 keV at 5.96 keV. The output from the preamplifier, with pulse pile-up rejection capability, was fed to a multi-channel analyser interfaced with a personal computer provided with suitable software for data acquisition and a peak analysis program.

The signal-to-noise ratios of the spectra of the samples were in an acceptable region for a good-quality sample. A very narrow collimator is ideal to avoid scattered radiations reaching the detector. However, the narrow collimator reduces the count rate and adversely affects the counting statistics. To reduce the error due to poor statistics, the collimator has to be wider. Obviously, one has to arrive at an optimum value between these two opposite trends. The incident



and transmitted intensities are determined by integrating the counts in the channels under a photo peak. While determining the transmitted intensity, one has to avoid counting photon that have suffered small angle and multiple scattering but yet reach the detector. In an experiment like this, these important parameters have to be judiciously selected to arrive at accurate values for attenuation coefficients [45].

The dead time of the multichannel analyser is a combination of the rise time of the pulse, the conversion time in the analogue-to-digital converter and the data processing time. There was a built-in provision for dead time correction in the multichannel analyser used in the present measurements. In the present measurements, the percentage dead time correction was always 2–3%.

The thickness of the samples that are suitable for transmission measurements is dictated by counting statistics. The sample thickness has been optimized so that multiple scattering effects are corrected for in the present experiment. Also, in the present measurements, the intensity of bremsstrahlung secondary photons would be very low. A detector of high resolution as well as a thin absorber was used; so the effects of photon build-up are expected to be negligible in the present measurements. In addition, the pulse pile-up effects were kept at a minimum by selecting an optimal count rate [43].

In this study are investigated that the mass attenuation and the effective atomic number measured for MgB<sub>2-x</sub>SiC<sub>x</sub> superconductor.

#### 4. Results and Discussion

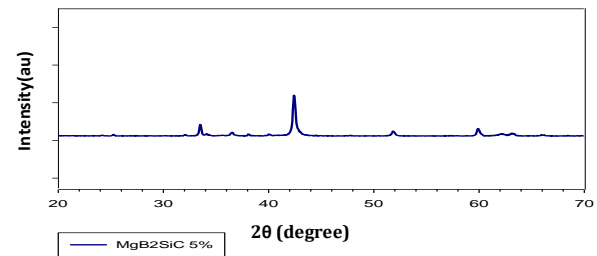
The nominal composition of MgB<sub>2</sub> superconducting powders were prepared by the mixing of magnesium and boron in the stoichiometric ratio 1:2 using the standard ceramic processing. SiC-doped MgB<sub>2</sub> powders were well mixed for 1.5 h with the ratio of %5. The mixtures of the 1 gram powder were pressed into the rectangle pellets form in the size of 10x15 mm

under the 10 ton pressure and put in a stainless steel tube under 8 barr argon gas pressures and sintered at 650 °C for 1.5 h. The stainless steel tube was removed from the furnace to air cool to room temperature.

In order to improve the homogeneity of reacted powders, the pellets were crushed into small pieces and ground using a mortar machine for 1.5 hours. Finally, the fine powders were pressed into the rectangle pellets form in the size of 10x15 mm under the 10 ton pressure and sintered at 750 °C for 0.5 hours under 8 bar argon gas pressures in same conditions.

The phase and crystal structure of all the samples was obtained from x-ray diffraction (XRD) patterns using a Rigaku D/Max-III diffract meter with CuK $\alpha$  radiation. Densities of samples were measurement using Archimedes method.

The x-ray diffraction (XRD) patterns of MgB<sub>2-x</sub>(SiC)<sub>x</sub> (x= 5%,) is shown in figure1. According to Fig.1. when the content of SiC increases some unknown peaks appear. This peaks belong to SiC and MgO.



**Figure 1.** XRD pattern of MgB<sub>2-x</sub>(SiC)<sub>x</sub> (x=5 %) sample

In Table 1, the experimental and theoretical mass attenuation coefficients ( $\mu/\rho$ ) for bulk MgB<sub>2-x</sub>SiC<sub>x</sub> superconductors sintered at 8 Ar pressure and a represented at 14.16, 15.77, 15.83, 17.66, 19.27, 21.65, 22.16, 24.94, 25.27, 27.47, 28.48, 30.99, 32.19, and 36.37 keV photon energies. Theoretical values of mass attenuation coefficients have been calculated by using XCOM computer program. Our results and theoretical values were shown at Table 1. According to values the

mass attenuation coefficients decreased with increasing photon energies.

**Table 1.** Experimental and theoretical values of the mass attenuation coefficients  $\mu/\rho$  (cm<sup>2</sup>g<sup>-1</sup>) for bulk MgB<sub>2-x</sub>SiC<sub>x</sub> superconductors

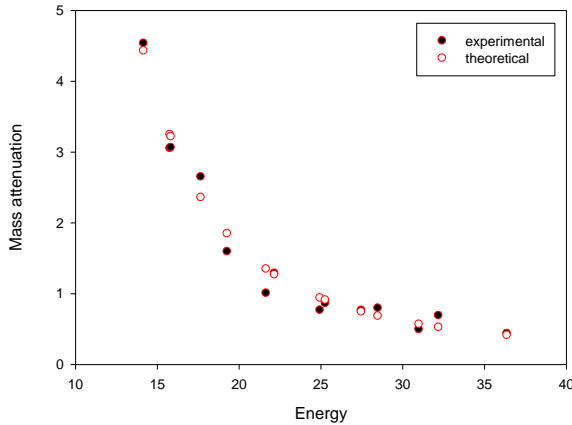
Energy (keV)	Exp.	Theo.
14.16	4.54	4.43
15.77	3.05	3.25
15.83	3.06	3.22
17.66	2.65	2.36
19.27	1.59	1.85
21.65	1.01	1.35
22.16	1.28	1.27
24.94	0.76	0.94
25.27	0.86	0.91
27.47	0.76	0.74
28.48	0.79	0.68
30.99	0.49	0.56
32.19	0.69	0.52
36.37	0.43	0.41

measurement are estimated to be 8–10% due to the evolution of peak areas ( $\leq 5\%$ ) sample thickness measurements ( $\approx 2\%$ ) and counting statistics ( $\leq 2\%$ ).

**Table 2.** Theoretical and experimental values of  $Z_{\text{eff}}$  for bulk MgB<sub>2-x</sub>SiC<sub>x</sub> superconductor

Energy (keV)	Theo.	Exp.
14.16	7.4	7.29
15.77	7.4	7.39
15.83	7.4	7.28
17.66	14.95	14.94
19.27	14.90	14.90
21.65	14.89	14.86
22.16	14.85	14.83
24.94	14.80	14.80
25.27	14.74	13.25
27.47	14.60	13.66
28.48	14.48	13.46
30.99	14.27	13.39
32.19	14.19	14.20
36.37	14.00	13.00

At the Fig. 2 is shown that experimental values and theoretical values are in good relation each other.



**Figure 2.** Comparison of theoretical and experimental data of Mass attenuation coefficients

By using experimental and theoretical data of  $\mu/\rho$ , the effective atomic numbers  $Z_{\text{eff}}$  were determined from Eq. (8) and the variation of  $Z_{\text{eff}}$  with photon energy for MgB<sub>2-x</sub>SiC<sub>x</sub> superconductors is tabulated in Table 2. These values show that  $Z_{\text{eff}}$  changes with the photon energy. Theoretical values and our results conform with each other. The errors in the present

In this paper the mass attenuation coefficients and the effective atomic numbers of MgB<sub>2-x</sub>SiC<sub>x</sub> superconductor were estimated at the energy range of 14.16–36.37. The results were compared with the theoretical values and there is qualitative agreement between each other. We have shown that a MgB<sub>2-x</sub>SiC<sub>x</sub> superconductor is a good absorber in the above mentioned energy range.

## REFERENCES

- [1] Angelone M. Bubba T. Esposito A. 2001. *Appl. Radiat. Isot.* 55. 505–511.
- [2] Baltas H. Cevik U. Tırasođlu E. Ertuđral. B. Apaydın G. Kobya A I. 2005. *Radiat. Meas.* 39. 33–37.
- [3] Feng Q R. Chen C. Xu J. Kong L W. Chen X. Wang Y Z. Zhang Y. Gao Z X. 2004. *Physica C* 411. 41–46.
- [4] Lorenz B. Meng Y Y. Xue R L. Chu C W. 2001. *Phys. Rev. B* 64 Art. No. 052513
- [5] Caplin A D. Bugoslavsky Y. Cohen L F. Perkins G.K. 2004. *Physica C* 401. 1–6.
- [6] Mickelson W. Cumings J. Han W Q. Zetti A. 2002. *Phys. Rev. B* 65 Art. No. 52505.

- [7] Pavindran R. Vajeeston P. Vidya R. Kjekshus A. Fjellvag H. 2001. *Phys. Rev. B* 64. 224509.
- [8] Wang X L. Soltanian S. James M. Qin M J. Horvat J. Yao. Q W. Liu H K. and Dou S X. 2004. *Physica C* 408-410 63-67.
- [9] Wang. X L. Zhou S H. Qin M. J. Munroe P R. Soltanian S. H K. and Dou S X. 2003. *Physica C* 385 461-465.
- [10] Shcherbakova O V. Pan A V. Soltanian S. H K. and Dou S X. 2007. *Physica C* 460-462 579-580.
- [11] Prasad S G. Parthasaradhi K and Bloomer W D 1998 *Radiat. Phys. Chem.* 53 449
- [12] Singh K. Singh H. Sharma V. Nathuram R. Khanna A. Kumar R. Bhatti S S and Sahota H S 2002 *Nucl. Instrum. Methods* B 194 1
- [13] Akkurt I. Mavi B. Akkurt A. Basyigit C. Kılıçarslan S and Yalim H A 2005 *J. Quant. Spectrosc. Radiat. Transfer* 94 379
- [14] Materna T H. Jolie J. Mondelaers W and Maschhaele B 2000 *Radiat. Phys. Chem.* 59 449
- [15] Singh C. Singh T. Kumar A and Mudahar G S 2004 *Ann. Nucl Energy* 31 1199
- [16] İçelli O. Erzeneoğlu S and Gürbulak B 2005 *J. Quant. Spectrosc. Radiat. Transfer* 90 399
- [17] Smale L F. Chantler C H. de Jonge M. Barnea Z and Tran C Q 2006 *Radiat. Phys. Chem.* 75 1559
- [8] Roy B. Chatterjee B K. Roy S C. Bhattacharya N and Choudhury N 1997 *Appl. Radiat. Isot.* 48 785
- [19] Singh K. Kaur G. Kumar V. Dhama A K and Lark B S 1998 *Radiat. Phys. Chem.* 53 123
- [20] Preseren R and Kodre A 1999 *Radiat. Phys. Chem.* 55 363
- [21] Alam M N. Miah M M H. Chowdhury M I. Kamal M. Ghose S and Rahman Runi 2001 *Appl. Radiat. Isot.* 54 973
- [22] Polat R. Budak G. Gürol A. Karabulut A and Ertuğrul M 2005 *Radiat. Meas.* 39 409
- [23] Moreira A C and Appoloni C R 2006 *Appl. Radiat. Isot.* 64 1065
- [24] Gowda S. Krishnaveni S and Gowda R 2005 *Nucl. Instrum. Methods* B 239 361
- [25] Manohara S R and Hanagodimath S M 2007 *Nucl. Instrum. Methods* B 258 321
- [26] Manohara S R. Hanagodimath S M and Gerward L 2008 *Phys. Med. Biol.* 53 377
- [27] Manohara S R. Hanagodimath S M. Thind K S and Gerward L 2008 *Nucl. Instrum. Methods* B 266 3906
- [28] İçelli O. Erzeneoğlu S and Sağlam M 2008 *Ann. Nucl. Energy* 35 432
- [29] El-Kateb A H. Rizk R A M and Abdul-Kader A M 2000 *Ann. Nucl. Energy* 27 1333
- [30] Manjunathaguru V and Umes T K 2006 *J. Phys. B: At. Mol. Opt. Phys.* 39 3969
- [31] Manjunathaguru V and Umes T K 2007 *J. Phys. B: At. Mol. Opt. Phys.* 40 3707
- [32] Çevik U. Baltas H. Çelik A and Bacaksız E 2006 *Nucl. Instrum. Methods* B 24 7173
- [33] İçelli O. Erzeneoğlu S. Karahan I H and Çankaya G 2005 *J. Quant. Spectrosc. Radiat. Transfer* 91 485
- [34] Shivaramu. Vijayakumar R. Rajasekaran L and Ramamurthy N 2001 *Radiat. Phys. Chem.* 62 371
- [35] Çevik U and Baltas H 2007 *Nucl. Instrum. Methods* B 256 619
- [36] Rizescu C. Besliu C and Jipa A 2001 *Nucl. Instrum. Methods* A 465 584
- [37] Duvauchelle P. Peix G and Babot D 1999 *Nucl. Instrum. Methods* B 155 221
- [38] Kumar T K and Reddy K V 1997 *Radiat. Phys. Chem.* 50 545
- [39] Shivaramu 2002 *Med. Dosim.* 27 1
- [40] Nayak N G. Vijaya M G and Siddappa K 2001 *Radiat. Phys. Chem.* 61 559
- [41] Baltas H. Çelik S. Çevik U and Yanmaz E 2007 *Radiat. Meas.* 42 55
- [42] Kaewkhao J. Laopai boon J and Chewpraditkul W 2008 *J. Quant. Spectrosc. Radiat. Transfer* 109 1260
- [43] Apaydın G. Cengiz E. Tıraşoğlu E. Aylıkçı V and Bakkaloğlu Ö F 2009 *Physica Scripta* 79 055302 (6pp)

[44] Berger M J. Hubbell J H. Seltzer S M. Chang J. Coursey J S. Sukumar R and Zucker D S 2005 *XCOM: Photon CrossSection Database (version 1.3)* (Gaithersburg, MD: National Institute of Standards and Technology) Online at <http://physics.nist.gov/xcom>

[45] Kaya N. Kobya A `I. Tırasođlu E and Apaydın G 2008 *J. Phys.B: At. Mol. Opt. Phys.* 41 225003

## Yapay Zekâ Tabanlı Büyük Veri Yönetim Aracı

Fatih ARSLAN\*<sup>1</sup>, Hamdi Tolga KAHRAMAN<sup>1a</sup>

<sup>1</sup> Karadeniz Teknik Üniversitesi, Teknoloji Fakültesi, Yazılım Mühendisliği Bölümü, 61830, Trabzon

(Alınış / Received: 06.08.2019, Kabul / Accepted: 22.08.2019, Online Yayınlanma / Published Online: 31.08.2019)

### Anahtar Kelimeler

Yapay Zekâ  
Eksik Veri Tamamlama  
Gürültü Tespiti ve Onarımı  
Yapay Sinir Ağları  
k-En Yakın Komşular(KNN)  
Sezgisel KNN

**Özet:** Yapay zekâ, günümüzde birçok problemin çözüme kavuşturulmasında başrol oynamaktadır. Şu an ki konum itibariyle en değerli maden haline gelen veri, bilginin oluşmasındaki asıl kaynaktır. Bilgiyi elde etme süreci göz önüne alındığında, kaliteli bir bilgiyi elde etmek için ise verinin incelenmesi, analiz edilmesi ve işlenmeye hazır hale getirilmesi gerekmektedir. Veri analiz sürecinde karşılaşılan problemlerin başında, eksik ya da gürültülü/hatalı verilerin tespit edilmesi ve düzeltilmesi gelmektedir. Bu çalışmada eksik ve gürültülü verilerin tespit edilmesi ve düzeltilmesi amacıyla yapay zekâ tabanlı çalışan, özgün ve güçlü bir veri yönetim aracı geliştirilmiştir. Bu araç sayesinde veri setlerinin analiz edilmesi, bu veri setlerindeki eksik ve gürültülü verilerin tespit edilmesi ve düzeltilmesi sağlanacaktır. Geliştirilecek yazılım aracının özgünlüğü ise eksik ve gürültülü verileri düzeltme sürecinde modern, güçlü ve melez yapay zekâ algoritmalarını kullanacak olmasıdır.

