JIET

Journal Of Investigation On Engineering & Technology

VOLUME/ CİLT: 5 (2022)

ISSUE: 1

ISSN: 2687-3052

KTÜ 1955 KARADENİZ TEKNİK ÜNİVERSİTESİ

Journal of Investigation on Engineering & Technology

http://dergipark.gov.tr/jiet

INDEXING:
İndeksler

Product Information
Ürün Bilgisi

Volume *Cilt* **5** **Issue** *Sayı* **1** **June/2022** *Haziran/2022*

Publisher *Yayıncı*	Karadeniz Technical University, Faculty of Technology *Karadeniz Teknik Üniversitesi Of Teknoloji Fakültesi*
Web Page *Web Sayfası*	http://dergipark.gov.tr/jiet
Date of Publication *Basım Tarihi*	June /2022 *Haziran/2022*
Language *Yayın Dili*	English/Turkish *İngilizce/Türkçe*
Frequency *Yayın Aralığı*	Published twice in a year *Yılda iki kez yayınlanır*
Type of Publication *Yayın Türü*	Periodical *Süreli yayın*
ISSN Number *ISSN Numarası*	2687-3052

Yazışma Adresi
Karadeniz Teknik Üniversitesi
Teknoloji Fakültesi
Çamlı M. Hacı Mehmet Baheddin Ulusoy
Cad. No:144 61830 Of / TRABZON
Telefon: +90 462 377 83 01
E-posta: jiet@ktu.edu.tr

Correspondence Address
Karadeniz Technical University
Faculty of Technology
Çamlı M. Hacı Mehmet Baheddin Ulusoy
St. No:144 61830 Of / TRABZON
Phone: +90 462 377 83 01
E-mail: jiet@ktu.edu.tr

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

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

CONTENTS

İçindekiler

Journal of Investigations on Engineering & Technology
Volume 5, Issue 1, 1-10, 2022

ISSN: 2687-3052

Araştırma Makalesi

Elektrik Güç Sistemlerindeki Kaçak Kullanımların Yapay Sinir Ağları ile Tahmini

Enes YILDIZ [1], Nurettin ÇETİNKAYA [2]

(Alınış / Received: 19.03.2022, Kabul / Accepted: 30.06.2022, Online Yayınlanma / Published Online: 30.06.2022)

Anahtar Kelimeler
Kayıp
Kaçak
Dağıtım hatları
Güç sistemleri
Enerji
Yapay sinir ağları (YSA)

[3] Öz: Üretimi maliyetli olan elektrik enerjisinin doğru kullanılması önemli bir konudur. Ülkemizdeki önemli sorunlardan birisi olan kaçak elektrik kullanımların tahmin edilmesi, enerji piyasasında çözülmesi gereken bir problemdir. Birlikte değerlendirilen kayıp ve kaçak kullanım miktarları, birbirinden farklı terimlerdir ve ayrı ayrı değerlendirilmesi gerekmektedir. İki terimin ayrı değerlendirilmesi, herhangi bir şebeke üzerinde alınacak önlemlerin veya yapılacak yatırımların daha sağlıklı olmasını sağlayacak ve elektriğe bakış açısını da olumlu yönde değiştirecektir. Bu sayede bölgelere yapılacak yatırımların, kaçak ile mücadele üzerine mi yoksa kayıpların azaltılması üzerine mi yapılması gerektiği hakkında bilgi verecektir. Kaçak kullanımların tahmin edilmesi sosyo-ekonomik yönden ve kaçak kullanımlarla mücadele kapsamında büyük gelişmelerin önünü açacaktır.
Bu çalışmada kaçak kullanım miktarlarının tahmin edilmesi yapay sinir ağları ile gerçekleştirilmiştir. Tahmin yapılması istenilen şehirlerin şebekeleri tek bir trafo bölgesine indirgenerek yapay sinir ağlarında tüm kayıpların ve kaçak kullanım miktarlarının tahmin edilmesi sağlanmıştır.
Literatürde kaçak kullanımların doğal kayıplardan ayrı değerlendirilmesiyle ilgili pek çalışma olmasa bile, bu çalışma ile kaçak kullanımların tahmin edilebilmesi önerilen algoritmalar ile mümkün hale gelmiştir. Önerilen algoritmalar yardımıyla doğal kayıplar ve kaçak kullanımlar birbirlerinden ayrılacak şekilde tahmin edilmiştir. Sonuçlar farklı eğitim fonksiyonları ile eğitilen yapay sinir ağları yardımıyla da desteklenmiştir. Hem şehirlerin kaçak kullanım miktarları tahmin edilmiş, hem de eğitim fonksiyonlarının performansları karşılaştırılmıştır.

Forecasting Of İllegal Consumption İn Electrical Power Systems

Keywords
Loss
Illegal use
Distrubition lines
Power systems
Energy
Artificial neural networks (ANN)

Abstract: Correctly using electrical energy, which is costly to produce, is essential. Estimating illegal electricity, which is one of the critical problems in our country, usage is a problem that needs to be solved in the energy market. Loss and illegal use, considered together, are different terms and should be evaluated separately. Evaluating each term separately will ensure that the measures to be taken or the investments to be made on any network will be healthier and positively change the perspective on electricity. In this way, it will provide information on whether the investments to be made in the regions should be made to fight against illegality or

[1] Afyon Kocatepe Üniversitesi, Mühendislik Fakültesi, Elektrik Mühendisliği, Ahmet Necdet Sezer Kampüsü, Afyonkarahisar, Türkiye, enesyildiz@aku.edu.tr
[2] Konya Teknik Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Elektrik-Elektronik Mühendisliği, Alaeddin Keykubat Kampüsü, Konya, Türkiye, ncetinkaya@ktun.edu.tr

Sorumlu yazar: Enes YILDIZ, enesyildiz@aku.edu.tr

reduce losses. Estimating illegal uses will pave the way for significant developments in socio-economic terms and the fight against illicit uses.
The estimation of the amount of illegal usage was carried out with artificial neural networks. Reducing the cities' networks to be estimated to a single transformer zone ensures that the amount of illegal use in artificial neural networks is estimated.
Although there are not many studies evaluating illegal uses separately from natural losses, this study has made it possible to predict illegal uses with the proposed algorithms. With the help of the proposed algorithms, natural losses and illegal uses are estimated separately. The results are also supported with the help of artificial neural networks trained with different training functions. Both the amount of illicit use in the cities were estimated, and the performances of the education functions were compared.

1. Giriş

Elektrik enerjisi tüketimi geçmişten günümüze kadar ülke ekonomilerinde çok önemli rol oynamaktadır. Özellikle son zamanlarda gelişen ve büyüyen dünyada, elektrik enerjisi her alanda önemli bir rol üstlenmektedir. Enerjiye ihtiyacın yüksek seviyelerde olduğu günümüzde enerjinin kaçak kullanım şeklinde yok olması ekonomik bir yüktür. Engellenmesi ve tespit edilmesi zor olan kaçak kullanımlar, ülkemize her yıl maddi olarak ciddi anlamda yük olmaktadır [1]. Kaçak kullanım miktarı altında kaybolan enerjinin değerlendirilmesi, başta ekonomik anlamda olmak üzere birçok alanda büyük önem arz etmektedir [2]. Ekonomik Kalkınma ve İşbirliği Örgütü (OECD) ülkeleri ile karşılaştırıldığında ülkemizde meydana gelen kayıp kaçak kullanım miktarları ortalamanın yaklaşık iki katı kadardır [3,4]. Kayıp ve kaçak kullanımlar hem değerlendirme bakımından hem de elektriksel (terimsel) ifade bakımından birbirinden ayrılması gereken iki farklı terimdir. Kayıplar terimi; şebekelerden kaynaklı meydana gelen ve sıfır olması mümkün olmayan gerçekleşmesi beklenen doğal kayıplardır.

Kaçak kullanım terimi ise; sıfıra indirilmesi mümkün olan, şebekenin özelliklerine bağlı olmayan kayıplardır [5-7]. İki terimin birbirinden ayrı değerlendirilmesi, şebekelere yapılacak yatırımların daha sağlıklı olmasını sağlayacaktır. Şebekeler üzerinde yapılacak iyileştirmelerin teknik kayıplar üzerinde mi yoksa kaçak kullanımlarla mücadele üzerine mi olması gerektiğinin anlaşılmasını sağlayacaktır. Her yıl gerçekleştirilen yatırımlar, kayıp ve kaçak kullanımların ikisinin de birlikte azaltılması yönünde gerçekleştirilmektedir [7,8]. Ancak kaçak kullanımı olmayan bir bölgeye, kaçak kullanımla mücadele için bütçe ayırmak gereksizdir. Tam tersi durum doğal kayıplar içinde geçerlidir. Doğal kayıpların az, kaçak kullanımların fazla olduğu bölgeler için de şebekelerin iyileştirilmesinden ziyade, kaçak kullanımlar ile mücadele edilebilecek bir yatırım gerekmektedir. Bir şebeke üzerinde kaybolan enerjinin ne kadarının kayıp ne kadarının kaçak kullanım olduğunun bilinmesi şebekeyi daha iyi tanımlamamızı sağlayacaktır.

Bir şebeke üzerinde meydana gelen kaçak kullanım miktarı direk hesaplanamazken, şebeke üzerinde kaybolması beklenen teknik (doğal) kayıpların hesaplanabilmesi, şebekelerde kullanılan malzemelerin elektriksel özelliklerine bağlı olduğu için mümkündür [8-10]. Birlikte değerlendirilen kayıp ve kaçak kullanımların toplam miktarları (şebekede boşa giden enerji miktarı) göz önünde bulundurulduğunda, kaçak kullanımların da tahmin edilebilmesi mümkün olacaktır. Kaçak kullanım miktarı ile şebekeden kaynaklı meydana gelen teknik kayıpların en azından yaklaşık olarak bilinmesi şebekelerin sağlamlığı hakkında da bize bilgi verecektir. Kaçak kullanımlarla mücadele de ağırlık verilmesi gereken bölgeler belirlenerek daha spesifik çalışmaların yapılması mümkün olacaktır.

Kaçak kullanımların noktasal olarak belirlenmesi de birkaç yöntemle mümkün olabilir. Şebekelere yerleştirilen akıllı sayaçlar, akıllı sensörler veya haberleşme altyapıları ile PLC kullanılarak kaçak kullanımlar noktasal olarak tespit edilebilir. Ancak bu yöntemler çok maliyetlidir. Özellikle kayıp ve kaçak kullanımların aynı değerlendirildiği göz önüne alındığında; kaçak kullanımların noktasal tespiti için kurulması gereken sistemin nereye kurulacağı çok yanıltıcı olabilir. Bu çalışmada şebekelerdeki kayıp ve kaçak kullanım miktarlarının ayrı ayrı değerlendirilmesi, yapılacak olan yatırımların daha sağlıklı olmasını da hedeflemektedir.

2. Meteryal ve Metot

2.1. Kaçak kullanımların tahmini için önerilen algoritma gösterimi

Ülkemizin en büyük sorunlarından biri olan kaçak kullanım miktarlarının, ayrıştırılması ile ilgili literatürde çok az çalışmalar bulunmaktadır. Güç sistemleri üzerinde yapılan en önemli konulardan birisi meydana gelen doğal kayıpların azaltılması yönündedir. Ancak kaçak kullanımların miktarlarının da tahmin edilebilmesi için birkaç

yöntem uygulanabilir [11,12]. Enerjinin korunumu kanunu "Var olan enerji yok olamaz" kapsamında tüketicilerin kullandığı enerji başta olmak üzere teknik ve teknik olmayan kayıpların toplam miktarı, harcanan enerji miktarına eşit olmak zorundadır. Bu kapsamda kaçak kullanım miktarlarının tahmini için, abonelerin tüketmiş oldukları enerji miktarı ve teknik kayıpların hesaplanması gerekmektedir. Abonelerin harcamış oldukları enerji miktarı faturalandırılan enerji miktarı anlamına gelmektedir.