## Artificial Intelligence Based Big Data Management Tool

### Keywords

Artificial intelligence  
Missing Data Completion  
Noise Detection and Repair  
Artificial neural networks  
K nearest neighborhood(KNN)  
Intuitive KNN

**Abstract:** Today, artificial intelligence plays a leading role in solving many problems. Data, which has become the most valuable mine in terms of the current location, is the main source of information. When the process of obtaining information is taken into consideration, in order to obtain a quality information, the data must be examined, analyzed and made ready for processing. One of the problems encountered during the data analysis process is the identification and correction of missing or noisy / incorrect data. In this study, a unique and powerful data management tool based on artificial intelligence will be developed in order to detect and correct missing and noisy data. With this tool, data sets will be analyzed, missing and noisy data in these data sets will be detected and corrected. The originality of the software tool to be developed is that it will use modern, powerful and hybrid artificial intelligence algorithms in the process of correcting missing and noisy data.

## 1. Giriş

Teknolojik gelişmelerin en hızlı yaşandığı ve hissedildiği alanların başında yapay zekâ gelmektedir. Yapay zekânın gelişimi üzerinde etkili olan başlıca öğelerden biri ise veridir. Veri, geçmişten günümüze bilginin oluşmasındaki asıl kaynaktır. İnsanın bir hücrelerinde tutulan verinin terabaytlar mertebesinde olduğu düşünülürken biyolojik olarak insanın veri depolama kapasitesinin günümüz teknolojilerinden üstün olduğu söylenebilir. Bunun yanında veri işleme teknolojilerinde son yıllarda yaşanan gelişmeleri göz ardı etmek mümkün değildir. Çok çekirdekli, paralel işlem yapabilen, yüksek hızlı ve büyük kapasiteli yeni nesil işlemciler yanında veriyi bilgiye dönüştürmede insan ve diğer canlılara benzer karar mekanizmaları kullanan algoritmalar sayesinde çok güçlü veri madenciliği araçları geliştirilmektedir. Şirketler ürünlerinden daha fazla kazanç elde edebilmek, çalışmalarında verimliliği artırmak ve karar mekanizmalarını güçlendirebilmek için veri madenciliği için geliştirilmiş araçlardan faydalanmaktadırlar. Bu süreçte veri kalitesi ve bütünlüğü için harcadıkları zamanı ve bütçeyi artırmaktadırlar. Bunun nedeni, veri kalitesi ve bütünlüğünün firmalara katmakta olduğu değerin fark yaratmasıdır [1].

Yapay zekânın bir alt disiplini olan veri madenciliği teknolojik gelişmelerden etkilenerek büyük veri madenciliğine evrilmiştir. Bu değişimin temelinde ise internet-tabanlı alış-verişten uzay araçları ile yapılan haberleşmeye, endüstriyel otomasyon uygulamalarından sosyal medya uygulamalarına kadar sayısız alanda büyük bir veri yığının ortaya çıkması ve bunların elektronik ortamda saklanması ihtiyacı gelmektedir. Yarı iletken teknolojilerindeki gelişmeler, veriyi depolama ve işleme kapasitesi yüksek elektronik malzemelerin üretilmesini sağlamıştır. Verideki ve işlem yapma kapasitesindeki artış ise veriyi işleyerek anlamlı bilgiye dönüştürmede kullanılan yapay zekâ algoritmalarının tekrar gözden geçirilmesine neden olmuştur. Geçmişte, günümüze kıyasla küçük sayılabilecek veri yığınları üzerinde etkili sonuçlar üreten algoritmalar bugün ise büyük veri işleme gereksinimini aynı şekilde karşılayamamaktadırlar. Günümüzde büyük veriyi işleyerek en kısa sürede en uygun sonucu bulmak giderek zor bir hale gelmiştir. En uygun sonuç, kabul edilebilir bir maliyet ile en kaliteli olan sonuçtur. Başarılı bir veri madenciliği için veriyi işleyecek algoritmadan daha öncelikli verinin kendisidir. Verinin, problem uzayını homojen bir şekilde örneklemesi ve doğru olması gerekir. Yani başarılı bir veri madenciliği etkisi elde etmek için kaliteli bir veriye ihtiyaç vardır. Kaliteli bir veriyi elde etmek için ise verinin incelenmesi, analiz edilmesi ve işlenmeye hazır hale getirilmesi gerekir. Veri analiz süreci, bilimsel araştırma sürecinin ve veri madenciliğinin en önemli basamaklarından biridir. Bu süreçte toplanan

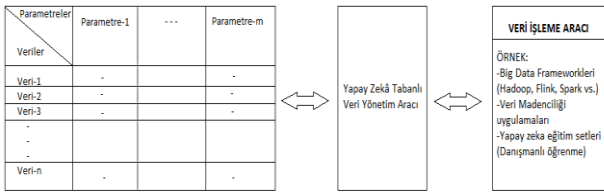
veriler amaca en uygun istatistiksel, matematiksel ya da yapay zekâ teknikleriyle işlenir veya analiz edilir.

Veri analiz sürecinde karşılaşılan problemlerin başında, eksik ya da gürültülü/hatalı verilerin tespit edilmesi ve düzeltilmesi gelmektedir. Bu sorunun çözümünde kullanılmak üzere geçmişten günümüze farklı yöntemler geliştirilmiştir. Eksik veri ile analize devam etme, eksik gözlemleri analiz dışı bırakma, eksik gözlemler yerine veri atama veya çeşitli istatistiksel yöntemlerle eksik verileri tamamlama gibi yöntemler bu durumlarda sıkça kullanılmaktadırlar [2]. Bu yöntemler içerisinde araştırmacılar tarafından en çok kullanılan yöntemler, liste bazında silme ve çiftler bazında silme gibi eksik verileri analiz dışı bırakma yöntemleridir. Ancak yapılan çalışmalar bu yöntemlerin örnekleme kaybı, güvenilirlikte azalmaya, tahminlerde yanlışlığa neden olduğunu [3] ve yanlışlıktan kaynaklı olarak da örneklemin evreni temsil etme derecesinin düştüğünü göstermektedir [4]. Belirtilen bu sebeplerden dolayı, son yıllarda, bu yöntemler yerine, beklenti maksimizasyonu ve çoklu atama gibi modern yöntemler önerilmektedir. Çünkü bu yöntemler, silme yöntemleri gibi geleneksel kayıp veri yöntemlerinin aksine, yanlışlığın azaltılması, etkili parametre tahminlerinin yapılması ve daha büyük istatistiksel gücün sağlanması hususunda daha etkili sonuçlar vermektedir [2].

Bu proje çalışmasında eksik ve gürültülü verilerin tespit edilmesi ve düzeltilmesi amacıyla yapay zekâ tabanlı çalışan, özgün ve güçlü bir veri yönetim aracı geliştirilmiştir. Bu araç sayesinde veri setlerinin analiz edilmesi, bu veri setlerindeki eksik ve gürültülü verilerin tespit edilmesi ve düzeltilmesi sağlanmaktadır. Geliştirilmiş olan yazılım aracının özgünlüğü ise eksik ve gürültülü verileri düzeltme sürecinde modern, güçlü ve melez yapay zekâ algoritmalarını kullanmış olmasıdır. Bu süreçte sezgisel optimizasyon algoritmalarından (genetik algoritma [5], yapay arı kolonisi algoritması [6], ortak yaşayan organizmalar [7] ve meta-sezgisel tahmin ve sınıflandırma algoritmalarından [8-9]) faydalanılmaktadır. Geliştirilmiş uygulamanın basit bir ara yüz ile kullanılması ve problemlere ait veri setlerinin kolaylıkla düzenlenebilmesi sağlanmaktadır. Veri setindeki eksiklikler ve gürültülü veriler veri yönetim yazılımı tarafından otomatik olarak tespit edilip araştırmacılara rapor halinde sunulmaktadır. Hata tespiti yapıldıktan sonra, program hatalı verilerin yerine geçebilecek en uygun değerleri bulup değiştirme işlemi yapılmaktadır. Bu süreçte melez bir tahmin ve sınıflandırma tekniği olan "meta-sezgisel k-NN algoritması" ve doğrusal olmayan regresyon problemlerinin çözümlenmesi amacıyla literatürde en yaygın kullanılan algoritma olan yapay sinir ağları kullanılmıştır.

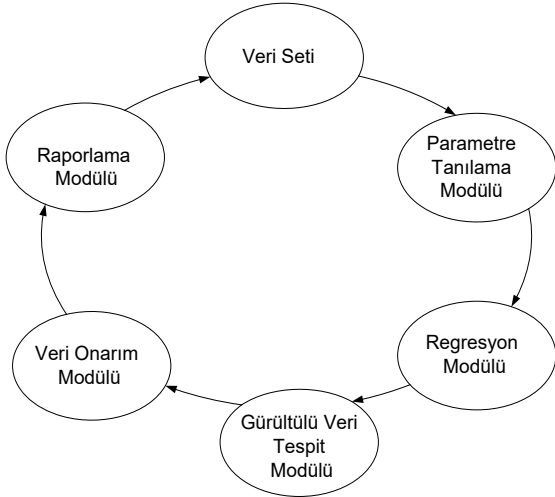
## 2. Yöntem

Bu çalışmada geliştirilen yapay zekâ tabanlı veri düzenleme aracı, meta-sezgisel k-nn algoritması ve yapay sinir ağlarının çok katmanlı bir mimari yapıda melezlenmesi ile geliştirilmiştir. İhtiyaç halinde diğer tahmin ve sınıflandırma algoritmaları sisteme dahil edilebilmektedir. Geliştirilmiş olan yapay zekâ tabanlı veri yönetim aracı web sunucusu üzerinde çalışmaktadır. Araştırmacılar web ara yüzü ile etkileşime girip veri setlerini kolay bir şekilde onarabilmektedirler. Şekil 1’de verildiği gibi, yapay zekâ tabanlı veri yönetim aracı veri işleme araçları ile veri seti arasına konumlandırılmaktadır. Yazılım aracı veri seti onarımını yaptıktan sonra, onarılmış veri setini araştırmacılara sunmaktadır.



**Şekil 1.** Yapay zekâ tabanlı veri yönetim aracının dağılım diyagramı

Yapay zekâ tabanlı veri yönetim aracı Şekil 2’de verildiği gibi 5 adet modülden oluşmaktadır. Bunlar: problem parametrelerini tanımlama modülü, regresyon (tahmin) modülü, gürültülü (eksik veya hatalı) veri tespit modülü, veri onarım modülü ve raporlama modülüdür.



**Şekil 2.** Veri yönetim süreci yaşam döngüsü

Problem parametrelerini tanımlama modülünde; kullanıcıdan alınan veri seti dosyası (excel tablosu formatında) okunup, problem parametreleri bağımlı ve bağımsız değişkenler olarak etiketlenilmektedir. Programlama araçlarının sahip olduğu işlevler kullanılarak her bir parametre için veri türü tespiti ve doğrulaması yapılmaktadır (sadece sürekli ya da ayrık

sayısal değerli niteliklerden oluşan veri setleri kabul edilecektir). Bu modül bir sonraki regresyon işlemi için probleme ait niteliklerin girişler ve çıkışlar şeklinde etiketlendirmesini sağlayacaktır.

Regresyon (tahmin) modülü: regresyon modülünün giriş verisi, bir önceki modülün çıktısı olan probleme ait etiketlenilmiş ve veri türleri doğrulanmış niteliklerden oluşan çok boyutlu bir veri setidir. Regresyon modülünün işlevi, problemlere ait bağımlı değişken(ler) ile bağımsız değişkenler arasındaki ilişkileri modellemektir. Bu amaçla modülde, literatürde yer alan iki güçlü yapay zekâ tekniği kullanılarak alternatif modeller oluşturulmaktadır. Bu teknikler yapay sinir ağı (YSA) ve sezgisel k-nn algoritmalarıdır. YSA ile yapılmak istenen, probleme ait bütün nitelikler için bir regresyon modeli yaratmaktır. Çünkü veri setinde herhangi bir veri örneğinin herhangi bir parametresi (niteliği) gürültülü veri içeriyor olabilir. Bu durumda probleme ait bütün parametreler için bir tahmin modeli yaratılması gerekir. Bu aşamada kullanılacak veri seti için birkaç senaryo üzerinden tartışma yapılabilir. İlk senaryo en olumsuz şartlardan biri olsun. Örneğin, veri seti gürültülü veri içeriyorsa bununla oluşturulacak bir regresyon modelinin performansı kötü olur mu? Literatürde YSA algoritması gürültülü veri örneklerine karşı güçlü ve etkili bir teknik olarak bilinmektedir. Bu yönüyle YSA ile oluşturulan modelden tatmin edici bir performans elde etmek mümkündür. Bunun yanında ideal ve olağan senaryo probleme ait kaliteli bir örnek veri setinin (gürültülü verilerden arındırılmış ve probleme ait bilgi alanındaki uzman tarafından doğrulanmış) bu aşamada kullanılmasıdır. Çünkü endüstriyel uygulamalarda ve bilimsel çalışmalarda öncelikle problem uzayını homojen bir şekilde temsil eden veri örneklerinden oluşan bir veri seti hazırlanır. Bu veri seti kullanılarak problem modeli oluşturulur. Daha sonraki aşamalarda ise gerçek bir sistem üzerinden alınan veri kullanılarak hedef parametreye yönelik tahmin işlemi gerçekleştirilir. Bu proje çalışmasında geliştirilecek olan regresyon modülünün amacı kaliteli veri üzerinden probleme ait bütün nitelikler için (sadece bağımlı değişken için değil) tahmin modellerini oluşturmak ve bir sonraki aşamayı gerçekleştiren “gürültülü veri tespit modülü” için hazır hale getirmektir. Böylelikle kullanıcıların “gürültülü veri tespit modülüne” yükleyeceği veri setindeki gürültülü verilerin tespit edilmesi ve düzeltilmesi sağlanacaktır. YSA dışında bu modülde parametre tahmini için ayrıca sezgisel k-nn algoritması kullanılmaktadır. Sezgisel k-nn algoritmasının seçilmesinin nedeni YSA’ya alternatif olabilecek nitelikte ve modern bir yapay zekâ tekniği olmasıdır. Regresyon modülünde sezgisel k-nn algoritması ile ilgili yapılacak çalışma bu algoritmanın probleme ve veri setine bağlı olarak değişen parametrelerinin (problem parametrelerinin ağırlık değerleri, k-nn algoritmasının k-değeri, uzaklık

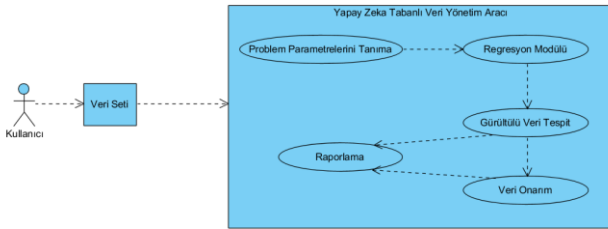


bağıntısı ve oylama yöntemi gibi) ideal değerlerini tespit etmektir [9].

Gürültülü veri tespit modülünde, bir önceki aşamada oluşturulan tahmin modelleri kullanılmaktadır. Bu tahmin modelleri, veri setindeki her bir veri örneği nesnesi için gürültülü değer içeren nitelikleri (bağımlı/bağımsız değişkenleri) tespit etmek için kullanılmaktadırlar. Bu süreçte veri setindeki boş ve dolu hücrelerin kontrolü yapılmaktadır. Boş verilere sahip hücreler işaretlenerek kayıt altına alınmaktadır. Dolu hücrelerdeki veriler için ise tahmin modelleri (YSA ve sezgisel k-nn) ile elde edilen değerler ile bu hücrelerin değerleri karşılaştırılacak ve veri setinde ilgili nitelik için belirlenen standart sapmanın çok üzerinde bir anlamlı farklılık olması durumunda bu hücrenin değeri gürültülü olarak etiketlendirilecektir. Bu süreçte “şüpheli” ve “kesin” şeklinde iki düzeyli bir etiketlendirme yapılmaktadır. Şüpheli durumlarda tahmin sonucunun mu yoksa hücredeki değer mi kullanılacağına ya da bu veri örneğinin silinme durumuna kullanıcı karar vermektedir. Bu durumlar kullanıcının onayına sunulmaktadır.

Veri onarım modülünde; hata tespit modülünden dönen sonuçlar alınıp, ilgili hücreler üzerinde gerekli işlemler yapılmaktadır. Boş hücrelerin doldurulması için ilgili nitelik için geliştirilen tahmin modeli kullanılarak değer oluşturulmaktadır. Dolu hücrelerdeki veriler için ise “şüpheli” durumda olanlar kullanıcı kararıyla ve “kesin” durumda olanlar da ilgili tahmin modellerinin üreteceği değerler ile düzeltilmektedir.

Hata raporu modülünde; onarımı yapılan hücrenin kayıtları tutulup, kullanıcıya rapor olarak sunulmaktadır. Araç gerçek ortamda çok büyük veriyle karşılaştığında kaç adet değişiklik yapıldı, hangi niteliklerde yüzde kaç problem yaşandı gibi ayrıca istatistiksel olarak raporlama yapılmaktadır.



**Şekil 3.** Yapay zekâ tabanlı veri yönetim aracının temel öğeleri

### 3. Deneysel Çalışma

#### 3.1. Veri seti: Enerji Verimliliği Veri Seti (Energy Efficiency Dataset)

##### 3.1.1. Veri Seti Bilgisi:

Ecotect'te simüle edilmiş 12 farklı yapı şeklini kullanarak enerji analizi yapılmaktadır. Veri kümesi, iki gerçek değerli bağımlı değişkeni tahmin etmeyi amaçlayan 768 örnek ve 8 özelliğinden (bağımsız değişken) oluşur [10].

##### 3.1.2. Özellik Bilgisi:

Veri seti, sekiz bağımsız değişken ( $X_1 \dots X_8$  ile gösterilen özellikler) ve iki bağımlı değişken ( $Y_1$  ve  $Y_2$  ile gösterilen sonuçlar) içerir. Amaç, iki bağımlı değişkenin her birini tahmin etmek için sekiz bağımsız değişken kullanmaktır [10].

##### 3.1.3. Problem Parametre Karşılıkları:

**Tablo 1.** Problem parametreleri

$X_1$	Göreceli Kompaktlık
$X_2$	Yüzey Alanı
$X_3$	Duvar Alanı
$X_4$	Çatı Alanı
$X_5$	Genel Yükseklik
$X_6$	Yönelim
$X_7$	Camlama Alanı
$X_8$	Camlama Alanı Dağılımı
$Y_1$	Isıtma Yüğü
$Y_2$	Soğutma Yüğü

#### 3.1.4. Problemin Yapay Sinir Ağları ile Modellenmesi ve Tahmin Modeli Hata Sonuçları

Problem ait veri seti sisteme yüklendikten sonra, sistem veri setine ait her bir problem parametresi (bağımlı veya bağımsız değişken) için bir tahmin modeli (Yapay Sinir Ağı) oluşturmaktadır. Oluşturulan tahmin modeli için MAPE (Ortalama Mutlak Yüzde Hata), MAE (Ortalama Mutlak Hata), MSE (Ortalama Kare Hata), Ortalama Tahmin Hatası ve tahmin modeli sonuçları ile gerçek sonuçların karşılaştırıldığı grafik çıktı olarak verilmektedir.

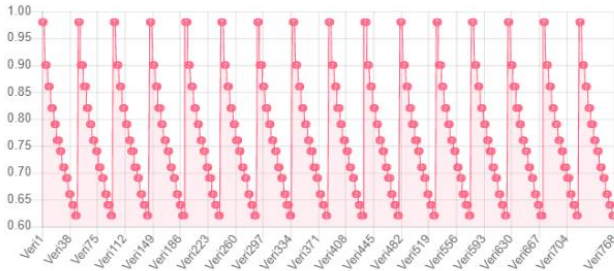


**Tablo 2.** Tahmin modelleri hata sonuçları

Problem Parametreleri	Hatalar			
	MAPE (%)	MAE	MSE	Ortalama Tahmin Hatası
X1	0	0	0	0
X2	0	0	0	0
X3	0	0	0	0
X4	0	0	0	0
X5	0	0	0	0
X6	33.44	1	1.25	0.01
X7	3.99	0.01	0	0
X8	59.18	1.15	1.87	0.02
Y1	0.95	0.18	0.08	0.02
Y2	1.12	0.24	0.12	0.01

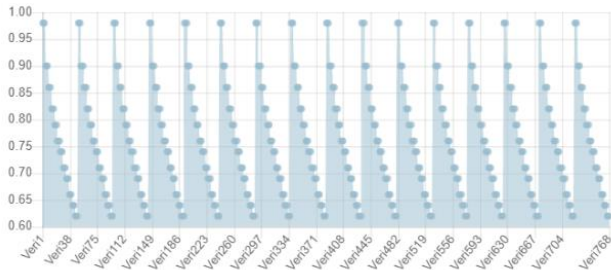
### 3.1.4.1. X1 parametresine ait tahmin modeli grafikleri

Veri setinde bulunan 768 veri örneği için “Göreceli Kompaktlık (X1)” parametresi veri dağılım grafiği aşağıdaki gibidir.



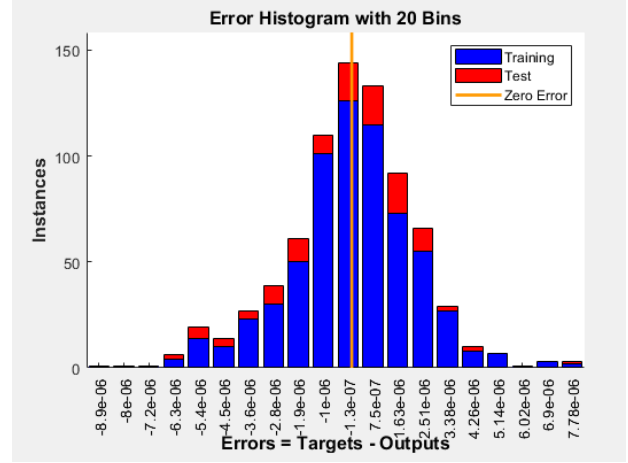
**Şekil 4.** X1 (1. Problem Parametresi) veri dağılım grafiği

X1 problem parametresini tahmin etmek için diğer 9 nitelikten (X2, X3, X4, X5, X6, X7, X8, Y1, Y2) faydalanılmaktadır. Veri setindeki 768 veri örneği baz alınarak X1 parametresi için bir tahmin modeli (Yapay Sinir Ağı) oluşturulmuştur. Eğitim için kullanılan veri seti, tahmin modeline gönderilmiş ve ele alınan sonuçlar Şekil 5'te ki grafik üzerinde gösterilmiştir.



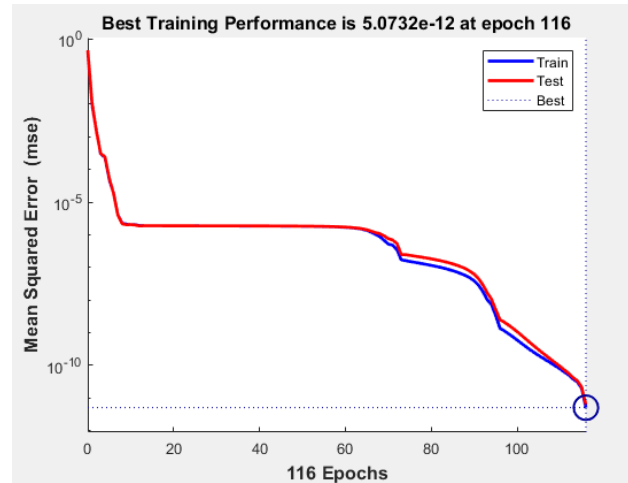
**Şekil 5.** X1 (1. Problem Parametresi) tahmin modeli sonucu veri dağılım grafiği

X1 problem parametresi için oluşturulan tahmin modelinin eğitimi sırasındaki hata değişimi Şekil 6'da ki histogramda gösterilmektedir.



**Şekil 6.** X1 (1. Problem Parametresi) tahmin modeli hata histogramı

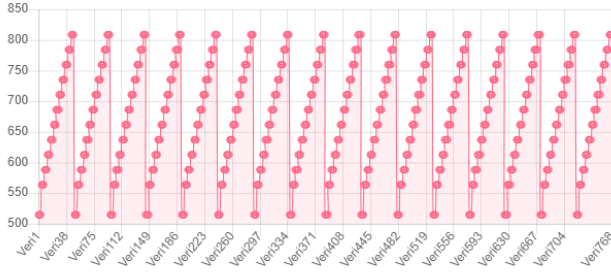
X1 problem parametresi için oluşturulan tahmin modelinin veri seti üzerindeki performansı Şekil 7'de ki grafikte gösterilmektedir.



**Şekil 7.** X1 (1. Problem Parametresi) tahmin modeli performans grafiği

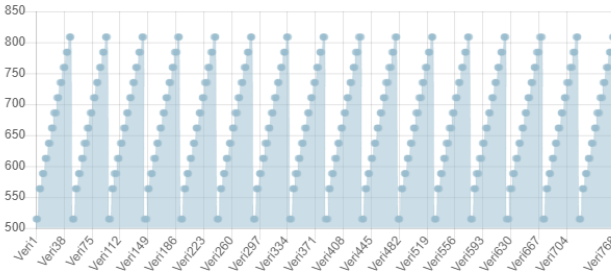
### 3.1.4.2. X2 parametresine ait tahmin modeli grafikleri

Veri setinde bulunan 768 veri örneği için “Yüzey Alanı (X2)” parametresi veri dağılım grafiği aşağıdaki gibidir.



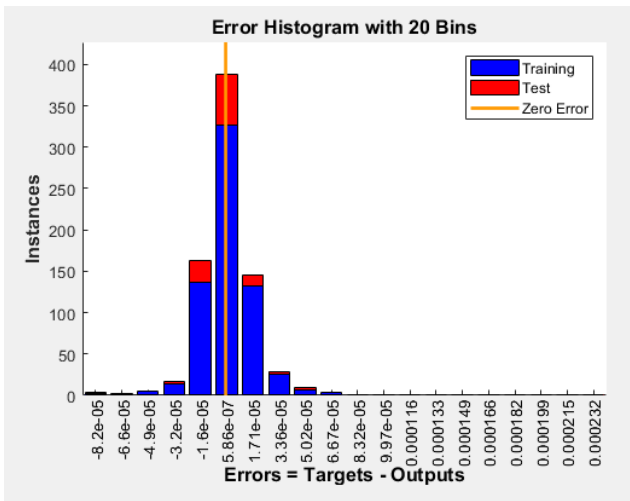
**Şekil 8.** X2 (2. Problem Parametresi) veri dağılım grafiği

X2 problem parametresini tahmin etmek için diğer 9 nitelikten (X1, X3, X4, X5, X6, X7, X8, Y1, Y2) faydalanılmaktadır. Veri setindeki 768 veri örneği baz alınarak X2 parametresi için bir tahmin modeli (Yapay Sinir Ağı) oluşturulmuştur. Eğitim için kullanılan veri seti, tahmin modeline gönderilmiş ve ele alınan sonuçlar Şekil 9'da ki grafik üzerinde gösterilmiştir.



**Şekil 9.** X2 (2. Problem Parametresi) tahmin modeli sonucu veri dağılım grafiği

X2 problem parametresi için oluşturulan tahmin modelinin eğitimi sırasındaki hata değişimi Şekil 10'da ki histogramda gösterilmektedir.