Çalışmada faturalandırılabilen (abonelerin harcamış oldukları) ve faturalandırılamayan (kayıp ve kaçak yollarla harcanan) enerji miktarları "W" ile ifade edilmiştir ve aralarındaki ilişki Denklem (1)'de belirtilmiştir. Teknik olmayan kayıplar ve teknik kayıplar ise faturalandırılamayan enerji miktarı olarak kaybolan kayıpların tamamıdır. Teknik olmayan kayıplar Denklem (2)'de belirtildiği gibi tanımlanabilir.

$$W_2\ (MWh) = W_1\ (MWh) - W_3\ (MWh) \quad (1)$$

$$W_4\ (MWh) = W_2\ (MWh) - W_5\ (MWh) \quad (2)$$

W_1: Toplam Harcanan Enerji Miktarı
W_2: Faturalandırılamayan Enerji Miktarı
W_3: Faturalandırılabilen Enerji Miktarı
W_4: Teknik Olmayan Kayıplar
W_5: Teknik Kayıplar

Teknik olmayan kayıpların tahmin edilebilmesi direk olarak mümkün değildir. Teknik olmayan kayıpların tahmin edebilmesi için öncelikle teknik kayıpların hesaplanması gerekmektedir. Bir şebekede meydana gelen teknik kayıpların büyük bir bölümünü iletken kayıpları ve trafo kayıpları oluşturmaktadır. Teknik olmayan kayıplar ise şebekelerden bağımsız, insan hatalarından dolayı meydana gelen kayıplardır. Teknik olmayan kayıpları oluşturan etmenler; faturalandırma hataları, müşteri takipsizliği, sayaç okuma hataları, dosya hataları vb. gibi insan hatalarından meydana gelen hatalar olduğu için tüm bu kayıpları kaçak kullanım olarak ele alabiliriz [13-15]. Dolayısıyla teknik olmayan kayıplar yaklaşık olarak kaçak kullanımlara eşittir diyebiliriz ve teknik olamayan kayıpları Denklem (3)'te, teknik kayıpları ise Denklem (4)'te gösterildiği gibi ifade edebiliriz.

$$W_4\ (MWh) = E_{Kaçak\ Kullanım}\ (MWh) \quad (3)$$

$$W_5\ (MWh) = E_1\ (MWh) + E_2\ (MWh) \quad (4)$$

E_1: İletkenler üzerinde meydana gelen kayıplar
E_2: Trafo üzerinde meydana gelen kayıplar

İletken kesiti ve üzerinden geçen akım değeri bilinen iletken üzerinde kaybolan zaiyat gücü Denklem (5) yardımıyla hesaplanabilir.

$$P_Z = 3 \cdot I^2 \cdot R \quad (5)$$

P_Z: Zaiyat Güç
I: Hat akımı
R: Hattın direnci
İletkene ait direnç değeri ise Denklem (6)'da verilmiştir [9,17].

$$R = \frac{L}{x \cdot q} \quad (6)$$

L: Hattın Uzunluğu (m)
x: Hattın Özgül İletkenliği ($m/\Omega mm^2$).

Konut abonelerinin kullanmış oldukları sayaç tiplerine göre de güç kayıpları meydana gelmektedir. 2001 yılında yayınlanan (15.02.2001 tarih ve 24319 sayılı resmî gazetede) düzenlenmeye göre konutlardaki abonelerin elektronik sayaç kullanması gerekmektedir. Elektronik sayaçlarda meydana gelen güç kayıpları ortalama olarak 1-2 W arasında değişmektedir [11]. Sayaçlar üzerinde kaybolan enerji miktarı da faturalandırılamayan enerji miktarı içerisinde teknik kayıplara dahil edilebilir.

Trafo kayıpları için ise transformatör gücü 1000 kVA'nın altında ise, bakır kayıpları yaklaşık olarak transformatör gücünün %3'ü veya %4'ü oranında değişiklik göstermektedir. Sargılar üzerinde oluşan bu kayıplar kısa devre

deneyi ile hesaplanmaktadır. Primer ve sekonder sargıları üzerinde kaybolan toplam bakır kayıpları Denklem (7)'de belirtilmiştir [12,13].

Toplam bakır kayıpları;

$$P_{CU} = I_1^2 \cdot R_1 + I_2^2 \cdot R_2 \quad (7)$$

Trafolar üzerinde kaybolan demir kayıpları ise boşta çalışma deneyleriyle hesaplanmaktadır. Boşta çalışma deneylerinde bakır kayıpları ihmal edilir ve demir kayıpları sabit olarak nitelendirilir. Demir kayıpları Histeresiz ve Fuko kayıpları olarak ikiye ayrılmaktadır. Histeresiz ve Fuko kayıpları Denklem (8) ve Denklem (9)'da belirtildiği gibidir.

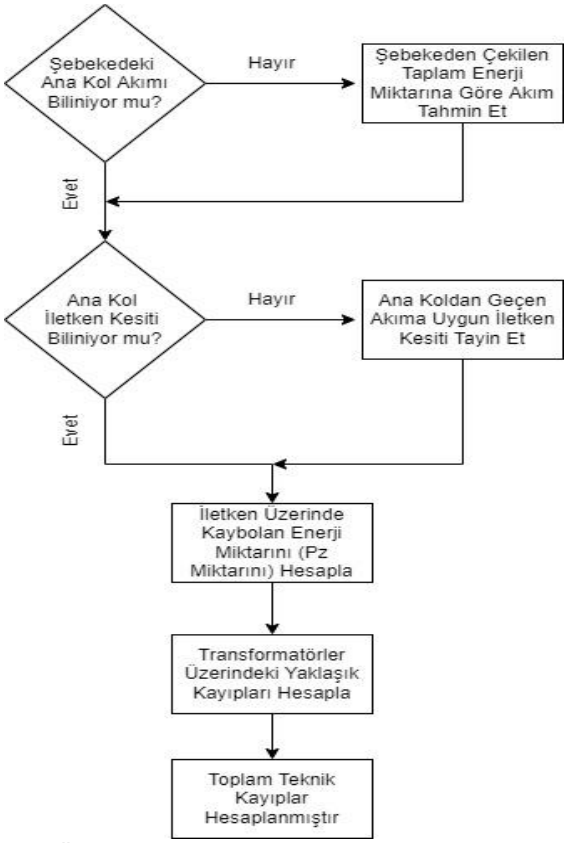
$$P_{his} = K_{his} \cdot f \cdot (B_{max})^{1,6} \ (W) \quad (8)$$

$$P_{fu} = K_f \cdot f \cdot (B_{max})^2 \ (W) \quad (9)$$

Transformatörün boştaki kayıpları Denklem (10)'da gösterildiği gibidir.

$$P_b = P_{his} \cdot P_{fu} \ (W) \quad (8)$$

Büyük ve karmaşık bölgeler için tek tek teknik kayıpların hesaplanması zor ve vakit alacak bir iştir [14,15]. Karmaşık bölgelerde hesap yapabilmek için önerilen algoritmalar takip edilebilir. Şekil-1'de teknik kayıpların hesabı için önerilen bir algoritma gösterilmektedir.



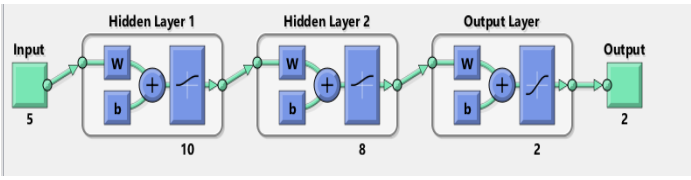
Şekil 1. Şebeke Üzerindeki Teknik Kayıpların Tespit Edilme Algoritması

Şekil 1'de önerilen algoritma (algoritma 1) yardımıyla bir şebekede meydana gelen toplam kayıplar, yaklaşık olarak hesaplanabilir. Karmaşık bölgelerde ise genel hesaplama yapılarak yaklaşık teknik kayıplar tahmin edilebilir. Bir şehirde meydana gelen teknik kayıpların tek tek hesaplanması yerine, şehir tek bir trafo bölgesine indirgenerek algoritma 1 uygulanabilir. Bir şehri veya karmaşık bir bölgeyi tek bir trafo bölgesine indirgemek için şekil 2'de verilen algoritma 2 uygulanabilir.

Şekil 2. Bir bölgenin tek trafo bölgesine indirgemesi ve kayıpların tahmin edilmesi

Önerilen algoritma-2 yardımıyla karmaşık bölgeler tek tip trafo bölgesine indirgenebilir ve bölgenin yaklaşık teknik kayıpları hesaplanabilir. Güç sistemlerinde teknik kayıplar, hat uzunlukları, trafo yapısı ve bilgileri, gerilim ve güç değerleri bilinen şebekeler için tahmin edilebilir [16,17]. Bir dağıtım şebekesinde harcanan toplam enerji miktarı, tüm trafo bölgelerine eşit miktarda paylaştırılarak, eş değer trafo bölgeleri elde edilir. Paylaştırılan güç miktarlarına uygun bir iletken tayin edilerek yeni bir trafo bölgesi oluşturulur. Herhangi karmaşık bir şebekeyi veya bir şehri tek trafo bölgesine indirgeyerek, teknik kayıpların analizleri bu sayede daha kolay bir şekilde gerçekleştirilir. Trafo kayıplarını hesaplarken, şebekenin veya bir şehrin genel durumunu veya bölgede kullanılan trafo tiplerinin yaygınlık durumunun göz önüne alınması, yapılacak tahminin doğruluk oranını artıracaktır. Trafo bölgesinin çok çeşitli olduğu bölgelerde yaygın kullanılan her çeşit trafo için, yeni trafo bölgeleri oluşturulabilir. Ancak harcanan enerji miktarları, hat uzunlukları, tayin edilen iletken özellikleri eşit (her bölge için eşdeğer) olacağından dolayı değişiklik gösterecek tek durum trafolar üzerinde meydana gelen kayıplar olacaktır (E1 değeri sabit, E2 değeri değişken olacaktır).

Algoritmalar ile elde edilen sonuçlar yapay sinir ağları (YSA) ile de desteklenmiştir [18-25]. Kullanılan YSA modelinde 5 adet giriş ve 2 adet çıkış bulunmaktadır. Girişler sırasıyla; şehrin toplam trafo gücü, toplam trafo sayısı, toplam hat uzunluğu, kayıp-kaçak oranı, faturalandırılan ve faturalandırılamayan enerjinin (tüm enerjinin) toplam miktarı şeklindedir. Çıkışlar ise toplam teknik kayıplar ve tahmin edilen kaçak kullanım miktarıdır. Oluşturulan YSA ağ yapısı şekil 3'te gösterildiği gibidir.



Şekil 3. Kullanılan YSA ağ yapısı

İleri beslemeli geri yayılımlı (Feed-Foward Backprop) ağ tipi kullanılmıştır. En yüksek doğruluk oranı, nöron sayıları sırasıyla 10 ve 8 olan iki gizli katmanın kullanılmasıyla elde edilmiştir. Nöron sayılarının belirlenmesinde, kabul görmüş herhangi bir yöntem olmadığı için nöron sayıları deneme yanılma yoluyla belirlenmiştir. Aktivasyon fonksiyonu olarak en yaygın kullanılan aktivasyon fonksiyonu sigmoid fonksiyon seçilmiştir. Birden fazla eğitim fonksiyonları kullanılmış ve eğitim fonksiyonlarının performansları birbirleri ile karşılaştırılmıştır.

2.2. Kullanılan eğitim fonksiyonları

Önerilen algoritmalar birden fazla eğitim fonksiyonu ile eğitilmiş; eğitim ve test sonuçları birbirleri ile kıyaslanmıştır.

Trainlm eğitim algoritması; "Luvenberg-Marquardt" geri yayılımı olarak bilinen ve ANN uygulamalarında kullanımı yaygın olan bir eğitim fonksiyonudur. Fazla belleğe ihtiyaç duyması dezavantaj olmasına rağmen ANN uygulamalarında sık tercih edilmektedir. Luvenberg-Marquardt optimizasyonuna göre bias ve ağırlıkları güncelleme prensibi ile çalışmaktadır [26].

Traincgb eğitim algoritması; Powerll/Beale yeniden başlatmalı geri yayılım optimizasyonuna göre bias ve ağırlıkları güncelleme prensibi ile çalışmaktadır [26].

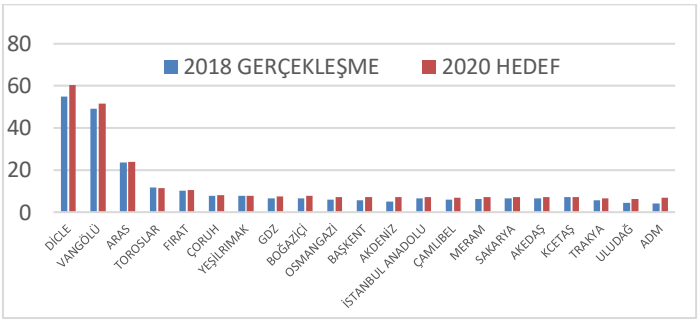
Traingdx eğitim algoritması; değişen öğrenme ve geri yayılıma sahip eğitim fonksiyonudur. En önemli avantajlarından bir tanesi değişken öğrenme hızının yüksek olmasıdır [26].