**Şekil 10.** X2 (2. Problem Parametresi) tahmin modeli hata histogramı

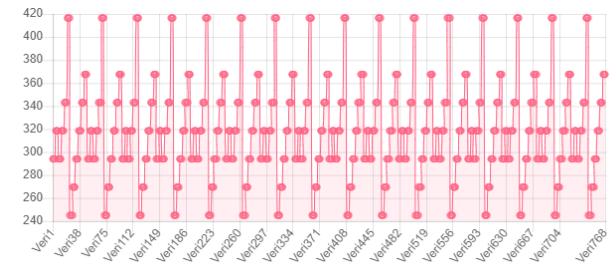
X2 problem parametresi için oluşturulan tahmin modelinin veri seti üzerindeki performansı Şekil 11'de ki grafikte gösterilmektedir.



**Şekil 11.** X2 (2. Problem Parametresi) tahmin modeli performans grafiği

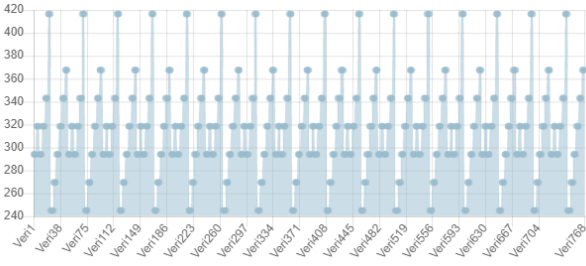
### 3.1.4.3. X3 parametresine ait tahmin modeli grafikleri

Veri setinde bulunan 768 veri örneği için "Duvar Alanı (X3)" parametresi veri dağılım grafiği aşağıdaki gibidir.



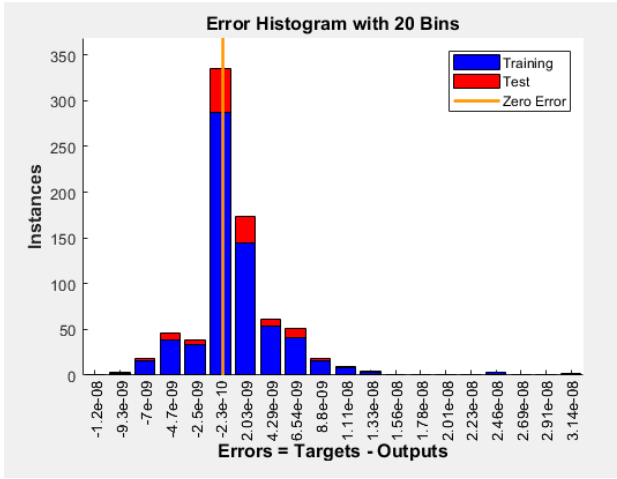
**Şekil 12.** X3 (3. Problem Parametresi) veri dağılım grafiği

X3 problem parametresini tahmin etmek için diğer 9 nitelikten (X1, X2, X4, X5, X6, X7, X8, Y1, Y2) faydalanılmaktadır. Veri setindeki 768 veri örneği baz alınarak X3 parametresi için bir tahmin modeli (Yapay Sinir Ağı) oluşturulmuştur. Eğitim için kullanılan veri seti, tahmin modeline gönderilmiş ve ele alınan sonuçlar Şekil 13'te ki grafik üzerinde gösterilmiştir.



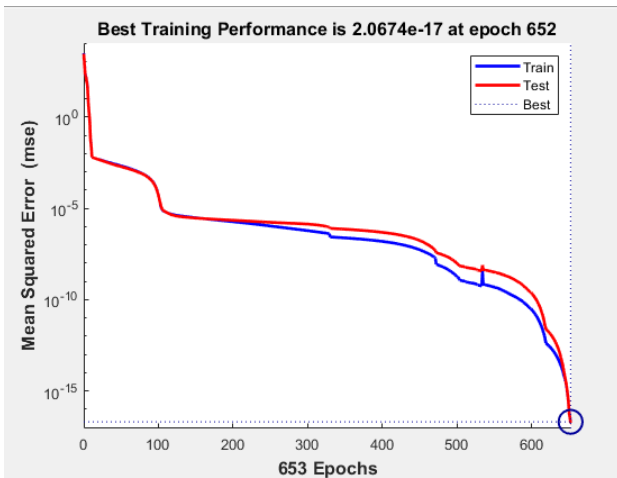
Şekil 13. X3 (3. Problem Parametresi) tahmin modeli sonucu veri dağılım grafiği

X3 problem parametresi için oluşturulan tahmin modelinin eğitimi sırasındaki hata değişimi Şekil 14'te ki histogramda gösterilmektedir.



Şekil 14. X3 (3. Problem Parametresi) tahmin modeli hata histogramı

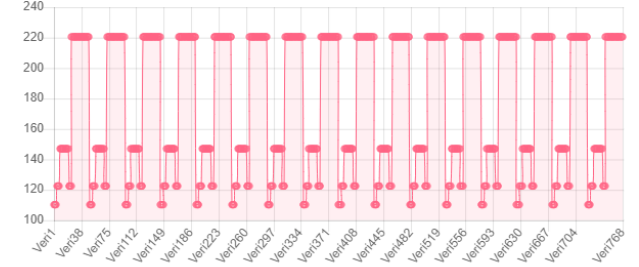
X3 problem parametresi için oluşturulan tahmin modelinin veri seti üzerindeki performansı Şekil 15'te ki grafikte gösterilmektedir.



Şekil 15. X3 (3. Problem Parametresi) tahmin modeli performans grafiği

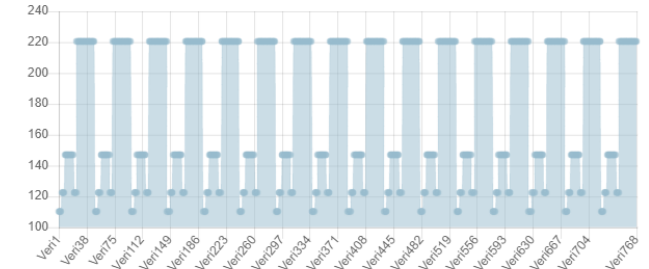
### 3.1.4.4. X4 parametresine ait tahmin modeli grafikleri

Veri setinde bulunan 768 veri örneği için "Çatı Alanı (X4)" parametresi veri dağılım grafiği aşağıdaki gibidir.



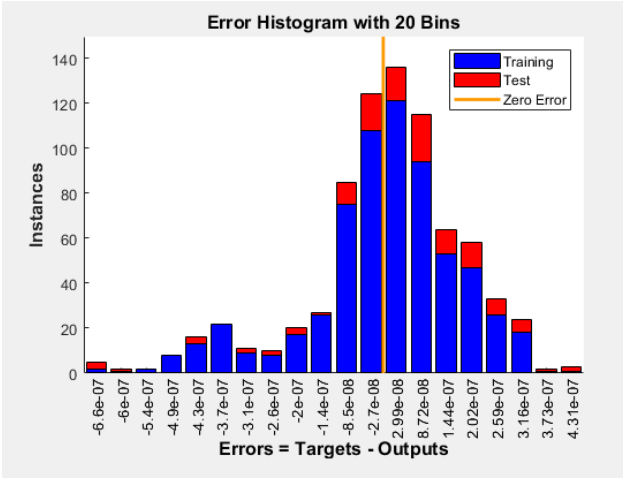
Şekil 16. X4 (4. Problem Parametresi) veri dağılım grafiği

X4 problem parametresini tahmin etmek için diğer 9 nitelikten (X1, X2, X3, X5, X6, X7, X8, Y1, Y2) faydalanılmaktadır. Veri setindeki 768 veri örneği baz alınarak X4 parametresi için bir tahmin modeli (Yapay Sinir Ağı) oluşturulmuştur. Eğitim için kullanılan veri seti, tahmin modeline gönderilmiş ve ele alınan sonuçlar Şekil 17'de ki grafik üzerinde gösterilmiştir.



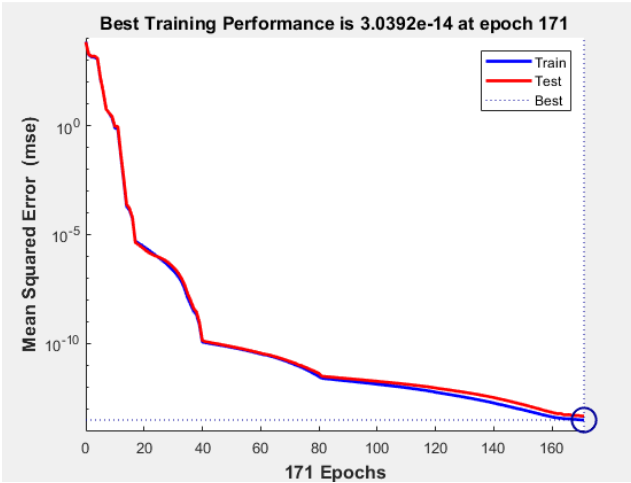
Şekil 17. X4 (4. Problem Parametresi) tahmin modeli sonucu veri dağılım grafiği

X4 problem parametresi için oluşturulan tahmin modelinin eğitimi sırasındaki hata değişimi Şekil 18'te ki histogramda gösterilmektedir.



Şekil 18. X4 (4. Problem Parametresi) tahmin modeli hata histogramı

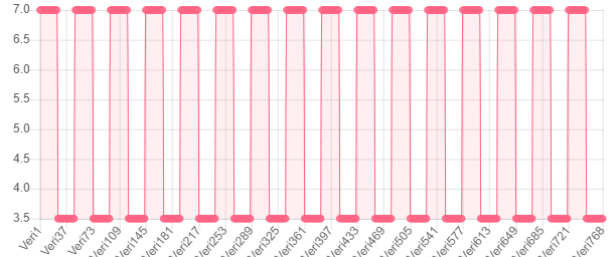
X4 problem parametresi için oluşturulan tahmin modelinin veri seti üzerindeki performansı Şekil 19'da ki grafikte gösterilmektedir.



Şekil 19. X4 (4. Problem Parametresi) tahmin modeli performans grafiği

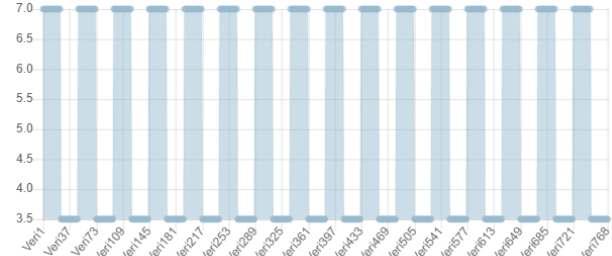
### 3.1.4.5. X5 parametresine ait tahmin modeli grafikleri

Veri setinde bulunan 768 veri örneği için "Genel Yükseklik (X5)" parametresi veri dağılım grafiği aşağıdaki gibidir.



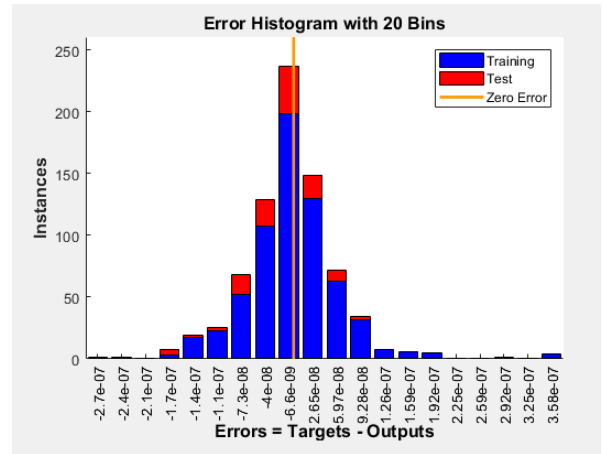
Şekil 20. X5 (5. Problem Parametresi) veri dağılım grafiği

X5 problem parametresini tahmin etmek için diğer 9 nitelikten (X1, X2, X3, X4, X6, X7, X8, Y1, Y2) faydalanılmaktadır. Veri setindeki 768 veri örneği baz alınarak X5 parametresi için bir tahmin modeli (Yapay Sinir Ağı) oluşturulmuştur. Eğitim için kullanılan veri seti, tahmin modeline gönderilmiş ve ele alınan sonuçlar Şekil 21'de ki grafik üzerinde gösterilmiştir.



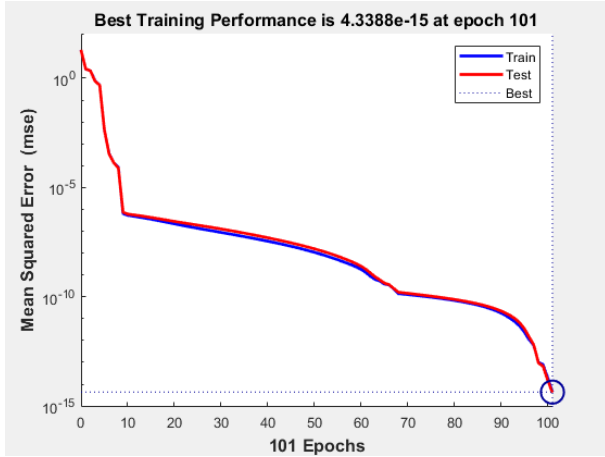
Şekil 21. X5 (5. Problem Parametresi) tahmin modeli sonucu veri dağılım grafiği

X5 problem parametresi için oluşturulan tahmin modelinin eğitimi sırasındaki hata değişimi Şekil 22'de ki histogramda gösterilmektedir.



Şekil 22. X5 (5. Problem Parametresi) tahmin modeli hata histogramı

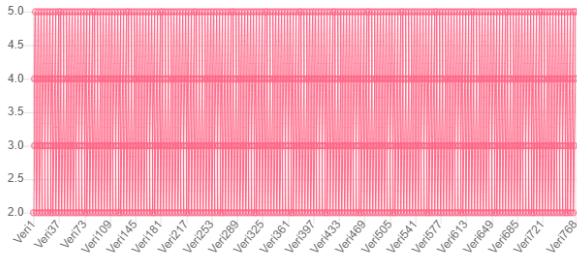
X5 problem parametresi için oluşturulan tahmin modelinin veri seti üzerindeki performansı Şekil 23'da ki grafikte gösterilmektedir.



Şekil 23. X5 (5. Problem Parametresi) tahmin modeli performans grafiği

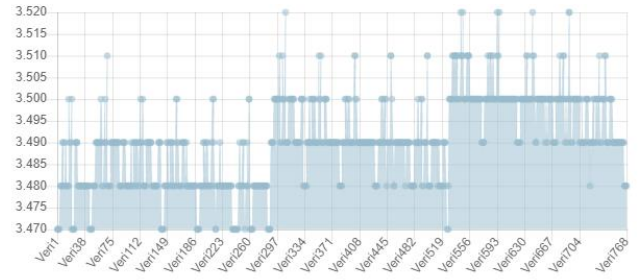
### 3.1.4.6. X6 parametresine ait tahmin modeli grafikleri

Veri setinde bulunan 768 veri örneği için "Yönelim (X6)" parametresi veri dağılım grafiği aşağıdaki gibidir.