Bu çalışmada yapay sinir ağı; birden fazla eğitim fonksiyonları yardımıyla ayrı ayrı oluşturulmuştur. Bu sayede eğitim fonksiyonlarının tahminleme üzerindeki performansları da karşılaştırılmıştır.

3. Bulgular

Ülkemizde ve dünyada kayıp ve kaçak kullanımlar beraber değerlendirilmektedir. Konuyla alakalı çalışmalar olsa bile net bir sonuç yoktur [7]. Ülkemizde kaçak kullanım miktarları; kayıp ve kaçak kullanım miktarlarının toplam miktarına bakılarak yorumlanmaktadır. Ancak bu kesin bir sonuç vermemektedir [27-28].

Enerji Piyasası Düzenleme Kurumu (EPDK) verilerine göre; dağıtım firmalarının kayıp kaçak oranları şekil 4'te gösterilmiştir [29-30].



Şekil 4. Dağıtım firmalarına ait 2018 yılı kayıp kaçak oranları

Çalışmada 81 şehrin kaçak kullanım miktarları farklı güç katsayılarıyla denenmiştir. YSA için şehirlerin mevcut harcadıkları güç miktarları örnek sayısının fazla olması için 0.1, 0.25, 0.5, 0.75, 1.1 gibi katsayılarla çarpılmış ve örnek sayısı artırılmıştır. Her şehir için 6 adet örnek olmak üzere toplamda 486 adet örnek elde edilmiştir. 486 örneğin yaklaşık %60'ı eğitimde %40'ı test için kullanılmıştır. Kaçak kullanım tahmini yapılırken, MATLAB üzerinde tahminleme metodu ve K-En Yakın Komşu (kNN) sınıflandırma metotları denenmiştir. Eğitim fonksiyonlarına ait başarı oranları Tablo 1'de gösterildiği gibidir.

Tablo 1. Eğitim fonksiyonlarına ait başarı oranları

Eğitim Fonksiyonu	Training	Validation	All (R)
TrainLM	0.998	0.998	0.995
TrainCGB	0.997	0.922	0.989
TrainGDX	0.858	0.794	0.854
TrainRP	0.988	0.961	0.988

Test veriler sonucunda kaçak kullanım miktarı tahmininde oluşan MSE (Ortalama Kare Hata) oranları ise tablo 2'de gösterilmiştir.

Tablo 2. Eğitim fonksiyonlarına ait hata oranları

Eğitim Fonksiyonu	MSE	RMSE	MPE
TrainLM	5.364E+17	7.32E+08	18.88588
TrainCGB	0.1784273	0.422407	67.16684
TrainGDX	0.0748873	0.273655	49.26689
TrainRP	0.0471565	0.217155	-10.16977

Eğitim sonuçlarında trainlm eğitim fonksiyonu daha iyi bir performans göstermesine rağmen test sonuçlarında iyi bir performans göstermediği gözükmektedir. En iyi performans MSE (ortalama kare hata) dikkate alındığında "traingdx" eğitim fonksiyonunda elde edilmiştir.

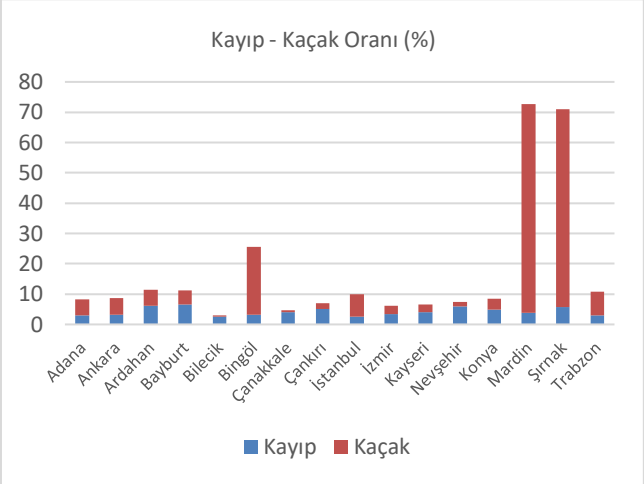
Algoritmalardan elde edilen sonuçlara göre, bazı şehirlere ait kaçak kullanım miktarları tablo 3'te verilmiştir.

Tablo 3. Eğitim fonksiyonlarına ait başarı oranları

Şehirler	W_1 (MWh)	W_2 (MWh)	E_2 (MWh)	Şehirler	W_1 (MWh)	W_2 (MWh)	E_2 (MWh)
Adana	36997.36	13562.34	23435.02	Eskişehir	8859.39	5164.11	3695.27
Adıyaman	4944.96	2793.26	22574.47	Erzurum	9365.82	3213.82	6151.99
Afyonkarahisar	7664.47	5614.77	2049.69	Gaziantep	18313.86	7350.76	10963.09
Ağrı	25008.06	1848.51	23159.55	Hakkâri	30175.74	1292.38	28883.35
Aksaray	4226.54	2860.09	1366.44	İstanbul	374710.33	94795.45	279914.87
Ankara	90127.81	33479.56	56648.24	İzmir	66098.54	36495.01	29603.53
Antalya	80276.03	23331.52	56944.50	Karaman	2516.59	1876.37	640.21
Ardahan	659.42	360.18	299.23	Kayseri	12483.12	7416.23	506.88
Artvin	4500.92	1175.25	3325.67	Kocaeli	40434.20	18018.56	22415.63
Aydın	19104.89	7751.33	11353.55	Konya	28756.79	16824.77	11932.01
Balıkesir	26674.11	8859.41	17814.70	Malatya	10128.78	4266.02	5862.75
Bartın	3934.02	1449.02	2484.99	Mardin	180895.17	9302.94	171592.22
Batman	29214.34	2268.24	26946.09	Sakarya	14901.31	6022.72	8878.59
Bayburt	596.96	346.55	250.39	Samsun	16260.82	6778.85	9481.97
Bilecik	3314.60	2822.42	492.17	Şanlıurfa	220564.36	17518.78	203042.58
Bingöl	5784.90	735.95	5048.95	Şırnak	39901.35	3227.57	36673.77
Bitlis	18984.16	1414.11	17570.05	Trabzon	13803.36	3917.11	9886.24
Bolu	3993.30	2479.52	1513.77	Van	91799.32	4006.70	87792.61
Burdur	3890.94	2176.69	1654.25	Yozgat	9312.12	2715.81	6596.30

Bu çalışmada önerilen algoritmalar ve yapay sinir ağları yardımıyla şebekeler tek bir trafo bölgesine indirgenmiştir. Öncelikle faturalandırılabilen enerji miktarları bilinen şehirlerin kayıp kaçak oranları da dikkate alınarak faturalandırılamayan (toplam kayıp ve kaçak) enerji miktarları tahmin edilmiştir. Tek tip trafo bölgelerine eşit miktarda paylaştırılan enerji miktarına göre; trafo bölgesine hayali bir iletken tayin edilip, iletken üzerinde meydana gelen teknik kayıplar (doğal kayıplar) hesaplanmıştır. Çalışmada her bölge (her şehir) için eşdeğer (tek tip) trafo kullanılmıştır. Trafo kayıpları yaklaşık olarak trafodan çekilen enerjinin %3'ü ile %4'ü civarında kabul edilmiştir.

Bazı şehirlere ait kayıp ve kaçak kullanım miktarının grafiksel gösterimi şekil 5'de gösterildiği gibidir.



Şekil 5. Bazı şehirlere ait kayıp-kaçak oranlarının grafiksel gösterimi

3. Sonuçlar

Önerilen algoritmalar ile istenilen herhangi basit bir bölgenin veya büyük elektrik güç sistemlerinin yaklaşık olarak kayıp ve kaçak kullanım miktarlarının birbirinden ayrılması hedeflenmiştir. Kayıp ve kaçak faktörleri birbirinden farklı konulardır ve ayrı ayrı incelemek enerjiye bakış açımızın değişmesi bakımından önemli bir konudur. Kaçak kullanımların ayrı değerlendirilmesi yapılacak olan çalışmaların daha verimli hale gelmesine yardımcı olacaktır. Herhangi bir bölgeye yatırım yapılacağı zaman, kayıpları azaltmak için mi yoksa kaçak kullanımların engellenmesi için mi yapılması gerektiği önemli bir konudur. İki terimin birbirinden ayrı değerlendirilmesi, yatırımların önünü açacaktır. Bu çalışmada iletken kesitinin bilinmediği bölgelerde, iletkenler çekilen yük miktarına göre ortalama olarak seçilmiştir. Trafolar bölgelere ve sık kullanılan trafo güçlerine göre seçilmiştir, trafo kayıpları yaklaşık olarak hesaplanmıştır. Bölgelerde kullanılan iletkenlerin ve trafoların durumuna göre tahminler ile gerçek değerlerin farklı çıkması muhtemel durumdur. Kaçak kullanım hesabı için önerilen algoritma, bilgileri bilinen tüm şebekeler için kullanılabilir. Kayıp ve kaçak kullanım miktarına göre, şebekeler yenilenerek kayıplar azaltılabilir veya kaçak kullanımlarla mücadeleye destek verilerek kaçak kullanım miktarı azaltılabilir. Engellenen her kayıp enerjinin, farklı alanlarda verimli bir şekilde kullanılması sağlanabilir.

Kayıp-Kaçak oranları ülkemizde doğu bölgelerinde yüksek, batı bölgelerde düşük seyrederken; kaçak kullanım miktarları yer yer değişiklik göstermektedir. Örneğin İstanbul bölgesi için kaçak kullanım miktarı batı bölgelerine göre yüksek olduğu tahmin edilmiştir. Bilecik için ise neredeyse kaçak kullanım yok denilecek kadar az olduğu gözükmektedir. Bu çalışmanın öncelikli amacı kaçak kullanımların tahmin edilmesidir. Ve bununla birlikte şehirlere yapılacak olan yatırımların nasıl yapılması gerektiği, kaçak kullanımlarla mücadelede nerelere yoğunlaşmak gerektiği konusunda yapılacak olan çalışmalara kolaylık sağlaması hedeflenmiştir. Kayıp ve kaçak kullanımlar elektriksel açıdan birbirinden farklı terimlerdir ve ayrı değerlendirmeleri gerekmektedir.

Yazar Katkı Oranları

Gerçekleştirilen çalışmada Enes YILDIZ, modelin oluşturulması, literatür taraması, verilerin toparlanması ve işlenmesinde, Nurettin ÇETİNKAYA fikrin oluşması, elde edilen sonuçların değerlendirilmesi ve makalenin imla denetiminde katkıda bulunmuştur.

Çıkar Çatışması

Yazarlar, bilinen herhangi bir çıkar çatışması veya herhangi bir kurum/kuruluş ya da kişi ile ortak çıkar bulunmadığını onaylamaktadırlar. Hazırlanan makalede etik kurul izni alınmasına gerek yoktur. Hazırlanan makalede herhangi bir kişi/kurum ile çıkar çatışması bulunmamaktadır.

Kaynakça

[1] Örsek B. (2016). Türkiye'de Kayıp Kaçak Oranı Düşme Eğiliminde. www.dogrulukpayi.com. (Erişim Tarihi: 25.10.2020).

[2] Sargın İ. (2006). Tavşanlı Enerji Dağıtım Sistemlerinde Meydana Gelen Kayıpların Oranlarının Belirlenmesi ve Azaltılması İçin Alınabilecek Önlemler ve Elde Edilen Enerji Kazanımları. *Türkiye 10. Enerji Kongresi, İstanbul, Türkiye*.

[3] Düzgün B. (2018). Türkiye Elektrik Iletim ve Dağıtım Şebekesinin Enerji Verimliliğinin Değerlendirilmesi ve 2023 Projeksiyonları. *Politeknik Dergisi*, 21(3): 621-632.

[4] World Bank Group (2016). Türkiye Cumhuriyeti: Elektrik Dağıtım Şirketlerinin Hizmet Kalitesinin İyileştirilmesine Yönelik Adımlar. *World Bank Group Türkiye, ACS20668.*

[5] Yaşar C., Aslan Y., Biçer T. (2010). Bir Dağıtım Transformatörü Bölgesindeki Kayıpların İncelenmesi. *Dumlupınar Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22, 9-22.

[6] Alperöz N. (1984). Elektrik Enerjisi Dağıtımı, Elektrik Enerjisi Dağıtımı, Nesil Matbaa, birinci baskı, Türkiye.

[7] Yıldız E., Çetinkaya N. (2018). The Proposed Forecasting Algorithm in Power Systems for Separating of Losses and Illegal Consumptions*. Third International Symposium on Industrial Design & Engineering 2018 (ISIDE)*, Antalya, Türkiye.

[8] Bhatt MS. (2003). Energy Efficiency Improvement of Electrical Transmission Distribution Networks". *Journal of Scientific & Industrial Research*, 62, 473-490.

[9] Balcı H., Esener İ., Kurban M. (2012). Regresyon Analizi Kullanılarak Kısa Dönem Yük Tahmini. *ELECO 2012 Elektrik- Elektronik ve Bilgisayar Mühendisliği Sempozyumu, Bursa, Türkiye.*

[10] Alcan Y., Öztürk A., Dirik H, Demir M. (2017). Güç Şebekelerinde Minimum Kayıpları Sağlayan STATCOM Konumunun ve Değerinin Belirlenmesinde Farklı Sezgisel Algoritmaların Karşılaştırılması. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 23(5), 550-558.

[11] Özel K. (2006). *Losses in Electric Distribution Systemes* (Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi, Ankara, Türkiye)

[12] Smith T.B. (2004). Electricity Theft: a Comparative Analysis. *Deparment of Social and Behavioral Science.* 32, 2067-2076.

[13] TEDAŞ (2018). 2017 Faaliyet Raporu. https://www.tedas.gov.tr. (Erişim Tarihi: 01.11.2020)

[14] Amjady N., Keynia F. (2008). Mid-term Load Forecasting of Power Systems by a Newprediction Method. *Energy Conversion and Management,* 49, 2678-2687

[15] Gürsoy E., Kaypmaz A. (2000). *Yük Tahmini yöntemleri ve Çukurova Elektrik A.Ş, , Kepez Elektrik T.A.Ş. bölgelerine uygulaması* (Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, Türkiye).

[16] İşcan S., Kaplan O., Lokman G. (2021). Güç sisteminde meta-sezgisel algoritmalarla güç kaybı ve gerilim kararlılığı optimizasyonu. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 27(2), 199-209.

[17] Uche O., Madueme T.C. (2015). A Power Flow Analysis of the Nigerian 330 kV Electric Power System. *The International Organization of Scientific Research Journal of Electrical and Electronics Engineering*, 10(1), 46-57.

[18] Viegas J.L., Esteves P.R., Melicio R, Mendes V.M.F., Vieira S.M. (2017). Solutions for detection of non-technical losses in the electricity grid: A review. *Renewable and Sustainable Energy Reviews*, 80(1), 1256-1268.

[19] Han W., Xiao Y. (2017). A novel detector to detect colluded non-technical loss frauds in smart grid. *Computer Networks*, 117, 19-31.

[20] Henriques H.O., Correa R.L.S., Fortes M.Z., Ferreira V.H. (2020). Monitoring technical losses to improve non-technical losses estimation and detection in LV distribution systems. *Measurement*, 161(1), 107840.

[21] Filho M.B.D.C., Silva A.M.L., Falcao D.M. (1990). Bibliography on Power System State Estimation (1968-1989). *IEEE Transsactions on Power Systems*, 5(3), 950-961.

[22] Nizar A.H., Dong Z.Y., Wang Y. (2008). Power utility nontechnical loss analysis with extreme learning machine method. *IEEE Transactions on Power Systems*, 23, 946-955.

[23] İnan A., Köroğlu S., İzgi E. (2005). Dengeli Elektrik Güç Sistemi Verilerini Kullanarak Dengesiz Sistem Kayiplarinin Yapay Sinir Ağlari Ile Belirlenmesi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 11(1), 47-52.

[24] Ghasemi A.A., Gitizadeh M. (2018). Detection of illegal consumers using pattern classification approach combined with Levenberg-Marquardt method in smart grid. *International Journal of Electrical Power & Energy Systems*, 99, 363-375.

[25] Monteiro R.V.A, Guimaraes G.C., Silva F.B., Teixeira R.F.S., Carvalho B.C., Finazzi A.P., Vasconcellos A.B. A medium-term analysis of the reduction in technical losses on distribution systems with variable demand using artificial neural networks: An Electrical Energy Storage approach. *Energy*, 164(1), 1216-1228.

[26] Çakır S.F., (2018). Yapay Sinir Ağları

[27] Neto E.A.C.A., Coelho J. (2013). Probabilistic methodology for Technical and Non-Technical Losses estimation in distribution system. *Electric Power Systems Research*, 97(1), 93-99.

[28] Navani J.P., Sharma N.K., Sapra S. (2017). Analysis of Technical and Non Technical Losses in Power System and its Economic Consequences in Power Sector. *International Journal of Advanced Electrical and Electronics Engineering*, 1(3), 396-405.

[29] TEİAŞ (2019). Faaliyet Raporu https://www.teias.gov.tr. (Erişim Tarihi: 05.04.2020)

[30] EPDK (2019). Faaliyet Raporu. https://www.epdk.org.tr. (Erişim Tarihi: 01.11.2020)

Journal of Investigations on Engineering & Technology
Volume 5, Issue 1, 11-20, 2022

ISSN: 2687-3052

Research Article

Analysis of Energy Raw Material Coal, Industrialization and Industrial Revolution Phenomena with N-gram

[1]Alaaddin VURAL, [2]M.Nuri URAL,[3]Ali ÇİFTÇİ

(Alınış / Received: 05.04.2022, Kabul / Accepted: 09.05.2022, Online Yayınlanma / Published Online: 30.06.2022)

Keywords
Please add at least three keywords
such as: *Fossil fuels*
Geopolitics
Coal mining
Mercantilism
Industrialization
Industrial Revolution.

Abstract: In this study, coal, which is one of the energy raw materials, and coal-related keywords and the relationship between industrialization and political/social events from past to present were investigated by n-gram analysis. Within the scope of the study, using the English works registered in the Google books directory, coal as energy raw material and related keywords "coal", "coal mine", "coal mining", "fossil fuels", "industrial revolution", "industrialization" "mercantilism", "geopolitics", and "geology" were used, and these words were subjected to n-gram analysis as subgroups that would form a unity within themselves.

With this study, coal-energy raw material, which is the most important engine power of industrialization in the industrial revolution and the following period, and the relationship of its related concepts with "Geo(logy)politics and its parallelism with political, military, social and economic events from past to present have been examined with n-gram analysis.

As a result, there has been an increase in the frequency of use of keywords after the 1820s in relation to the beginning of the Industrial Revolution and the increase in its effectiveness in world history, and it has been determined that the related words have significant co-changes with each other. Therefore, with the n-gram analysis, the Industrial Revolution/industrialization relationship between coal and related keywords was easily observed. When the digital environment, which is a large data source, is analyzed with the n-gram method using the appropriate keywords to be selected, it is understood that it will be possible to reach many sources and that such a large corpus will allow multi-purpose and multidimensional evaluations. The fact that this study reveals the origins of coal-related concepts and industrialization in the literature with the n-gram analysis method, which is one of the energy sources, also presented an extraordinary perspective with an analysis technique out of the ordinary.

1. Introduction

In today's world, where information technologies are an important part of our lives, and their use is spreading to all areas of our lives (Hussein, 2021; Raudeliuniene et al., 2020), digital media data analysis is gaining importance day by day and finds its place in many areas. While globalization and technological advances color our world, which has turned into an information universe, on the other hand, it also contains difficulties in every field at the point of management of such a large amount of information. Technological innovations also increase

[1] Gümüşhane University, Department of Geological Engineering, Gümüşhane/Türkiye. ORCID: https://orcid.org/0000-0002-0446-828X)
[2] Gümüşhane University, Software Engineering Department, Gümüşhane/Türkiye. ORCID: https://orcid.org/0000-0001-7011-401X
[3] Amasya University, Merzifon İİBF-Department of Political Science and Public Administration, Amasya, Türkiye, ORCİD: ID/0000-0002-1273-4867

Corresponding Author and e–mail address: Alaaddin Vural, alaaddinvural@hotmail.com

very creative demands (Raudeliuniene et al., 2020). Today, information technologies and technology-based social networks are used by more than 50% of the world's population. This has brought some innovations that seem impossible to return in our age, in many areas of our lives: communication tools, information exchange, products and services, etc.
Especially with the widespread use of the internet, accessing information has become faster and easier than ever before. The irrepressible increase in information has presented a new problem for researchers: How to use such a large source of information efficiently and easily. In this context, "big data" and "data mining", which are now widely used in the literature, have become popular concepts in almost every scientific field (Wang et al., 2020). The concept was first expressed in the journal Nature and defined as data on a very large scale that cannot be presented, processed and analyzed with existing technology, methods and theories (Goldston, 2008). Today, new research/evaluation methods have been started to be developed for such a large data collection that has been transferred to digital media and can be accessed from almost the most remote point of the world via the internet (Lin et al., 2019). What is important today is how effectively researchers can use these methods in the information universe rather than their access to information. Researchers gain a linear advantage over other researchers with the methods they use/develop.

In this study, one of the important energy sources and the most important engine power of industrialization, coal, and the relationship between coal-related concepts and economic and political events, using the data in the digital environment, were examined with n-gram analysis, which is a kind of natural language processing method. The study is also related to the fields of geopolitics and geology. Although geology is briefly defined as earth sciences, it covers many fields with its sub-branches and sheds light on underground resources and many issues that are especially important for countries (Bobbette & Donovan, 2019). Geopolitics, on the other hand, is the science that examines the relations between the geographical features of states and their politics. The name of the concept originates from Swedish Rudolf Kjellen (1864-1922). He introduced the concept for the first time in 1916 in his book The State as an Organism. C. Haushofer, one of the geopolitical theorists, also defines geopolitics as the relationship of the state with the place in which it lives. Although the concept of geopolitics, which became widespread between the two World Wars and was widely used during the Second World War, seems to have lost its importance today with the development of transportation and communication technologies and the decrease in the dependence of countries on geographical features, it still has an important place (Dodds, 2007). The concept of geopolitics also emphasizes the importance of natural resources and therefore, geographical location in the development. Along with the concept of geopolitics, the concepts of geography-geostrategy have also found use. Coal, one of the fossil fuels, which was the main energy source for the energy demand, which was a kind of engine of the Industrial Revolution, had a dominant place in the history of World Industrialization(Andrews, 2008; Beresford, 2018). Considering all the infrastructure explanations mentioned above, the ability to examine the relationship between industrialization and energy resources with each other and with political/social/economic events from past to present is the research question of the study, and the aim of this study is "*A conceptual retrospective analysis of coal in particular, an important motivational force of the Industrial Revolution, natural energy resources/fossil fuels in general, and the industrialization/industrialization process, from the past to the present, with political/social events, by making use of the English works registered in the Google books directory in the digital environment, by the n-gram method*".
N-gram analyzes have been used in the literature mostly in the informatics sector. There are very limited studies in the literature on the use of analysis with natural resources, which are important raw materials for geology/geopolitics and political, social and technological developments of countries (Çiftçi et al., 2019; Çiftçi, Ural, et al., 2020; Çiftçi, Vural, et al., 2020a, 2020b; Ural et al., 2020c, 2020b, 2020d; Vural et al., 2021b; Vural, Çiftçi, et al., 2020; Vural, Ural, et al., 2020a, 2020b; Vural et al., 2021a; Vural & Çiftçi, 2021b, 2021a). In the studies conducted by Vural et al. (2019; 2020c), the relationship between colonialism and precious metals (gold, silver) was analyzed with the n-gram method and a significant relationship was revealed. In the study by Vural et al.(2021a; 2020), it was emphasized how the colonial activities toward gemstones such as diamonds affected the fate of weak countries. In a similar study, a retrospective analysis of Pb-Zn-Cu elements (base metals for short), which has an important place in Industrialization, was carried out by Çiftçi et al. (2019; 2020). Ural et al. (2020a, 2019) tested the hypothesis that rare earth elements (REE), which is an important raw material of high-tech products, will relate to the supply, market and geopolitics with n-gram analysis and reached satisfactory results. In the aforementioned studies, researchers have shown that significant relationships can be determined by n-gram analysis in the evaluation of geopolitical processes, colonization and related events that accelerated after the Industrial Revolution.
Although fossil fuels, especially coal, have been partially recognized and have long been known by humanity, they have gained importance in the history of humanity both politically, socially, economically and technologically with the Industrial Revolution and have always maintained this importance until today. It is known that coal was used for heating and simple metallurgy activities (smelting copper) in China about 3000 years ago. In the 4th century AD, the Romans also used coal. In the 13th century, it is known that surface coal mining was carried out, especially in France (Ünalan, 2013). Until the 18th century, the use of coal was on a small

scale. In the 18th century, the most important energy source effective in the birth of the first Industrial Revolution in Europe was coal. It has maintained its importance from the Industrial Revolution until today without losing its importance. Today, it is the most important and most abundant fossil energy source in the world, and when it is considered as a reserve, its volume is quite high compared to oil and natural gas. It is thought that coal will still be available after the depletion of these raw materials. Considering that the regions where a large part of the world's coal reserves are located and the regions where the industry was born and developed are different (United Kingdom, Europe, industrialized-America, Asia, Africa raw material source), the supply-demand imbalance has increased the imperial appetite for the supply of this mine. Therefore, this imbalance has been the driving force of important exploitation movements in history and has led to social/political consequences. In this study, coal and other related keywords, which have an important place in the political-economic life of the mentioned societies, were analyzed using the n-gram method, and it was examined whether the frequency of use of these words overlapped with the social/political/economic events covering approximately 250 years. It has been investigated whether the n-gram analysis of the political/social/economic events in the mentioned period with the selected keywords works.