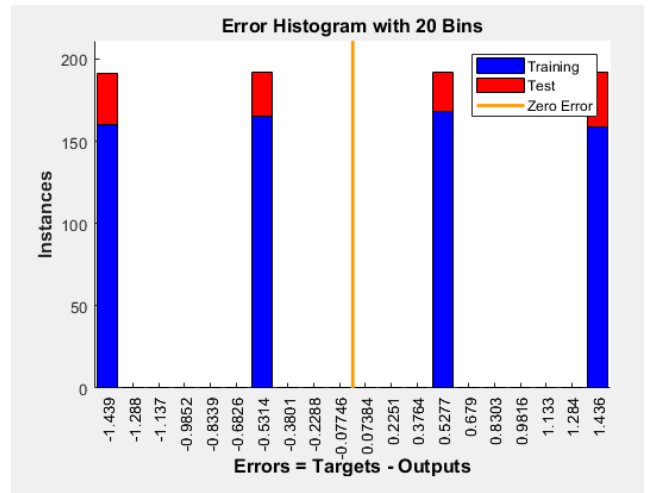
Şekil 24. X6 (6. Problem Parametresi) veri dağılım grafiği

X6 problem parametresini tahmin etmek için diğer 9 nitelikten (X1, X2, X3, X4, X5, X7, X8, Y1, Y2) faydalanılmaktadır. Veri setindeki 768 veri örneği baz alınarak X6 parametresi için bir tahmin modeli (Yapay Sinir Ağı) oluşturulmuştur. Eğitim için kullanılan veri seti, tahmin modeline gönderilmiş ve ele alınan sonuçlar Şekil 25'de ki grafik üzerinde gösterilmiştir.



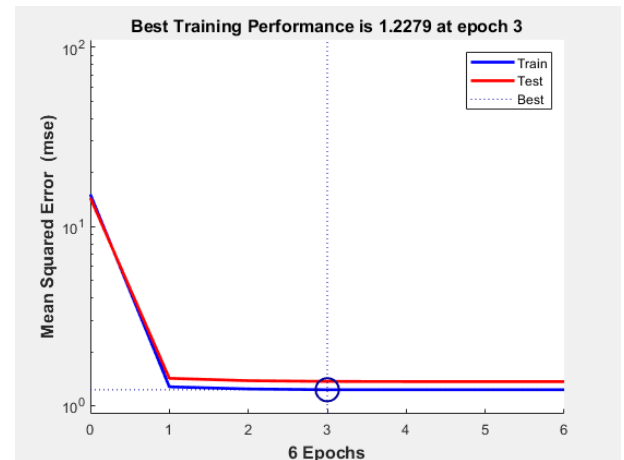
Şekil 25. X6 (6. Problem Parametresi) tahmin modeli sonucu veri dağılım grafiği

X6 problem parametresi için oluşturulan tahmin modelinin eğitimi sırasındaki hata değişimi Şekil 26'da ki histogramda gösterilmektedir.



Şekil 26. X6 (6. Problem Parametresi) tahmin modeli hata histogramı

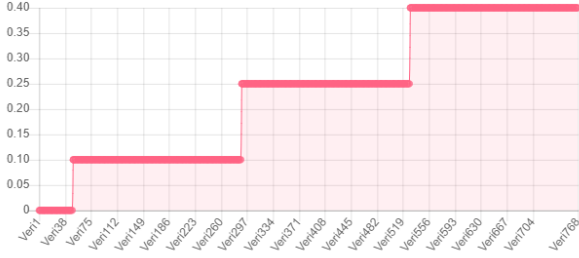
X6 problem parametresi için oluşturulan tahmin modelinin veri seti üzerindeki performansı Şekil 27'de ki grafikte gösterilmektedir.



Şekil 27. X6 (6. Problem Parametresi) tahmin modeli performans grafiği

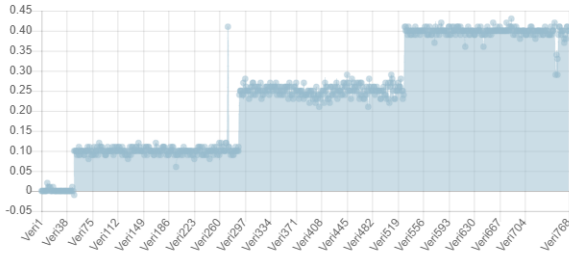
### 3.1.4.7. X7 parametresine ait tahmin modeli grafikleri

Veri setinde bulunan 768 veri örneği için “Camlama Alanı (X7)” parametresi veri dağılım grafiği aşağıdaki gibidir.



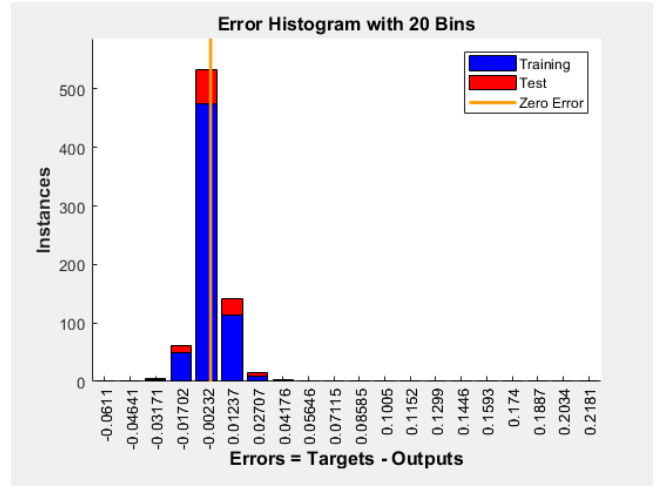
Şekil 28. X7 (7. Problem Parametresi) veri dağılım grafiği

X7 problem parametresini tahmin etmek için diğer 9 nitelikten (X1, X2, X3, X4, X5, X6, X8, Y1, Y2) faydalanılmaktadır. Veri setindeki 768 veri örneği baz alınarak X7 parametresi için bir tahmin modeli (Yapay Sinir Ağı) oluşturulmuştur. Eğitim için kullanılan veri seti, tahmin modeline gönderilmiş ve ele alınan sonuçlar Şekil 29'da ki grafik üzerinde gösterilmiştir.



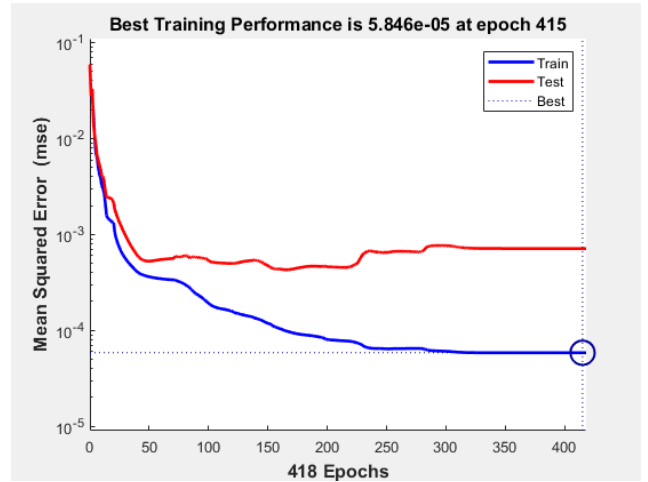
Şekil 29. X7 (7. Problem Parametresi) tahmin modeli sonucu veri dağılım grafiği

X7 problem parametresi için oluşturulan tahmin modelinin eğitimi sırasındaki hata değişimi Şekil 30'da ki histogramda gösterilmektedir.



Şekil 30. X7 (7. Problem Parametresi) tahmin modeli hata histogramı

X7 problem parametresi için oluşturulan tahmin modelinin veri seti üzerindeki performansı Şekil 31'de ki grafikte gösterilmektedir.

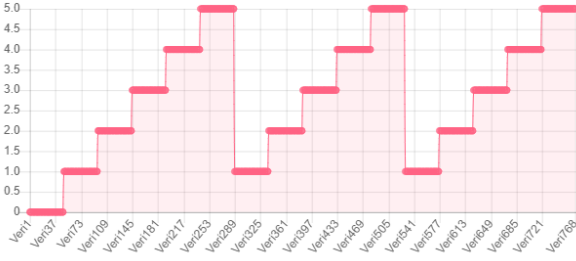


Şekil 31. X7 (7. Problem Parametresi) tahmin modeli performans grafiği

### 3.1.4.8. X8 parametresine ait tahmin modeli grafikleri

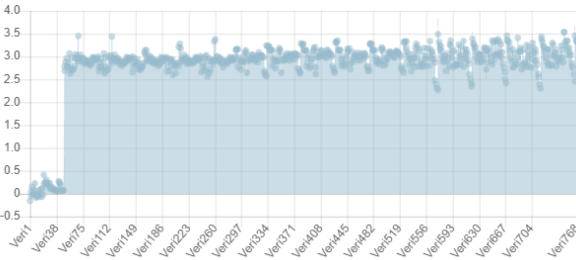
Veri setinde bulunan 768 veri örneği için “Camlama Alanı Dağılımı(X8)” parametresi veri dağılım grafiği aşağıdaki gibidir.





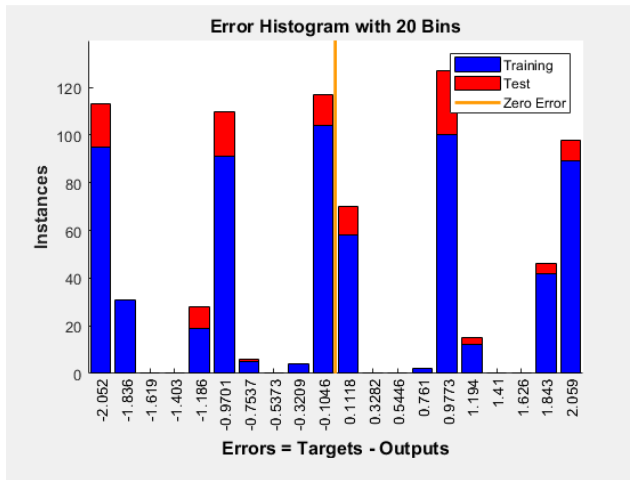
**Şekil 32.** X8 (8. Problem Parametresi) veri dağılım grafiği

X8 problem parametresini tahmin etmek için diğer 9 nitelikten (X1, X2, X3, X4, X5, X6, X7, Y1, Y2) faydalanılmaktadır. Veri setindeki 768 veri örneği baz alınarak X8 parametresi için bir tahmin modeli (Yapay Sinir Ağı) oluşturulmuştur. Eğitim için kullanılan veri seti, tahmin modeline gönderilmiş ve ele alınan sonuçlar Şekil 33'de ki grafik üzerinde gösterilmiştir.



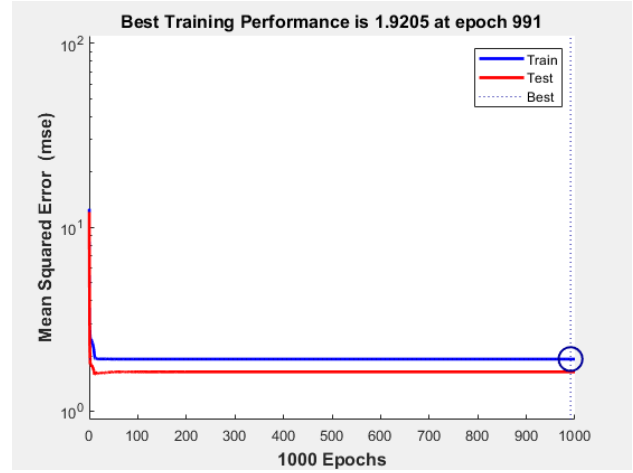
**Şekil 33.** X8 (8. Problem Parametresi) tahmin modeli sonucu veri dağılım grafiği

X8 problem parametresi için oluşturulan tahmin modelinin eğitimi sırasındaki hata değişimi Şekil 34'de ki histogramda gösterilmektedir.



**Şekil 34.** X8 (8. Problem Parametresi) tahmin modeli hata histogramı

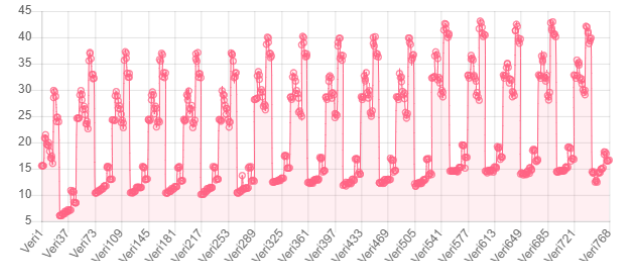
X8 problem parametresi için oluşturulan tahmin modelinin veri seti üzerindeki performansı Şekil 35'de ki grafikte gösterilmektedir.



**Şekil 35.** X8 (8. Problem Parametresi) tahmin modeli performans grafiği

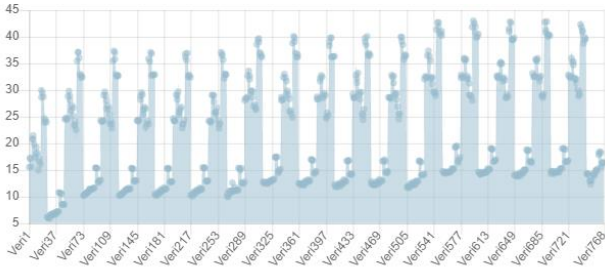
### 3.1.4.9. Y1 parametresine ait tahmin modeli grafikleri

Veri setinde bulunan 768 veri örneği için "Isıtma Yüğü (Y1)" parametresi veri dağılım grafiği aşağıdaki gibidir.



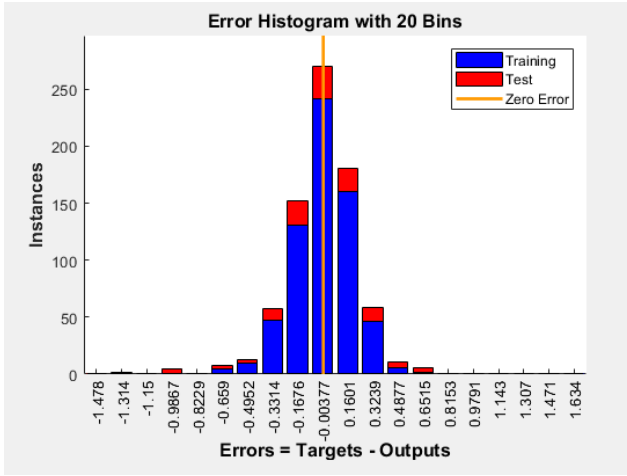
**Şekil 36.** Y1 (9. Problem Parametresi) veri dağılım grafiği

Y1 problem parametresini tahmin etmek için diğer 9 nitelikten (X1, X2, X3, X4, X5, X6, X7, X8, Y2) faydalanılmaktadır. Veri setindeki 768 veri örneği baz alınarak Y1 parametresi için bir tahmin modeli (Yapay Sinir Ağı) oluşturulmuştur. Eğitim için kullanılan veri seti, tahmin modeline gönderilmiş ve ele alınan sonuçlar Şekil 37'de ki grafik üzerinde gösterilmiştir.