2. Material-Method

In this study, the n-gram analysis method, which has been used very limitedly in the social/political field, has been used. N-grams are all combinations of adjacent words or letters of length n that can be found in the source text. (Çiftçi, Vural, et al., 2020b). If each word is considered as a class in the N-gram analysis method, language modelling can also be viewed as a classification problem (Aleahmad et al., 2007; Huang et al., 2012). This is used to create an index of how often words follow each other. The assumption that the probability of a word depends only on the previous word is also known as the Markov conjecture (Gagniuc, 2017). Markov models are a class of probabilistic models that assume that the probability of a future unit can be predicted without looking too far into the past.

In this study, n-gram analysis method was used. For N-gram analysis, the works registered in the Google books database were scanned with the interface prepared by Google. In the scanning, 3 was used as the smoothing factor and the English language was used as the corpus. Since the field of energy is a large phenomenon in itself, only "coal" from energy raw materials and related concepts and the phenomenon of industrialization were investigated in this study. By revealing the historical usage frequencies of these concepts in the literature with the n-gram analysis method, an extraordinary perspective has been tried to be presented with an analysis technique (n-gram) out of the ordinary.

As keywords in N-gram analysis industrialization, fossil fuels, the industrial revolution, coal mining, coal mine, geopolitics and mercantilism concepts are preferred. The graphs of n-gram analysis, in which all the concepts are used together, were not preferred because they present a complex appearance and in some cases, some concepts suppress other concepts, instead, it was preferred to create analysis graphs by considering the concepts that exhibit meaningful integrity with each other.

3. Findings and Discussion

Within the scope of the study, n-gram analyzes were carried out by using the keywords that characterize the energy raw materials coal and coal and other political/social events of interest (Figure 1-10). As stated in the method section, n-gram analysis was carried out separately for the related concept groups, since the graphics of the analyzed concepts were not very clear due to the dominance of some concepts compared to other concepts. In this context, first, n-gram analysis was created by using the concepts of coal and coal mine and coal mining (Figure 1). It is clearly seen in the graph that the word “coal” suppresses the other two words. When n-gram analysis is made with other keywords, it is observed that the word coal is always dominant, since the graphics in question are not meaningful, they are not included in the article, only the coal keyword and the graphics of the coal mining and coal mine keywords are given in the article. For this reason, it is preferred to interpret the graphic formed by the word coal. When the concept of coal is examined (Figure 1), it is seen very comfortably that it has had an upward trend since the 1750s. This upward trend showed itself more strikingly after the 1860s and reached its peak in the early 1900s (just before the First World War and the war years). The 1920s (1929 economic depression years) were the years of transition to a decreasing trend. On the eve of and after World War II, the downward trend was reversed again, and the aforementioned years became the years of an upward trend with the effect of the war. From the 1940s to the 1960s, a decreasing trend was observed in the frequency of use of the concept. The cold war years of 1970 and after being observed as the years in which the keyword tends to increase. It is thought that the economic struggle of the bipolar world is effective in this. The rise seen in the graph in the 1970s can also be associated with the 1973 Oil Crisis. The increase in the price of oil, which is an important energy source, and the decrease in its production and supply, brought about the discussion of alternative energy sources such as natural gas, nuclear energy, and wind energy. In this period, it can be easily thought that coal among alternative energy sources came to the fore and discussed. For example, at the

beginning of the 20th century, England made its coal-fired warships run on oil in order to maintain its dominance on the seas (Yılmaz & Kalkan, 2017). Probably the 1973 Oil Crisis has created a reversal trend here.

The 1980s and later and the 1990s were the years in which the frequency of use of the concept tended to decrease with the transformation of the cold war into a unipolar world order and with the effect of environmental concerns and awareness of carbon emissions (Figure 1).

When the graph of the n-gram analysis, in which the keywords fossil fuels and industrial revolution are evaluated together, is examined (Figure 2), it is understood that the concept of fossil fuels, which was rarely used in the past, began to be seen prominently in the literature after the 1920s. It is understood that the frequency of use of the concept increased after the 1940s, therefore, in the past, concepts such as coal and oil were used separately instead of fossil fuels (Figures 1 and 2). Similar to Figure 1, the 1980s and later and the 1990s were the years when environmental concerns and awareness increased, and a more increasing trend is observed in the concept of fossil fuels in this period (Figures 1 and 2). The concept of the industrial revolution, on the other hand, shows an increasing trend after the 1860s, as expected. After the 1920s, there was a remarkable upward trend between the years 1940-1970, following the downward trend. This time period is the period when the growth in the economies of the World's capitalist countries is the highest and the Industrialization accelerates. It was also found significant that it overlapped with the years corresponding to the revolutionary youth movements (1968 generation etc.). When the concepts of industrialization and industrial revolution are analyzed together (Figure 3), it is seen that the concept of industrial revolution dates back to the 1750s, while the concept of industrialization has a remarkable increase after the 1920s.

At the beginning of the Second World War, the concept of the industrial revolution was horizontal, while the concept of industrialization decreased slightly in the period coinciding with the war years. It is also possible to see this as a reflection of the decline in industrial production during the war years and the necessity of industrialization. Economic historians and social scientists state that the fastest growth and development in industrialized countries occurred between the 1950s and 1970s. For example, the average growth rate in industrial production between 1953 and 1975 was 6% (Kennedy, 1996). The increase in the frequency of use of the concept of industrialization observed in Figure 3 has parallels with this period. The growth between 1950-1970 is traced to the peak point in the 1970s on the graph (Figure 3). The concept of “industrialization” has a decreasing trend in terms of usage frequency after the 1970s, and it has started an upward trend since the 2000s, especially with the concept of industry 4.0 (Jones, 2021; Mavropoulos & Nilsen, 2020).

Considering the n-gram analysis of the concepts of industrialization and fossil fuels, it is seen that the 1980s had a generally parallel trend, confirming Figures 2 and 3 (Figure 4).

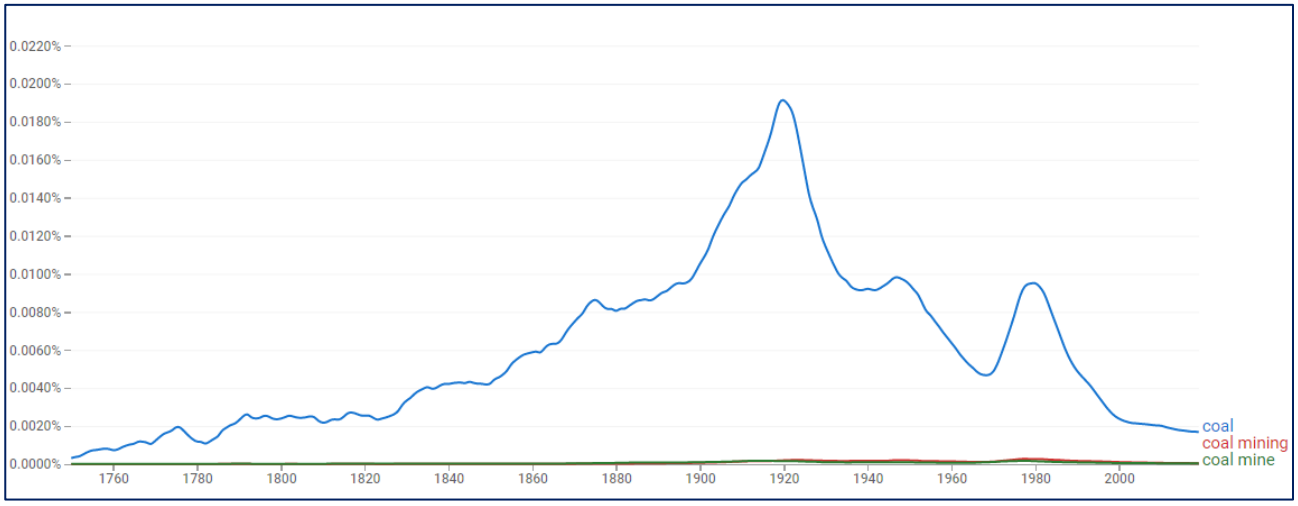


Figure 1. n-gram analysis graph of coal keyword

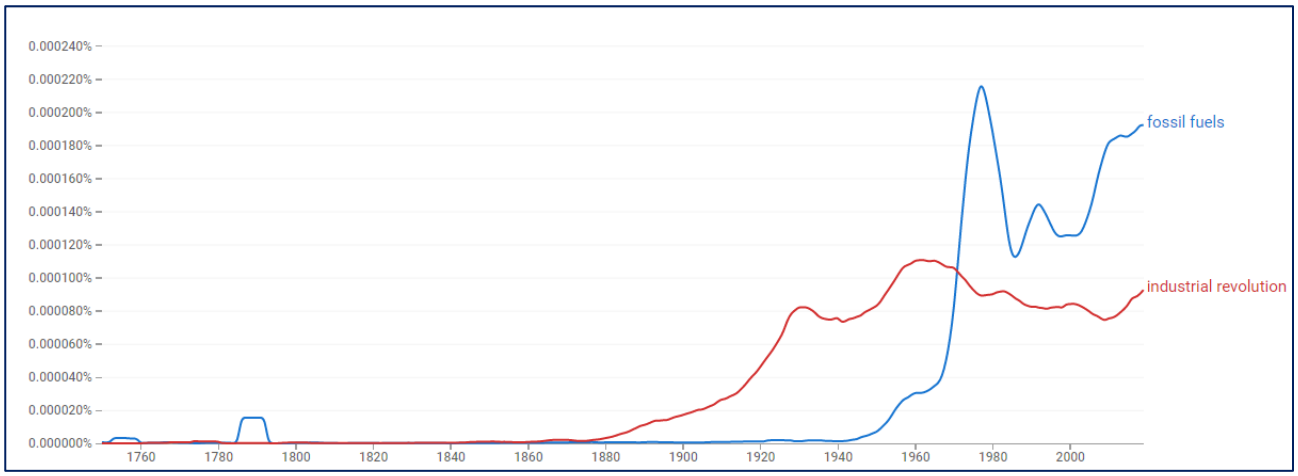


Figure 2. Analysis of Industrial Revolution and Fossil Fuels concepts together

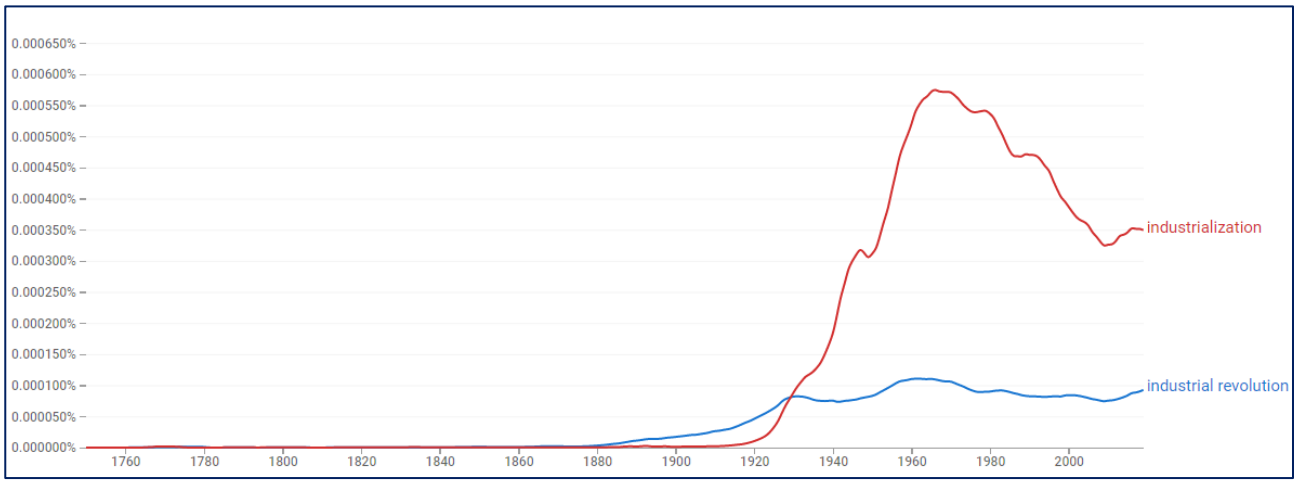


Figure 3. n-gram analysis of industrialization and industrial revolution keywords