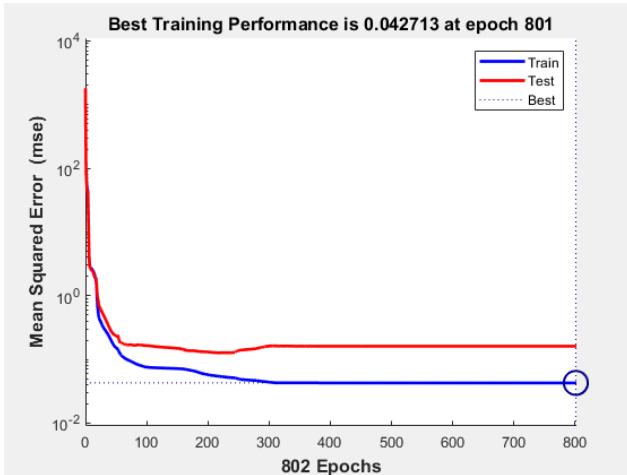
**Şekil 37.** Y1 (9. Problem Parametresi) tahmin modeli sonucu veri dağılım grafiği

Y1 problem parametresi için oluşturulan tahmin modelinin eğitimi sırasındaki hata değişimi Şekil 38'de ki histogramda gösterilmektedir.



**Şekil 38.** Y1 (9. Problem Parametresi) tahmin modeli hata histogramı

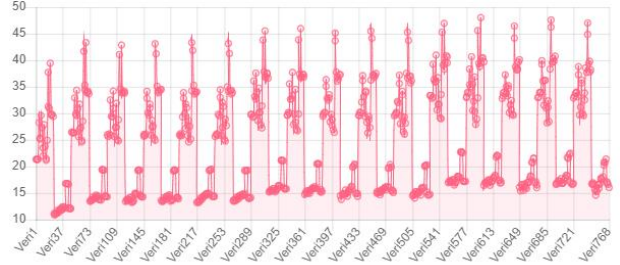
Y1 problem parametresi için oluşturulan tahmin modelinin veri seti üzerindeki performansı Şekil 39'da ki grafikte gösterilmektedir.



**Şekil 39.** Y1 (9. Problem Parametresi) tahmin modeli performans grafiği

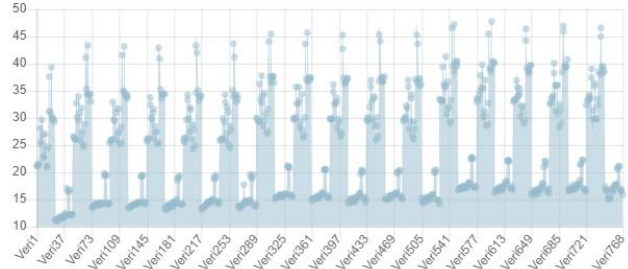
### 3.1.4.10. Y2 parametresine ait tahmin modeli grafikleri

Veri setinde bulunan 768 veri örneği için "Soğutma Yüğü (Y2)" parametresi veri dağılım grafiği aşağıdaki gibidir.



**Şekil 40.** Y2 (10. Problem Parametresi) veri dağılım grafiği

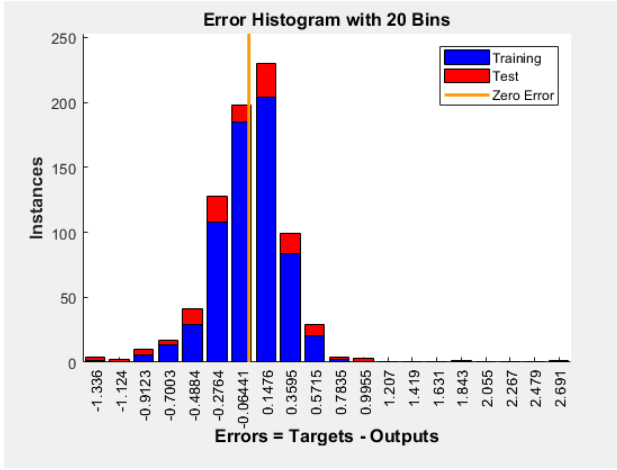
Y2 problem parametresini tahmin etmek için diğer 9 nitelikten (X1, X2, X3, X4, X5, X6, X7, X8, Y1) faydalanılmaktadır. Veri setindeki 768 veri örneği baz alınarak Y2 parametresi için bir tahmin modeli (Yapay Sinir Ağı) oluşturulmuştur. Eğitim için kullanılan veri seti, tahmin modeline gönderilmiş ve ele alınan sonuçlar Şekil 41'de ki grafik üzerinde gösterilmiştir.



**Şekil 41.** Y2 (10. Problem Parametresi) tahmin modeli sonucu veri dağılım grafiği

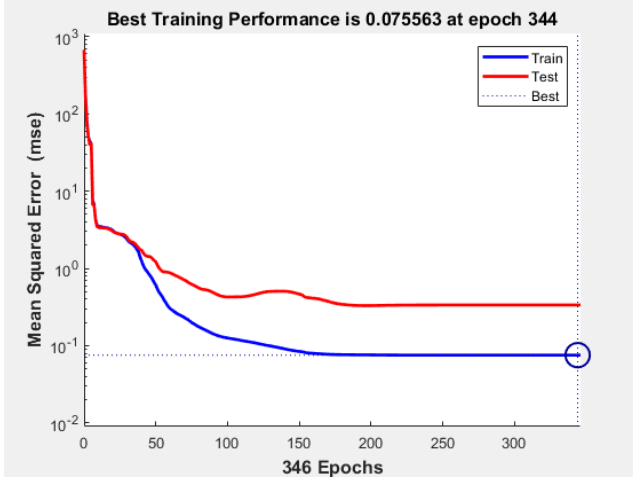
Y2 problem parametresi için oluşturulan tahmin modelinin eğitimi sırasındaki hata değişimi Şekil 42'de ki histogramda gösterilmektedir.





Şekil 42. Y2 (10. Problem Parametresi) tahmin modeli hata histogramı

Y2 problem parametresi için oluşturulan tahmin modelinin veri seti üzerindeki performansı Şekil 43'de ki grafikte gösterilmektedir.



Şekil 43. Y2 (10. Problem Parametresi) tahmin modeli performans grafiği

#### 4. Sonuç

Bu proje önerisinde yapay zekâ tabanlı veri yönetim aracının geliştirilmesi konusunda detaylı bir planlama ve analiz yapılmıştır. Yapay zekâ tabanlı veri yönetim aracının temel öğeleri ve algoritmaları tanımlanmış ve tanıtılmıştır. Bu aracın geliştirilmesine ilişkin yöntem ise detaylı bir şekilde verilmiştir. Geliştirilmesi planlanan aracın literatürde hangi boşluğu dolduracağı yani araştırmacılar açısından önemi açıkça ortaya koyulmuştur. Ortaya koyulan hedefler, geliştirilecek yazılım aracının entegre edileceği sistemleri ve internet üzerinden bağımsız çalışabilen bir uygulamayı işaret etmektedir. Bu yönüyle somut ve ölçülebilir çıktılar tanımlanmıştır. Projenin kabul edilmesi biz proje çalışanlarının yapay zekâ ile veri madenciliği konusunda planladığımız çalışmalarını yapmamız açısından son derece önemlidir. Projeyi

hayata geçirmemiz, mezuniyet sonrası yapay zekâ alanında faaliyet gösteren yazılım firmalarında işe girmemiz açısından önemli bir referans ve motivasyon kaynağı olacaktır.

#### Teşekkür

Bu çalışmanın konusunun belirlenmesinde ve hazırlanma sürecinin her aşamasında değerli bilgilerini ve zamanını benden esirgemeyerek her fırsatta çalışmamla yakından ilgilenen, eleştirileriyle yol gösteren danışman hocam Doç. Dr. Hamdi Tolga KAHRAMAN' a teşekkür ve minnetimi özellikle belirtmek istiyorum.

#### Kaynakça

- [1] Deloitte. "Veri Analizi- Veri Kalitesi ve Bütünlüğü". <http://www.denetimnet.net/UserFiles/Documents/Makaleler/BT%20Denetim/Veri Analizi Veri Kalitesi ve Bütünlüğü.pdf> , 13 Ekim 2018.
- [2] Çelik, Y., Sezgin, E., "Veri Madenciliğinde Kayıp Veriler İçin Kullanılan Yöntemlerin Karşılaştırılması". Akdeniz Üniversitesi, <http://ab.org.tr/ab13/bildiri/184.pdf> , 14 Ekim 2018.
- [3] Çüm, S., Gelbal, S. (2015). "Kayıp Veriler Yerine Yaklaşık Değer Atamada Kullanılan Farklı Yöntemlerin Model Veri Uyumu Üzerindeki Etkisi", Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi, 35, 87-111.
- [4] Çüm, S., Demir, E.K., Gelbal, S., Kışla, T. (2018). "Kayıp Veriler Yerine Yaklaşık Değer Atamak İçin Kullanılan Gelişmiş Yöntemlerin Farklı Koşullar Altında Karşılaştırılması", Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi, 45, 230-249.
- [5] Kahraman, H. T., Bayindir, R., & Sagioglu, S. (2012). A new approach to predict the excitation current and parameter weightings of synchronous machines based on genetic algorithm-based k-NN estimator. Energy Conversion and Management, 64, 129-138.
- [6] Karaboga, D., & Akay, B. (2009). A comparative study of artificial bee colony algorithm. Applied mathematics and computation, 214(1), 108-132.
- [7] Cheng, M. Y., & Prayogo, D. (2014). Symbiotic organisms search: a new metaheuristic optimization algorithm. Computers & Structures, 139, 98-112.
- [8] Holland, J.H., (1975). "Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence". Q. Rev. Biol. 1, 211. <http://dx.doi.org/10.1086/418447>.

- [9] Kahraman, H. T. (2016). A novel and powerful hybrid classifier method: Development and testing of heuristic k-nn algorithm with fuzzy distance metric. *Data & Knowledge Engineering*, 103, 44-59.
- [10] A. Tsanas, A. Xifara: 'Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools', *Energy and Buildings*, Vol. 49, pp. 560-567, 2012.
- [11] Arslan, F.,(2019), 'Yapay Zekâ Tabanlı Büyük Veri Yönetim Aracının Tasarımı ve Uygulaması', Karadeniz Teknik Üniversitesi Lisans Bitirme Tezi.

## Discharge Estimation by Drainage Area Ratio Method at Some Specific Discharges for 2251 Stream Gauging Station in East Black Sea Basin, Turkey

Fatih SAKA<sup>\*1</sup>, Hasan Törehan BABACAN<sup>2a</sup>

<sup>1</sup>Karabük University, Faculty of Engineering, Department of Civil Engineering, 78000, Karabük, Turkey

<sup>2</sup> Amasya University, Taşova Yüksel Akın Vocational School, Transportation and Traffic Services Program, 05845, Amasya, Turkey

(Alınış / Received: 31.12.2018, Kabul / Accepted: 30.08.2019, Online Yayınlanma / Published Online: 31.08.2019)

### Keywords

*Drainage area ratio method  
Flow duration curves  
Regression analysis  
Discharge estimation  
Ungauged basins*

**Abstract:** Flow duration curves (FDC) showing exceedance probabilities across flow value frequencies at a specific time can be used for determining flow values having certain exceedance probability. But FDC has not been obtained at the place where flow measurements are not existed or insufficient, so discharges having certain exceedance probabilities have been estimated by different methods. In this study, drainage area ratio (DAR) method was considered due to obtaining discharge values at ungauged sites for feasibility studies. Three basins having similar hydrologic characteristics were chosen for this purpose. Discharges having %3, %20, %30, %50 and %90 exceedance probabilities were selected for the estimation. Three basins are 2251 stream flow gauging station (SGS), 22-88 SGS, and 22-61 SGS in the Black Sea Basin, Turkey. The 22-88 and 22-61 SGSs are included in the area of 2251 SGS. If discharges having special exceedance probabilities need to be determined, firstly it is considered to investigate whether the DAR method can be used. DAR method is mainly depended on a regression analysis between the project site and the other basins/subbasins which have long-term data. %3, %20, %30, %50 and %90 exceedance probabilities on yearly scale discharges were obtained from 10 years of data from each SGS. Regression analysis was applied between flow records and drainage areas of the SGSs. Flows having exceedance probabilities of 3, 20, 30, 50 or 90 percent and their regression equations were analyzed and absolute relative error of them were investigated. After analysis, determination coefficients of discharges having %3, %20, %30, %50 and %90 exceedance probabilities were obtained as 0.898, 0.807, 0.834, 0.782 and 0.892, respectively. Mean absolute relative errors were 18.88, 7.28, 5.83, 2.14 and 1.53, respectively. According to discharges for the considered exceedance probabilities, errors can be dangerous level for %3 exceedance probability, but the other exceedance probabilities, DAR method can be useful for feasibility studies. Consequently, it has been found that errors of the DAR method can't be convenient for all percent values of exceedance probabilities. Using for higher discharges having smaller exceedance probabilities can cause more errors.

### 1. Introduction

Many water resources projects and models are applied by researchers and practitioners worldwide. These models and projects are needed to simulate reservoir/river system operations for flood control, water supply, hydropower generation, etc. The streamflow input datasets capture the hydrologic

characteristics of a river basin, including severe multiple-year droughts, major floods, and the full range of more normal flow fluctuations [1].

In most of the water resources engineering applications, water quantities at the project site or discharge probabilities having certain values are required but flow measurements are limited and

available only at the streamflow gauging stations (SGS). Flow duration curves (FDC) showing the exceedance probabilities across flow value frequencies at a specific time can be used for determining flow values having certain exceedance probability. However, FDC or discharges having certain exceedance probabilities need to be estimated by using different methods [2] at places where FDC has not been obtained or flow measurements do not exist sufficiently.

In 2003, the International Association for the Hydrological Sciences (IAHS) launched an initiative focused on Prediction in Ungauged Basins (PUB). The PUB initiative is aimed at engaging the scientific community in a cohesive effort to advance the understanding and prediction capability of hydrologic parameters in ungauged basins. By defining ungauged basins as those that lack sufficient length or quality of recorded data, the prediction is understood to include reconstruction of past events, prediction of future and passed magnitudes and forecasting, the coupling of certain magnitudes with particular points in time [3].