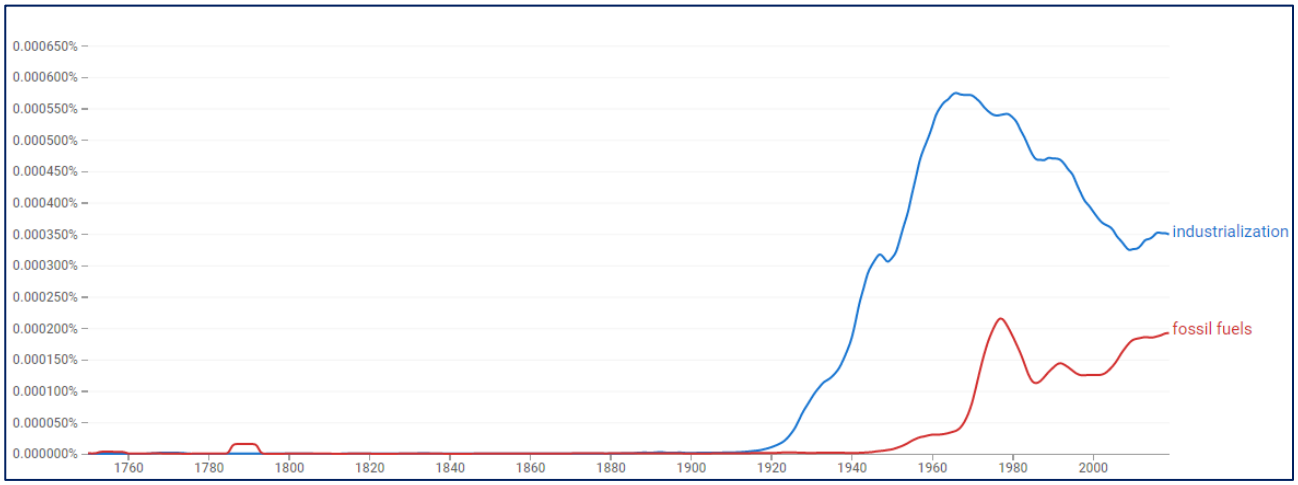


Figure 4. n-gram graph of fossil fuels and industrial revolution keywords

When the concepts of “industrial revolution”, “coal mining” and “coal mine” are subjected to n-gram analysis as a group (Figure 5), it is seen that the frequency of use of coal mine and coal mining keywords in the literature dates back to before the 1800s, although there are small fluctuations. After the 1800s, coal-fired engines came to the fore (Bruere, 1922), and coal gained importance, and in parallel, a relative increase was observed in coal and related keywords. When these two concepts are evaluated together with the concept of the industrial revolution, the increase in their use in the literature in the 1820s is remarkable. Especially after the 1870s, the increase shows itself more with small fluctuations (Simmons, 1976). It is known that since the beginning of the 1900s, mass production began in the industry, and as a result, the amount of product produced has increased many

times over the previous decades. Therefore, the use of coal in the industry and the frequency of use as a concept in the literature tend to increase in this period, which was confirmed in Figure 5. The rise seen in the graph from the 1900s to the 1940s can also be explained by the mass production in the industry. Among the reasons for the increase seen after the 1920s, the development of India, especially based on the coal industry, should be mentioned. There is a parallelism with this situation (Simmons 1976). Likewise, their monopolization by acquiring European coal mines is in parallel with this process. Considering the aforementioned data, the increase in the frequency of use of the concept seems more significant.

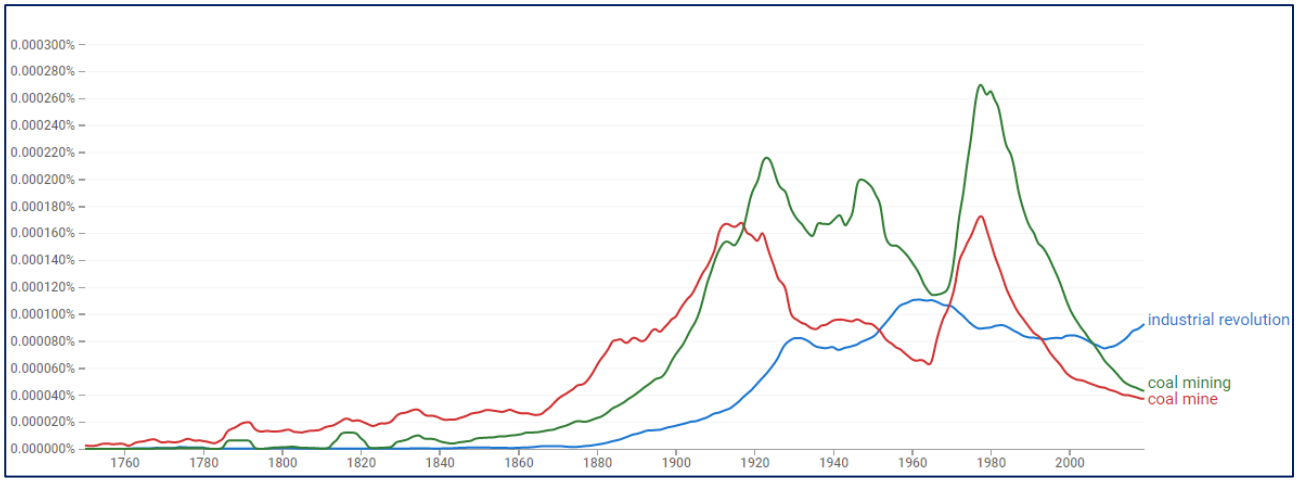


Figure 5. n-gram graph for keywords industrial revolution, coal mine and coal mining

The concept of mercantilism, which argues that the economic power of states is related to the mines they own (Magnusson, 2015; Stern & Wennerlind, 2014), is also directly related to concepts such as industrialization and industrial revolution. (Çiftçi et al., 2019; Çiftçi, Ural, et al., 2020; Vural et al., 2019). Considering that the energy, and therefore the possession of energy raw materials (especially coal) in the industrialization of countries, is related to mercantilism, the concepts of mercantilism and coal mining and coal mining were also analyzed with n-grams (Figure 6). When Figure 6 is examined, it is seen that the concepts of coal mining and coal mining show parallelism with each other, while the concept of mercantilism in general terms has been on an upward trend since the 1880s. When the concepts of geopolitics and mercantilism are evaluated together, it is seen that the concepts of mercantilism and geopolitics have entered an upward trend together since the late 1930s. The concept of geopolitics was used a lot in the literature during the Second World War, and it peaked after the war. While the two concepts showed a downward trend, although not parallel, in the 1960s, the concept of geopolitics had a remarkable upward trend after the 1990s (Figure 7). Since the end of the Cold War and the bipolar world order at the beginning of the 1990s, and the establishment of a new order or disorder, the concept of geopolitics has come to the fore in the literature, as can be seen in the graphic, a sharp rise has been observed since the 1990s. Figure 7 clearly demonstrates this. Today, the concept of geopolitics has a high frequency of use in line with its importance.

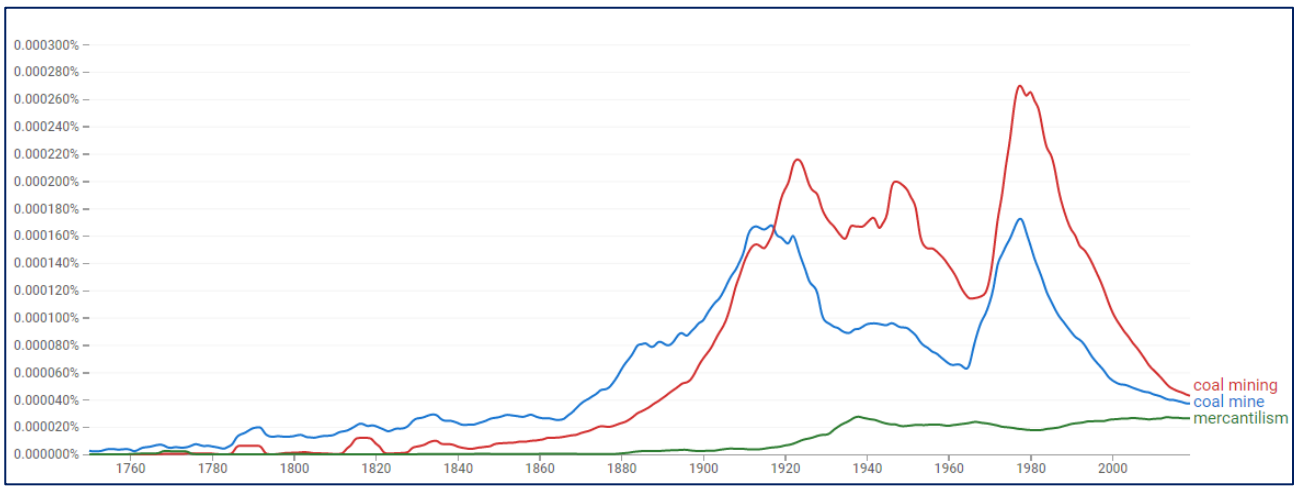


Figure 6. Graph of n-gram analysis of coal mining, coal mine, mercantilism concepts

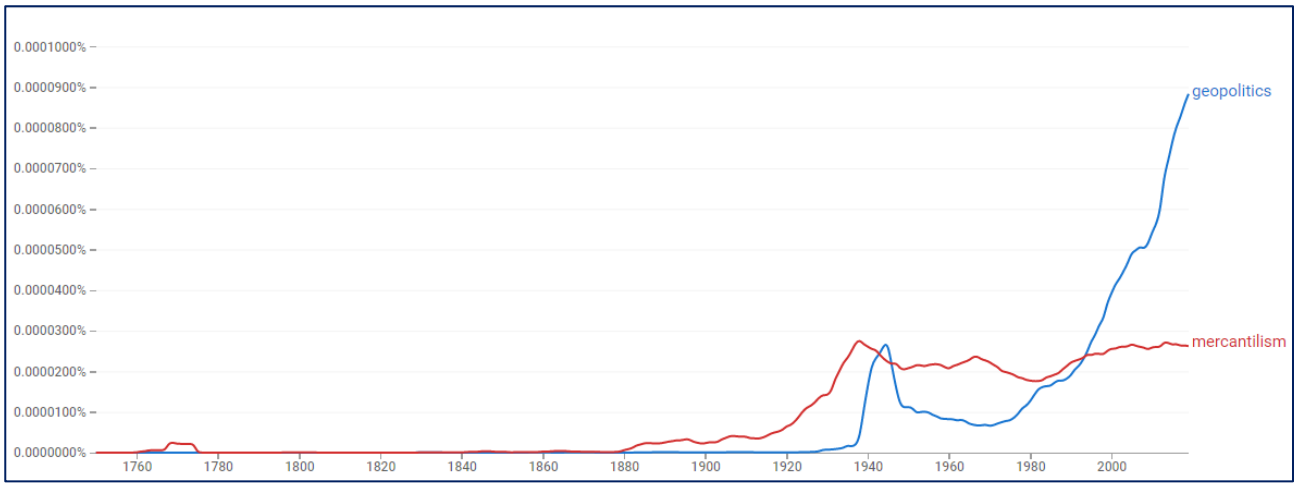


Figure 7. n-gram analysis of geopolitics and mercantilism

The science of geology (especially economic geology), which also deals with the exploration and discovery of natural resources, has gained importance, especially with industrialization and, subsequently, colonialism. Coal, which has an important place in industrialization and the industrial revolution, can be chosen as a suitable keyword group with the science of geology in this sense. Therefore, in the study, the concepts of industrialization, the industrial revolution, coal mine and coal mining were subjected to n-gram analysis in subgroups, taking into account the multiplier effects of the concepts (Figure 8-10).

Although the mining activities of human beings started with the existence of humanity on earth (Vural, 1998, 2006; Vural et al., 2009; Vural & Aydal, 2016; Vural & Ünlü, 2020, 2016), the concept of "geology", which has become synonymous with earth sciences and benefiting from natural resources, began to take place in the literature at the end of the 1700s, and the frequency of use always had an increasing trend with temporal fluctuations in the period until the end of the 1900s (Figure 8-10). The concept showed an upward trend until the end of the 1980s in parallel with the concept of industrialization after the 1920s. After the 1980s, with the concept of economic geology replacing the concept of geology, the frequency of use of the concept decreased. The concept of industrialization, on the other hand, entered an upward trend again, albeit slightly, after the 2000s, as stated earlier (Figure 8).

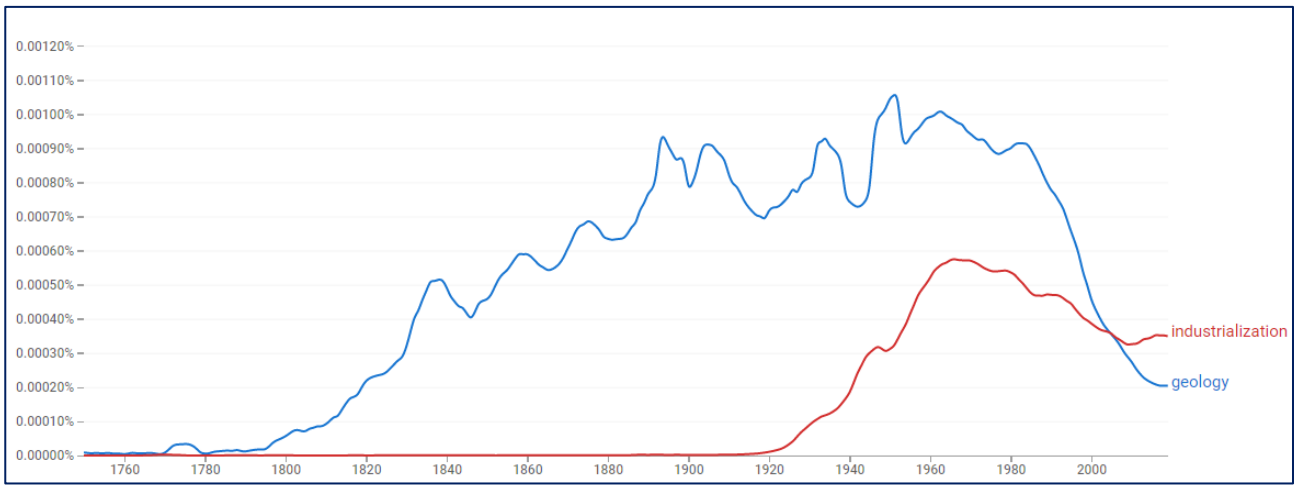


Figure 8. n-gram analysis of industrialization and geology concepts

Although the concepts of "geology" and "industrial revolution" do not show clear parallelism (Figure 9), both concepts are in an increasing trend from the early 1900s to the 1980s (2000s). Since the frequency of use multipliers are different, clear relationships are not seen, but the fluctuation trends of both concepts show relative parallelism, especially from the 1920s to the 2000s (Figure 9). The same is true for the concepts of geology, coal mining and coal mine. (Figure 10). Therefore, there are significant usage frequency relationships

between the concepts of geology, the industrial revolution, coal mine, coal mining and even industrialization.

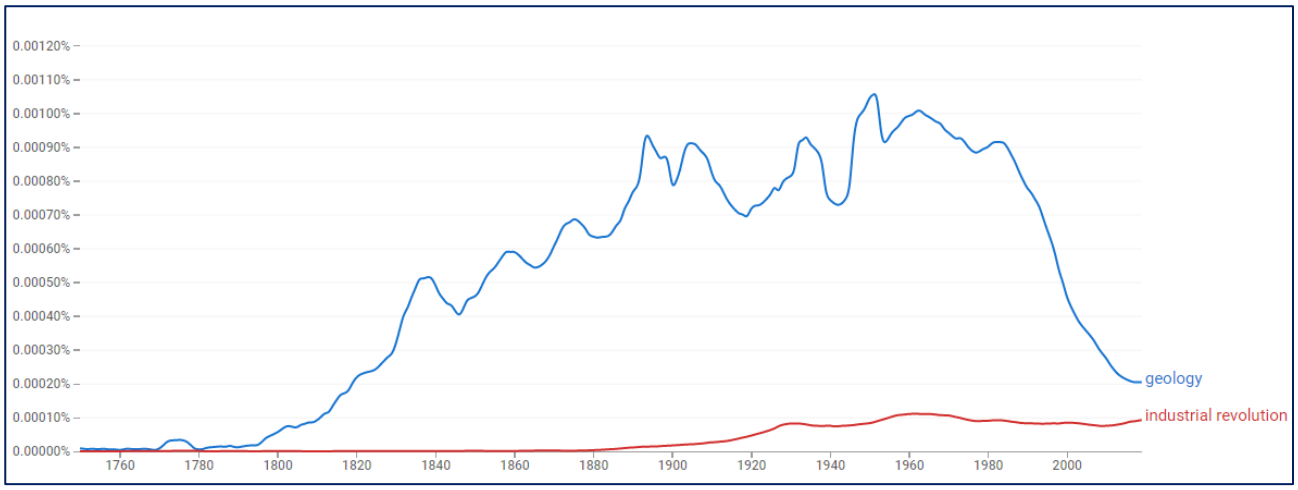


Figure 9. n-gram analysis of the concepts of geology and industrial revolution

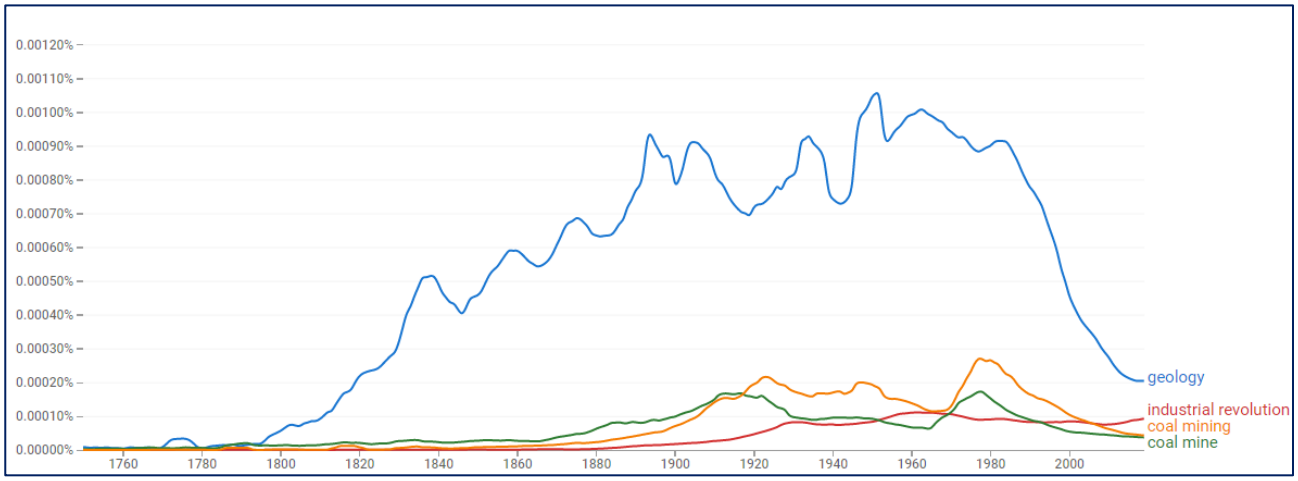


Figure 10. n-gram analysis of the concepts of geology, industrial revolution, coal mine and coal mining

3. Results

In this study, the concept of coal, which is an important energy raw material, has been analyzed with the n-gram method by associating it with industrialization and the industrial revolution and using related keywords that are expected to contribute to the analysis of the subject. For this purpose, the keyword “coal” was chosen as “coal mine” and “coal mining” keywords to be the distinction, and the related keywords were “fossil fuels”, “industrial revolution”, “industrialization”, mercantilism”, “geopolitics” and “geology” has been determined. Since it would be difficult to do an n-gram analysis of all keywords together, these words were subjected to n-gram analysis under subgroups. When the n-gram analysis graphs obtained are evaluated together, it is understood that the n-gram analysis of coal, which has an important place in the industrial revolution and early industrialization period, with related keywords, shows a usage frequency trend associated with political, social, military and economic events in history. Therefore, the possibility of deep analysis on the targeted subject with n-gram analysis with appropriate keywords has been confirmed/shown in this study.

Author Contribution

In this study, the selection of keywords, the determination of appropriate keywords and the analysis of appropriate keywords by reducing them to subgroups were carried out by Alaaddin Vural. The n-gram analyzes of the keywords were made by M. Nuri Ural. The evaluation of the data obtained and its interpretation in relation to political, social, military and economic events was mainly carried out by Ali Çiftçi and partially by Alaaddin Vural and M. Nuri Ural. The translation of the text into English was carried out by M. Nuri Ural.

Conflict of Interest

The author(s) confirm that there is no known conflict of interest or common interest with any institution/organization or person.

References

Aleahmad, A., Hakimian, P., Mahdikhani, F., & Oroumchian, F. (2007). N-gram and local context analysis for Persian text retrieval. *2007 9th International Symposium on Signal Processing and Its Applications, ISSPA 2007, Proceedings*, 1–4. https://doi.org/10.1109/ISSPA.2007.4555345

Andrews, T. G. (2008). *Kiling for Coal*. Harvard University Press.

Beresford, Q. (2018). *Adani and the War over Coal*. Newsouth.

Bobbette, A., & Donovan, A. (2019). Political Geology. In *Political Geology: Active Stratigraphies and the Making of Life*. Palgrave mcmilan publishment. https://doi.org/10.1007/978-3-319-98189-5_3

Bruere, R. (1922). *The Coming of Coal*. Association Press.

Çiftçi, A., Ural, M. N., & Vural, A. (2020). Baz metallerİn dünya sİyasİ tarİhİndekİ önemlİ olaylarla bağlantısının retrospektİf lİteratür taramasi yöntemİ İle araştırılması. *International Social Sciences Studies Journal*, *6*(60), 1453–1461.

Çiftçi, A., Ural, M. N., & Vural, A. (2019). Investigation of the Relationship between Base Metal (Pb-Zn-Cu) and Social/Political/Economical Events by N-gram Analysis. *4. Uluslararası GAP Sosyal Bilimler Kongresi*, 116–124.

Çiftçi, A., Vural, A., & Ural, M. N. (2020a). Analysis of Environmental and Healt Related Concepts with N-Gram Method. *5.Uluslararası Sağlık Bilimleri ve Yönetimi Kongresi*.

Çiftçi, A., Vural, A., & Ural, M. N. (2020b). N-Gram Analysis of Industrial/Industrial Revolution Relation with Coal and Related Keywords. *Euroasia Summit Congress on Scientific Rearches and Recent Trend-6*, *May*, 587–591.

Dodds, K. (2007). Geopolitics: A Very Short Introduction. In *Journal of Contemporary European Studies*. Oxford University Press. https://doi.org/10.1080/14782804.2015.1067453

Gagniuc, P. A. (2017). *Markov Chains: From Theory to Implementation and Experimentation*. John Wiley & Sons.

Goldston, D. (2008). Big Data Wrangling. *Nature*, *455*(7209), 15.

Huang, Y. C., Lin, H., Hsu, Y. L., & Lin, J. L. (2012). Using n-gram analysis to cluster heartbeat signals. *BMC Medical Informatics and Decision Making*, *12*(1). https://doi.org/10.1186/1472-6947-12-64

Hussein, K. M. (2021). Review in Digital Data in Supporting of Information Technologies. *Journal of Digital Integratet Circuits in Electrical Devices*, *6*(1), 23–29.

Jones, S. M. (2021). *Advancing a Circular Economy. A future without waste?*

Kennedy, P. (1996). *Büyük Güçlerin Yükseliş ve Çöküşleri: 1500'den 2000'e Ekonomik Değişme ve Askeri Çatışmalar* (6. Baskı). Birtane Karanakçı (Çev.) Türkiye İş Bankası Yayınları.

Lin, Y., Wang, H., Li, J., & Gao, H. (2019). Data Source Selection for Information Integration in Big Data Era. *Information Sciences*, *479*, 197–213.

Magnusson, L. (2015). The political economy of mercantilism. *The Political Economy of Mercantilism*, 1–230. https://doi.org/10.4324/9781315694511

Mavropoulos, A., & Nilsen, A. W. (2020). *Industry 4.0 and Circular Economy*. Jhon Wiley&Sons Ltd.

Raudeliuniene, J., Albats, E., & Kordab, M. (2020). Impact of information technologies and social networks on knowledge management processes in Middle Eastern audit and consulting companies. *Journal of Knowledge Management*, *25*(4), 871–898. https://doi.org/10.1108/JKM-03-2020-0168

Simmons, C. P. (1976). Indigenous Enterprise in the Indian Coal Mining Industry. *Indian Economic and Social History Review*, *13*(2), 189–217.