When flow values are requested at an ungauged site, traditional streamflow transfer techniques can be applied. Mainly, four types of flow transfer techniques are named as standardizing flows by drainage area, standardizing flows by mean flows, standardizing with the maintenance of variance extension (MOVE) and the use of the FDC. The use of drainage area ratio (DAR) is the most common and is appealing as it requires no additional information other than the streamflows at an index site and the drainage areas of the index and ungauged sites, making it the easiest possible method that one could consider. [4]. That is, for any given month for two sites,

$$\frac{Q_y}{A_y} = \frac{Q_x}{A_x} \quad (1)$$

X and Y, with monthly streamflow Q and drainage area A. Traditionally, site X is considered the gauged site and site Y is the ungauged site [4].

Often the logarithms of streamflows are better behaved than the flows in real space. For this reason, it is important to consider the log-space transformation (?) of the DAR. It recognizes that one could standardize the logarithms of flows by the logarithm of the drainage areas such that, for any given month,

$$\frac{\ln(Q_y)}{\ln(A_y)} = \frac{\ln(Q_x)}{\ln(A_x)} \quad (2)$$

for the two sites X and Y. Solving the equation above yields,

$$Q_y = Q_x \frac{\ln(A_y)}{\ln(A_x)} \quad (3)$$

which allows for the streamflow at the ungauged site to be estimated directly [4].

The basic equation of the DAR had been modified with bias correction factor and exponent parameter as followings:

$$Q_y = K \left( \frac{A_y}{A_x} \right)^\phi Q_x \quad (4)$$

Where  $Q_y$  is streamflow at an ungauged site,  $Q_x$  is streamflow at a gauged site,  $A_y$  is the drainage area of the ungauged site,  $A_x$  is the drainage area of the gauged site,  $K$  is a bias correction factor, and  $\phi$  is an exponent parameter. The bias correction factor is estimated using the nonparametric method described by Duan (1983), and the exponent parameter is calculated by a regression equation between flow sequence and drainage area at a gauged site [5].

The DAR method is also one of the linear equations, and this can be simply represented as:

$$Y = BX \quad (5)$$

Assuming Y is  $Q_y$ , and X is  $Q_x$ . B can be written as:

$$B = K \left( \frac{A_y}{A_x} \right)^\phi \quad (6)$$

“B” can be estimated by linear regression equation without intercepts, made from selected primary control point pairs with high correlation coefficient (more than  $r=0.91$ ) [1].

All discharge transfer methods applications are needed to be highly correlated data between the discharge and drainage area, this also shows that, have same hydrologic characteristics areas, have same discharge transferring characteristics.

Flow Duration Curve (FDC) characterizes the relationship between the amount and frequency of the flows in an SGS for a certain period of time. FDCs can be obtained by plotting discharges on the vertical axis and time percentages on the horizontal axis by using flow course line belonging to the mentioned station, by calculating the time percentage in which the discharge equals or exceeds a certain value.

To obtain FDCs, discharges are ranged from the largest to the smallest and the probability of exceedance can be calculated through Weibull formula below.

$$p = i/(n + 1) \quad (7)$$

In the formula, i indicates the sequence number of sorted discharges and n indicates a total number of

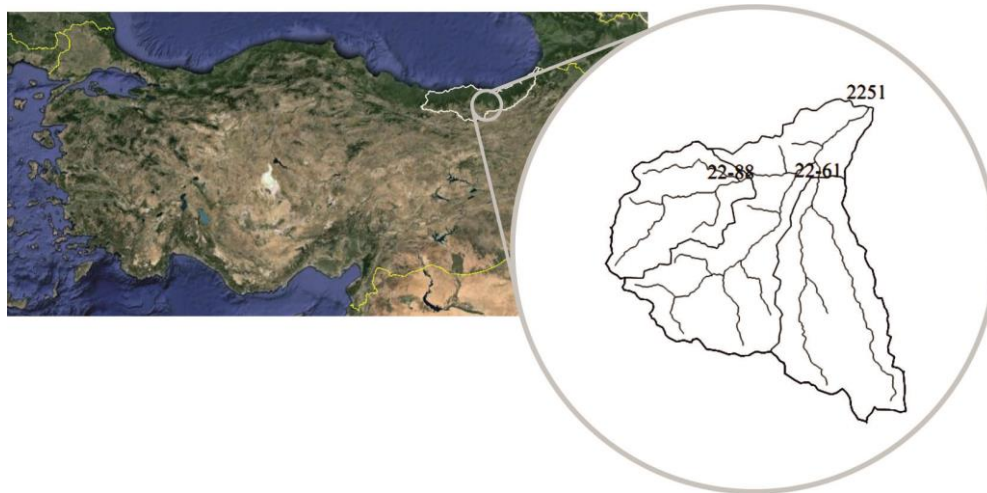
data. Duration characteristics of a river network are determined by using FDCs. However, the time period considered in the discharges to be used in the formation of FDC affects the shape of FDC, so FDCs differ from each other based on the duration (i.e daily, monthly or any other). The characteristics of discharge data also change depending on the time period of the study. FDC can be formed either by using all discharge values measured during recording period or by drawing separate FDC for each year, and these are called annual FDC., Comparing annual FDC enables to see year-by-year changes in the river flows [2].

## 2. Materials and Method

River and reservoir system analysis models start with homogeneous sequences of monthly or daily streamflow volumes covering a hydrologic period-of-

analysis at relevant sites. Homogeneous means that the flows represent a specified uniform condition of watershed and river system development, long-term climate, and water use. Non-homogeneities in historical gauged stream flows are typically caused primarily by the construction of reservoir projects, growth or changes in water use, and other changes in water management practices over time. However, watershed land-use changes, climate changes, and other factors may also affect the stationarity of recorded streamflow measurements [1]. Homogeneity of data used was carefully tested.

There are three basins in the study area. Those are indicated by streamflow gauging station (SGS) 2251, 22-88, and 22-61 in the Eastern Black Sea Basin, Turkey. The 22-88 and 22-61 SGSs form sub-basins of 2251 SGS (Figure 1). Input datasets are 10-yr mean daily discharge measurements and drainage areas for each basin.



**Figure 1.** The Eastern Black Sea Basin, Turkey, and the configuration of three basins

Determination of discharges having 3, 20, 30, 50 and 90 percent exceedance probabilities was done using FDC. To derive these, 10-yr daily discharges of 2251, 22-61, and 22-88 SGSs are used. Drainage areas and tributaries were determined using NetCAD. Daily mean discharge values ( $m^3 s^{-1}$ ) of each SGSs were gathered from relevant institutions (General Directorate of State Hydraulic Works, General Directorate of Electric Power Resources Survey and Development Administration). Periods of annual discharge records and drainage areas are presented in Table 1.

**Table 1.** Periods of annual daily discharges and drainage areas of SGSs

	Period of Records	Area (km <sup>2</sup> )
22-61	1990, 1992, 1995-2002	260.1
22-88	1987-1989, 1993-1999	154.9
2251	1996-2005	726.5

Discharges having exceedance probabilities of 3, 20, 30, 50 and 90 percent have been used in this research.

Minimum, maximum and mean values of these discharges are shown in Table 2.

**Table 2.** Discharges having exceedance probabilities of 3, 20, 30, 50 and 90 percent

	Discharges	
	Min.-Max	Mean
<b>Q3</b>	5.20-67.00	24.60
<b>Q20</b>	1.72-30.60	10.55
<b>Q30</b>	1.05-18.70	6.31
<b>Q50</b>	0.48-7.90	3.15
<b>Q90</b>	0.14-4.12	1.58

Table 2 shown the discharge values evaluation according to exceedance probabilities. If there is a correlation between discharge values and major basin characteristics, DAR is considered to be usable for these areas. According the same hydrological characteristics 2251, 22-61 and 22-88 SGSs were used in this paper.

### 3. Results

Correlation and regression analysis results are given below for the 2251 SGS. Determination coefficients of discharges having %3, %20, %30, %50 and %90 exceedance probabilities are obtained as 0.807, 0.834, 0.782, 0.838 and 0.892, respectively.

For the coefficient of determination values less than 0.700, relations between discharges and drainage areas are considered poor and regression equations are not derived. Solely coefficient of determination value is not sufficient to decide the success of the regression analysis. Therefore mean absolute relative error used for evaluating of regression analysis results. Mean absolute relative errors for discharges having %3, %20, %30, %50 and %90 exceedance probabilities are 18.88, 7.28, 5.83, 2.14 and 1.53, respectively and some characteristics of them are shown in Table 3.

**Table 3.** Values of % relative absolute errors for discharges having exceedance probabilities of 3, 20, 30, 50 and 90 percent

	% Relative Absolute Errors		
	Min.-Max	Mean	Max. Mean Rel. Error
<b>Q3</b>	0.56-113.64	28.18	18.88
<b>Q20</b>	0.60-157.34	23.80	7.28
<b>Q30</b>	2.10-176.32	31.17	5.83
<b>Q50</b>	0.40-203.00	27.05	2.14
<b>Q90</b>	0.11-294.15	37.19	1.53

Table 4 contains regression equations which were created using the independent variable drainage area (A). Table 4 can use correction for mean absolute error thus, it was aimed at improving the reliability of the study.

**Table 4.** Regression equations of discharges having exceedance probabilities of 20, 30, 50 and 90 percent

<b>Q20</b>	<b>Q30</b>	<b>Q50</b>	<b>Q90</b>
$0.221 + 0.027 \cdot A$	$0.558 + 0.015 \cdot A$	$0.283 + 0.008 \cdot A$	$-0.154 + 0.005 \cdot A$

### 4. Conclusions and Suggestions

It is found that according to discharges for the considered exceedance probabilities, errors can be dangerous level for %3, but for the other exceedance probabilities, the DAR method can be useful for feasibility studies. Thus, DAR method can not be convenient for all percent values of exceedance probabilities. Using for higher discharges having smaller exceedance probabilities can cause more errors. It is shown that the DAR method is suitable for discharge estimation but method can not suitable for estimation of higher discharge values. After all, the DAR method is useful for 2251 SGS for some discharges having exceedance probabilities of 20, 30, 50 and 90 percent.

### REFERENCES

[1] Ryu, "Developing homogeneous sequences of river flows and performing comparative analyses of flow characteristics," Ph.D. thesis, Texas A&M University, 2015.

[2] Saka, "Determination of synthetic flow duration curves by using mathematical methods and a case study in the eastern black sea," Ph.D. thesis, Graduate School of Natural and Applied Sciences, Karadeniz Technical University, Turkey (in Turkish).

[3] M. Sivapalan, K. Takeuchi, S.W. Franks, V.K. Gupta, H. Karambiri, V. Lakshmi, X. Liang, J.J. McDonnell, E.M. Mendiando, P.E. O'Connell, T. Oki, J.W. Pomeroy, D. Schertzer, S. Uhlenbrook and E. Zehe, "IAHS decade on prediction in ungauged basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences," Hydrological Sciences Journal, 48.6: 857-880, 2003.

[4] Farmer, "Estimating monthly time series of streamflows at ungauged locations in the united states," MEng. thesis, Department of Civil and Environmental Engineering, Tufts University, 2012.

[5] D.G. Emerson, A.V. Vecchia, and A.L. Dahl, "Evaluation of drainage-area ratio method used to estimate streamflow for the Red River of the North Basin, North Dakota and Minnesota," US Geological Survey Scientific Investigative Report 5017, 2005.

## Renk Kaydırmalı Anahtarlama Modülasyonunun Görünür Işık Haberleşme Sistemlerinde Başarımı

Emin TUĞCU\*<sup>1</sup>

<sup>1</sup>Karedeniz Teknik Üniversitesi, Of Teknoloji Fakültesi, Elektronik ve Haberleşme Mühendisliği Bölümü, 61830, Trabzon

(Alınış / Received: 04.08.2019, Kabul / Accepted: 26.08.2019, Online Yayınlanma / Published Online: 31.08.2019)

### Anahtar Kelimeler

Renk kaydırmalı anahtarlama, Görünür ışık haberleşmesi, Bit hata oranı

**Özet:** Bilgi ve iletişim teknolojilerindeki gelişmeler, kullanıcı yoğunluğunun hızlı bir şekilde artması ve bilgiye daha hızlı erişim ihtiyaçlarından dolayı mobil haberleşmeye ayrılan radyo frekans bantları hızla dolmakta ve üst sınırlarına doğru yaklaşmaktadır. Radyo frekans bantlarıyla girişim oluşturmayan görünür ışık bandını kullanan görünür ışık haberleşmesi bu problemin çözümüne uygun bir aday olarak görülmektedir. Bu çalışmada, görünür ışık haberleşme sistemlerinde kullanılan renk kaydırmalı anahtarlama modülasyonunun başarımları analizleri yapılmıştır.

## Performance of Color Shift Keying Modulation in Visible Light Communication Systems

### Keywords

Color shift keying, Visible light communication, Bit error rate

**Abstract:** Due to the developments in information and communication technologies, the rapid increase in user density and the need for faster access to information, the radio frequency bands allocated to mobile communications are rapidly filling and approaching to their upper limits. Visible light communication using visible light bands that do not interfere with radio frequency bands is seen as a suitable candidate for the solution of this problem. In this study, performance analysis of color shift keying modulation used in visible light communication systems is performed.

### 1. Giriş

İletişim teknolojilerinde yaşanan hızlı değişim ve gelişmeler yüksek hızlı kablosuz haberleşme sistemlerini günlük hayatımızın vazgeçilmezleri durumuna getirmiştir. Kullanıcı yoğunluğunun artması, bilgiye daha hızlı erişim ihtiyacının artması veri trafiğinin artmasına neden olmuş ve bunun sonucunda mevcut olarak kullanılan radyo frekans bandında kapasite problemini oluşturmuştur. Bu problemin bir çözümü olarak radyo frekans bandıyla girişim oluşturmayan lisansız görünür ışık bandının kullanımı ciddi bir aday olarak görülmektedir. Günlük hayatımızın önemli ve vazgeçilmez bir parçası olan ışığın haberleşme sistemlerinde bilgi taşımak amacıyla kullanılması yenilikçi, düşük maliyetli ve enerji tüketimi açısından etkin bir haberleşme

teknolojisinin ortaya çıkması anlamına gelmektedir. Görünür ışık bandının haberleşme amacıyla kullanımını amaçlayan görünür ışık haberleşmesi (Visible Light Communication, VLC) yeni nesil haberleşme sistemleri için önemli bir tamamlayıcı olarak görülmektedir [1]. VLC sistemleri yüksek veri hızı avantajına sahip, sağlığa zararı olmayan, güvenli, maliyet etkin, lisansız çok büyük bant genişliğine sahip, sinyal izolasyon özelliğine sahip kompakt yapısı ile veri iletiminde daha az güç harcayan yeni nesil bir teknolojidir.

Günümüzde hükümetler enerji tüketiminin azaltılması konusunda önemli rol üstlenmektedirler. Örneğin US hükümeti yakın zamanda güç tüketimi yüksek olan akkor ve floresan lambaları verimli, ve güç tüketimi düşük olan LED'ler ile değiştirme kararı almıştır. Verimliliği yaklaşık olarak 15 lm/W olan



geleneksel 60–100 W akkor ışık ampülleri ile karşılaştırıldığında LED'ler 150 lm/W değerinde bir verimliliğe sahiptir [2]. Dünyadaki tüm aydınlatma kaynaklarının LED'ler ile değiştirilmesi durumunda aydınlatma için harcanan toplam küresel güç tüketiminin 50 % oranında azalacağı tahmin edilmektedir.

VLC sistemlerinde de son yıllarda aydınlatmada yaygın bir kullanım alanı olan LED'ler verici ve foto-dedektörler ise alıcı olarak kullanılır. LED'lerin geleneksel aydınlatma kaynaklarına göre yüksek verimlilikleri aydınlatma alanında ilgi görmelerine neden olmaktadır. VLC sistemleri yoğunluk modülasyonu/doğrudan sezim (intensity modulation/direct detection, IM/DD) yöntemi kullanılarak gerçekleştirilmektedir. Veri iletimi LED'ler tarafından yayılan ışığın yoğunluğunun değiştirilmesi ile sağlanır. IM/DD tekniğinde bilginin taşınması için sadece ışık şiddetinin kullanılması, iletimde kullanılan optik sinyallerin, RF sistemlerinden farklı olarak, gerçel ve pozitif değerli olmasını gerektirmektedir.

VLC sistemlerinde, aydınlatma ile birlikte haberleşmenin de sağlanması, veri iletiminde kullanılan LED'ler tarafından yayılan ışığın beyaz ışık olmasını gerektirmektedir. Beyaz ışık LED'lerle temel olarak iki yöntem kullanılarak üretilmektedir. Birinci yöntemde beyaz ışığın üretimi mavi LED'lerin yüzeyinin fosfor ile kaplanmasıyla sağlanır. Bu yöntem basit olmasına rağmen haberleşme hızı açısından değerlendirildiğinde modülasyon hızı düşüktür ve fosforun oluşturduğu etkiden dolayı yaklaşık olarak 1 Mhz ile sınırlıdır [3]. Mavi filtreleme yöntemi kullanılarak bu LED'lerin iletim bant genişliği 20 Mhz seviyelerine çıkarılmıştır [4]. Ayrıca fosfor tabanlı LED'lerin aydınlatma verimliliği de düşüktür. İkinci yöntemde ise kırmızı, yeşil ve mavi ışık üreten üçlü kromatik LED'lerin (RGB LED) bir araya getirilmesiyle beyaz ışığın üretimi gerçekleştirilmektedir. Bu yöntemde görülebilir ışık spektrumu yüksek modülasyon hızlarına sahip üç ayrı renk bandına bölünmektedir. Her bir renk bandı Dalgaboyu Bölme Çoğullama (Wavelength Division Multiplexing, WDM) teknolojisi kullanılarak bağımsız olarak modüle edildiğinde yüksek veri hızlarına ulaşılmaktadır [5]. RGB bantlarının her biri bağımsız olarak modüle edildiğinde beyaz ışığın üretilmesinde problem oluşmaktadır [6]. Beyaz ışığın insan gözünü etkilemeyecek şekilde üretilmesi için IEEE 802.15.7 standardında [7] renk kaydırmalı anahtarlama (Color Shift Keying, CSK) VLC için tanımlanmış bir modülasyon türüdür.

Bu çalışmada CSK modülasyonunun AWGN kanallarda VLC sistemleri için BER başarımı incelenmiştir.

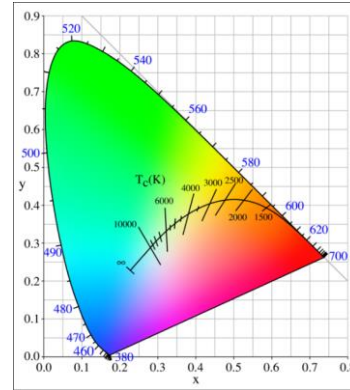
Çalışmanın sonraki kısmında ikinci bölümde CSK modülasyonu ve demodülasyonu incelenmiştir. Üçüncü bölümde CSK modülasyonu için AWGN kanallarda elde edilen BER başarımları sonuçları verilmiştir. Çalışmanın son bölümde ise elde edilen sonuçlar irdelenmiştir.

## 2. CSK Modülasyonu ve Demodülasyonu

CSK modülasyonunda veri iletimi için, kırmızı, yeşil ve mavi (RGB) gibi çok renkli LED'ler tarafından yayılan görünür ışığın yoğunluğu modüle edilir. CSK modülasyonu CIE 1931 renk uzayında [8] uluslararası aydınlatma komisyonu tarafından tanımlanan x-y renk koordinatları kullanılarak gerçekleştirilir. Şekil 1'de görüldüğü gibi tanımlanan renk uzayında tüm renkler 2 boyutlu uzayda x-y koordinatları ile temsil edilirler. x ve y koordinatlı düzlem renklerin kromatik değerlerine karşılık gelmektedir. CSK sinyalinin üretilmesi için üç farklı ışık kaynağı kullanılmalıdır. CSK modülasyonunda işaret yıldız diyagramları üçgen yapıya sahiptir.

VLC sistemlerinde haberleşmenin yanında aydınlatmanın da sağlanmasından dolayı CSK sinyallerinde bazı kısıtlamalar vardır. Verici kısımda ilgili renk bantlarındaki LED'lerden yayılan optik gücün daima pozitif değerli olması gerektiğinden iletimi yapılacak olan işaret vektörüne  $\mathbf{s} = [s_i, s_j, s_k]^T$  karşılık gelen güç değerleri  $[P_i, P_j, P_k]^T$  vektörü aşağıda (1) ile verilen şartları sağlamalıdır.

$$P_i \geq 0, P_j \geq 0, P_k \geq 0 \quad (1)$$

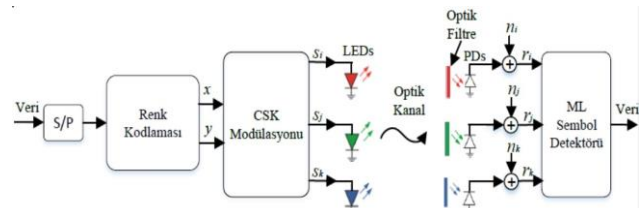


Şekil 1: CIE 1931 renk uzayı kromatik diyagramı [8]

Kırpışma etkisinin meydana gelmemesi için LED'lerden yayılan işaretin toplam optik gücünün bir sembol süresi boyunca (2) eşitliğinde verildiği gibi sabit olması gerekmektedir.

$$P_i + P_j + P_k = P_T \quad (2)$$

CSK sistemine ait blok diyagramı Şekil 2' de verilmiştir.



Şekil 2: CSK Sisteminin Blok Diyagramı



Verici tarafta ilk işlem olarak gelen ikili veri renk kodlaması yapılarak CIE 1931 renk uzayında x-y kromatik değerleri ile tanımlanan M seviyeli sembollere dönüştürülür [7]. Sonrasında ise dönüştürülen kromatik değerlere karşılık gelen R, G, B renk bantlarına  $\{i, j, k\}$  ait optik güç yoğunlukları (3), (4) ve (5) ile verilen denklem sistemi çözülerek elde edilir [9].

$$P_i x_i + P_j x_j + P_k x_k = x \quad (3)$$

$$P_i y_i + P_j y_j + P_k y_k = y \quad (4)$$

$$P_i + P_j + P_k = 1 \quad (5)$$

Denklem sisteminde yer alan  $(x_i, y_i)$ ,  $(x_j, y_j)$  ve  $(x_k, y_k)$  kromatik değerleri CSK sisteminde kullanılan 3 farklı ışık kaynağının (kırmızı/yeşil/mavi, RGB) merkez dalga boylarına karşılık gelen x-y kromatik değerleridir. Sembol uzayında bulunan her bir sembole ait hesaplanan  $[P_i, P_j, P_k]^T$  yoğunluk bilgileri işaret uzayında  $s_l = [s_{i,l}, s_{j,l}, s_{k,l}]^T$  ( $0 \leq l \leq M - 1$ ) olmak üzere 3-boyutlu sembolleri oluştururlar. Alıcı ile verici arasındaki optik kanal ortamı, kanallar arasındaki girişimin ve optik filtreleme kayıplarının ihmal edildiğinde AWGN kanal olduğu durum için alıcı tarafta alınan işaret (6)'da verildiği gibi yazılabilir.

$$\begin{bmatrix} r_i \\ r_j \\ r_k \end{bmatrix} = \begin{bmatrix} s_i \\ s_j \\ s_k \end{bmatrix} + \begin{bmatrix} \eta_i \\ \eta_j \\ \eta_k \end{bmatrix} \quad (6)$$

Burada  $\eta_i$ ,  $\eta_j$  ve  $\eta_k$  he birinin varyansı  $\sigma^2$  olan beyaz gürültüdür. Alıcıda ilk olarak RGB iletim bantlarından gelen işaretler foto detektörler tarafından alınarak optik işaretler elektriksel işaretlere dönüştürülürler. Alınan işaret vektörü  $r = [r_i, r_j, r_k]^T$  3-boyutlu işaret uzayında (7) eşitliği ile verildiği gibi en küçük öklid mesafesine göre en büyük olabilirlik (Maximum Likelihood, ML) sezici ile iletilen sembollere karar verilir.

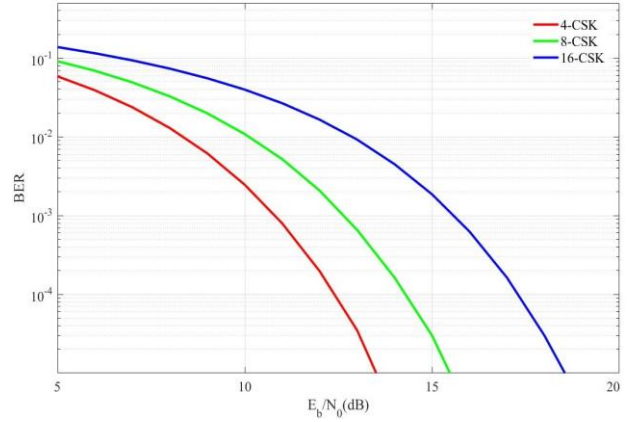
$$\hat{s} = \underset{s_l \in S}{\operatorname{argmin}} \|r - s_l\|_2 \quad (7)$$

Burada S, 3-boyutlu uzayda M seviyeli CSK sembollerinde oluşan kümeye karşılık gelir.

### 3. Başarım Sonuçları

Şekil 2' de 4-CSK, 8-CSK ve 16-CSK modülasyonları için elde edilen sonuçlar verilmiştir. Beklendiği gibi modülasyon seviyesi arttıkça semboller arasındaki öklid mesafeleri azalacağından gürültünün etkisi başarımlar üzerinde daha fazla etkili olmaktadır ve sistemin başarımlarının düştüğü BER sonuçlarından görülmektedir.  $1e-4$  BER seviyesine 4-CSK modülasyonu için yaklaşık işaret gürültü oranı (Signal to Noise Ratio, SNR) SNR=15.4 dB seviyesinde ulaşılmaktadır. 8-CSK ve 16-CSK modülasyonları

kullanıldığında ise aynı BER seviyesine yaklaşık olarak sırasıyla 19 dB ve 23.2 dB SNR değerlerinde ulaşılmaktadır. Aynı BER seviyesinde 4-CSK ile 8-CSK modülasyonu arasında yaklaşık 4 dB' lik bir fark görülmektedir. 4-CSK ile 16-CSK arasında ise yaklaşık 8dB'lik bir fark oluşmaktadır.



Şekil 2: 4-CSK, 8-CSK ve 16-CSK modülasyonları için benzetim BER değişimleri

### 4. Sonuçlar

Bu çalışmada yeni nesil haberleşme teknolojilerinden olan VLC sisteminde CSK modülasyonun AWGN kanallar için BER başarımları elde edilmiştir. Elde edilen sonuçlardan modülasyon seviyesi yükseldikçe semboller arasındaki öklid mesafesinin azalmasıyla beklenildiği gibi BER başarımlarının düştüğü görülmektedir.  $1e-4$  BER seviyesinde 4-CSK modülasyonu ile 8-CSK ve 16-CSK modülasyonları arasında sırasıyla yaklaşık olarak 4 dB ve 8 dB' lik bir SNR farkı oluşmaktadır.

### Kaynakça

- [1] Ghassemlooy, Z., Alves, L. N., Zvanovec, S., & Khalighi, M.-A. (2017). Visible Light Communications: Theory and Applications.
- [2] Sau Koh, Willem Van Driel, & G. Q. Zhang, Degradation of light emitting diodes: a proposed methodology (2011). Journal of Semiconductors, 32.
- [3] Chow, C.W., Yeh, C.H., Liu, Y.F., & Liu, Y. (2011). Improved modulation speed of LED visible light communication system integrated to main electricity network, Electronics Letters, 47,15,867-868.
- [4] Cui, K.; Chen, G.; He, Q., & Xu, Z., Indoor optical wireless communication by ultraviolet and Visible Light (2009). Proc. SPIE 7464, Free-Space Laser Communications IX.
- [5] Cossu, G., & Khalid, A. M., Choudhury, P., Corsini, R., & Ciaramella, E. (2012). 3.4 Gbit/s visible optical wireless transmission based on RGB LED, Optics Express, 20,26, B501-B506.

- [6] Jia, L., Wang, J.-Y., Zhang, W., Chen, M., & Wang, J.-B. (2015). Symbol error rate analysis for colour-shift keying modulation in visible Light communication system with RGB light-emitting diodes, *IET Optoelectronics*, 9,5,199-206.
- [7] IEEE Standard for Local and Metropolitan Area Networks-Part 15.7:Short-Range Wireless Optical Communication Using Visible Light (2011), IEEE Std 802.15.7-2011, 1-309.
- [8] CIE, Commission Internationale de l'Eclairage Proc. (1931). Cambridge University Press.
- [9] Tang, J., Zhang, L., & Wu, Z. (2017). Exact Bit Error Rate Analysis for Color Shift Keying Modulation, *IEEE Communications Letters*, 99,1-4.