Stern, P. J., & Wennerlind, C. (2014). Mercantilism Reimagined Political Economy in Early Modern Britain and Its Empire. In *Mercantilism Reimagined Political Economy in Early Modern Britain and Its Empire*. Oxford University Press, Inc. https://doi.org/10.1093/acprof:oso/9780199988532.001.0001

Ünalan, G. (2013). *Kömür Jeolojisi*. Maden Tetkik ve Arama Genel Müdürlüğü.

Ural, M. N., Vural, A., & Çiftçi, A. (2020a). Analysis of Rare Earth Elements (REE) in the Literature by Using N-gram Method and Comparison with Social/ Political/ Economic Even. *Journal of Social, Humanities and Administrative Sciences*, *6*(24), 369–379.

Ural, M. N., Vural, A., & Çiftçi, A. (2020b). Conceptual Analysis of the Change in the Historical Process in Thermal Tourism. *Journal of International Health Sciences and Management*, *6*(12), 67–73.

Ural, M. N., Vural, A., & Çiftçi, A. (2020c). Conceptual Development Analysis of Health Tourism Based on "N-Gram." *5.Uluslararası Sağlık Bilimleri ve Yönetimi Kongresi, 9-11 Temmuz 2020, Kırşehir/Türkiye*.

Ural, M. N., Vural, A., & Çiftçi, A. (2019). Analysis of Rare Earth Elements (REE) in the Literature by Using N-gram Method and Comparison with Social/Political/Economic Events. *4. Uluslararası GAP Sosyal Bilimler Kongresi*, 135–144.

Ural, M. N., Vural, A., & Çiftçi, A. (2020d). N-Gram Analysis of Nuclear and Alternative Energy Sources and Historical and Political Course. *Euroasia Summit Congress on Scientific Rearches and Recent Trend-6*, *May*, 582–586.

Vural, A. (1998). *Güneyköy ve Çevresi (Eşme-Uşak) Arsenopirit Cevherleşmelerinin Maden Jeolojisi*. Ankara Üniversitesi.

Vural, A. (2006). *Bayramiç (Çanakkale) ve Çevresindeki Altın Zenginleşmelerinin Araştırılması*. Ankara Üniversitesi.

Vural, A., & Aydal, D. (2016). Bayramiç ve Yakın Çevresindeki Altın Zenginleşmelerinin Araştırılması. *69. Türkiye Jeoloji Kurultayı*, 376–377.

Vural, A., & Çiftçi, A. (2021a). An analysis of some concepts related to environmental issues and development by N-gram. *Euroasia Journal of Social Sciences & Humanities*, *8*(2), 18–28. https://doi.org/http://dx.doi.org/10.38064/eurssh.158

Vural, A., & Çiftçi, A. (2021b). Analysis of Raw Material Supply-Demand Relationship Using N-Gram: Chrome Mine Example. *Euroasia Journal of Social Sciences and Humanities*, *8*(1), 1–9.

Vural, A., Çiftçi, A., & Ural, M. N. (2020). Kıymetli Taşlar ve Sömürgecilik: Dijital Veri Materyallerinin Analizi Örneğiyle (Precious Stones and Colonialism: Example of Analysis of Digital Data Materials). *Euroasia Journal of Mathematics, Engineering, Natural & Medical Sciences*, *7*(13), 122–134. https://doi.org/10.38065/euroasiaorg.404

Vural, A., Çiftçi, A., & Ural, M. N. (2021a). Kıymetli Taşlar ve Sömürgecilik. *73. Türkiye Jeoloji Kurultayı*, 237–241.

Vural, A., Çiftçi, A., & Ural, M. N. (2021b). Precious Stones and Colonialism. *73rd Geological Congress of Turkey*, 237–241.

Vural, A., Kaya, S., Başaran, N., & Songören, O. T. (2009). *Anadolu Madenciliğinde İlk Adımlar*. Maden Tetkik ve Arama Genel Müdürlüğü, MTA Kültür Serisi-3.

Vural, A., & Ünlü, T. (2020). The geology and mineralogical / petrographic features of Umurbabadağ and its surroundings (Eşme , Uşak - Turkey). *Journal of Engineering Research and Applied Science*, *9*(2), 1561–1587.

Vural, A., & Ünlü, T. (2016). Güneyköy ve Çevresindeki Kalıntı Altınlı Arsenopirit Cevherleşmelerinin Maden Jeolojisi Açısından İncelenmesi. *69. Türkiye Jeoloji Kurultayı*, 374–375.

Vural, A., Ural, M., & Çiftçi, A. (2020a). Evaluation of Historical Development of Some Concepts Related To Development And Environmental Issues with N-Gram Analysis. *International Black Sea Coastline Countries Scientific Research Symposium- V. November 28-29, 2020 / Zonguldak, Turkey*.

Vural, A., Ural, M. N., & Çiftçi, A. (2020b). N-Gram Analysis of Raw Material Supply-Demand Relationship: In Case of Chromium. *International Black Sea Coastline Countries Symposium-5*, 60–61.

Vural, A., Ural, M. N., & Çiftçi, A. (2019). N-gram Yöntemi İle Değerli Metallerin Sosyal/Siyasal/Ekonomik Olaylarla İlişkisinin Değerlendirilmesi. *4. Uluslararası GAP Sosyal Bilimler Kongresi*, 125–134.

Vural, A., Ural, N., & Çiftçi, A. (2020c). Değerli Metallerin Sosyal / Siyasal / Ekonomik Olaylarla İlişkisinin N- gram Yöntemi İle Değerlendirilmesi. *Social Mentality and Researcher Thinkers Journal*, *6*(29), 247–257.

Wang, J., Yang, Y., Wang, T., Simon Sherratt, R., & Zhang, J. (2020). Big data service architecture: A survey. *Journal of Internet Technology*, *21*(2), 393–405. https://doi.org/10.3966/160792642020032102008

Yılmaz, S., & Kalkan, D. K. (2017). Enerji Güvenliği Kavrami: 1973 Petrol Krizi Işığında Bir Tartışma. *Uluslararası Kriz ve Siyaset Araştırmaları Dergisi*, *1*(3), 169–199. http://dergipark.gov.tr/uksad/issue/33359/371272

Journal of Investigations on Engineering & Technology
Volume 5, Issue 1, 21-33, 2022

ISSN: 2687-3052

Research Article

The Heavy Metals and Minor Elements Effects of Mineralization and Alteration Areas with Buried Ore Deposits Potential on the Surface Waters

Alaaddin VURAL[1], Ali GÜNDOĞDU[2], Fatih SAKA[3], Volkan Numan BULUT[4], Mustafa SOYLAK[5], Selçuk ALEMDAĞ[6]

(Alınış / Received: 07.04.2022, Kabul / Accepted: 19.06.2022, Online Yayınlanma / Published Online: 30.06.2022)

Keywords
Heavy metal contamination,
Alteration area,
Surface water,
Gümüşhane

Abstract: This paper aims to elucidate the potential environmental risks associated with ore deposits and alteration areas related to these mineralizations by quantifying contamination in stream waters. For this purposes, 23 surface water samples were collected from the streams and analyzed by inductively coupled plasma-mass spectrometry (ICP-MS) for heavy metals and minor elements. Concentrations of heavy metals in surface waters reached 160 µg/L for Al, 288.6 µg/L for Mn, 13.2 mg/L for Fe, 58.2 µg/L for Co, 164.4 µg/L for Zn, 100.2 µg/L for Cu, 2.4 mg/L for Sr, 77.0 µg/L for Ba, 96.7 µg/L for Pb, 115.4 µg/L for V, 101.5 mg/L for Ca, 15.0 mg/L for Mg, 24.4 µg/L for Ni, and 16.0 mg/L for Na. In addition, high arsenic, cadmium values (137.9 µg/L for As, 6.7 µg/L for Cd), were detected in some locations. Some other heavy metals are below the limit of quantification (LOQ). At the end of the study, a heavy metal pollution risk has been reached in the surface waters examined from the instant data obtained, caused by various anthropogenic, agricultural and industrial activities, as well as alteration areas with buried ore deposit potential and ore deposit areas. More detailed studies should be carried out by increasing the number and period of sampling in such areas and precautions should be taken against possible heavy metal contamination risks.

1. Introduction

It is estimated that the water in the world water cycle is $1.380000.10^{12} m^3$ and 3/4 of the earth's surface is covered with water. Despite the abundance of water, most of it is salty and only 2.5% is freshwater [1]. A very small portion of freshwater is drinkable. 68.7% of these waters are in the form of ice and snow. 1.2% is in lakes, rivers, wetlands and polar regions and only 30.1% is in the form of fresh groundwater [1]. Especially with the rapid population growth and industrialization of the world, water resources (especially freshwater resources), today, are one of the

[1] Gümüşhane University, Department of Geological Engineering, Gümüşhane/Türkiye. ORCID: https://orcid.org/0000-0002-0446-828X
[2] Karadeniz Technical University, Macka Vocational School, Trabzon/Türkiye. ORCID: https://orcid.org/0000-0002-9594-4121
[3] Karabük University, Faculty of Engineering/Department of Civil Engineering, Karabük/Türkiye. ORCID: https://orcid.org/
[4] Karadeniz Technical University, Macka Vocational School, Trabzon/Türkiye. ORCID: https://orcid.org/0000-0003-2192-7043
[5] Erciyes University, Faculty of Science, Department of Chemistry, Kayseri/Türkiye. ORCID: https://orcid.org/0000-0002-1017-0244
[6] Gümüşhane University, Department of Geological Engineering, Gümüşhane/Türkiye. ORCID: https://orcid.org/0000-0003-2893-3681

Corresponding Author and e–mail address: Alaaddin Vural, alaaddinvural@hotmail.com

most vital natural resources for all countries. Water management policies are now at the top of the priorities of all countries. In the past, uncontrolled/unconscious industrialization moves have caused human-induced pollution of water resources over time. Now, nowadays water contamination/pollution is a major problem for all countries related to the economic/industrial growth of any country. Water pollution is not only caused by anthropogenic factors, but also by mixing trace elements released into the environment through natural processes, water pollution can occur depending on natural processes [2].

Contrary to popular misconception, Türkiye is not a water-rich country, on the contrary, it is among the countries in the middle risk group suffering from water stress [3]. Considering the amount of usable water per capita per year (1400 m^3 according to 2017 data), Türkiye is in the class of countries facing water scarcity [4]. It is predicted that Türkiye's population will reach 100 million in 2040, and the annual amount of usable water per person will decrease to 1120m^3. Therefore, the sustainable management of Türkiye's water resources is a national issue that is the responsibility of all sectors, both on the basis of society, the public and the individual. Based on the strategic importance of water for Türkiye, a Climate Change Strategy document has been prepared by the Government of Türkiye. In addition, the National Climate Change Adaptation Strategy and Action Plan were prepared by the Ministry of Environment and Urbanization and the targets in water resources management were set. In addition, the name of the ministry was changed to the Minister of Environment, Urbanization and Climate Change on October 29, 2021.

When evaluated in terms of water resources, Gümüşhane has a potential above Türkiye's average. In addition to large rivers and streams (Harşit River, Çitderesi River etc.) there are lakes, albeit on a small scale. Although it has decreased in recent years, the amount of precipitation it receives in the form of snow is at a significant level, and snow can remain until the end of June and July in the high areas. This ensures that the creeks of Gümüşhane are flowing in almost every season. The region is also in an important metallogenic belt and is home to many mineral deposits such as copper, lead, silver, gold etc. [5,6]. Although mineral deposits are important raw materials in the development of countries, they also bring some environmental problems. Such environmental problems bring the risk of pollution to the soil, plants and especially the waters of the region [7,8].

The Arzular-Kabaköy, Kaletaş, Dölek and its surrounding areas are situated at the east of the Gümüşhane city (NE Blacksea Region) and these areas host either mineral deposits (Arzular-Kabaköy gold deposit, Kaletaş gold mineralization) or hydrothermal alteration areas with buried mineralization potential. There are many large (Arzular Stream, Kabaköy Stream, etc.) and small streams (Yoncalık creek, Kalıyaz creek, Yetirmez Creek, etc.) in the area. City, town and especially village residents mostly get their drinking water from these streams. These streams are also used by the local people for irrigation activities for agricultural purposes. Considering the litho-chemical, geological and metallogenic characteristics of the region, the surface waters in the area have the risk of heavy metal pollution/contamination. The aim of the study is to investigate the heavy metal pollution risk of the streams and creeks in Arzular-Kabaköy, Kaletaş, Dölek and their surroundings and to determine the water quality of the surface waters.

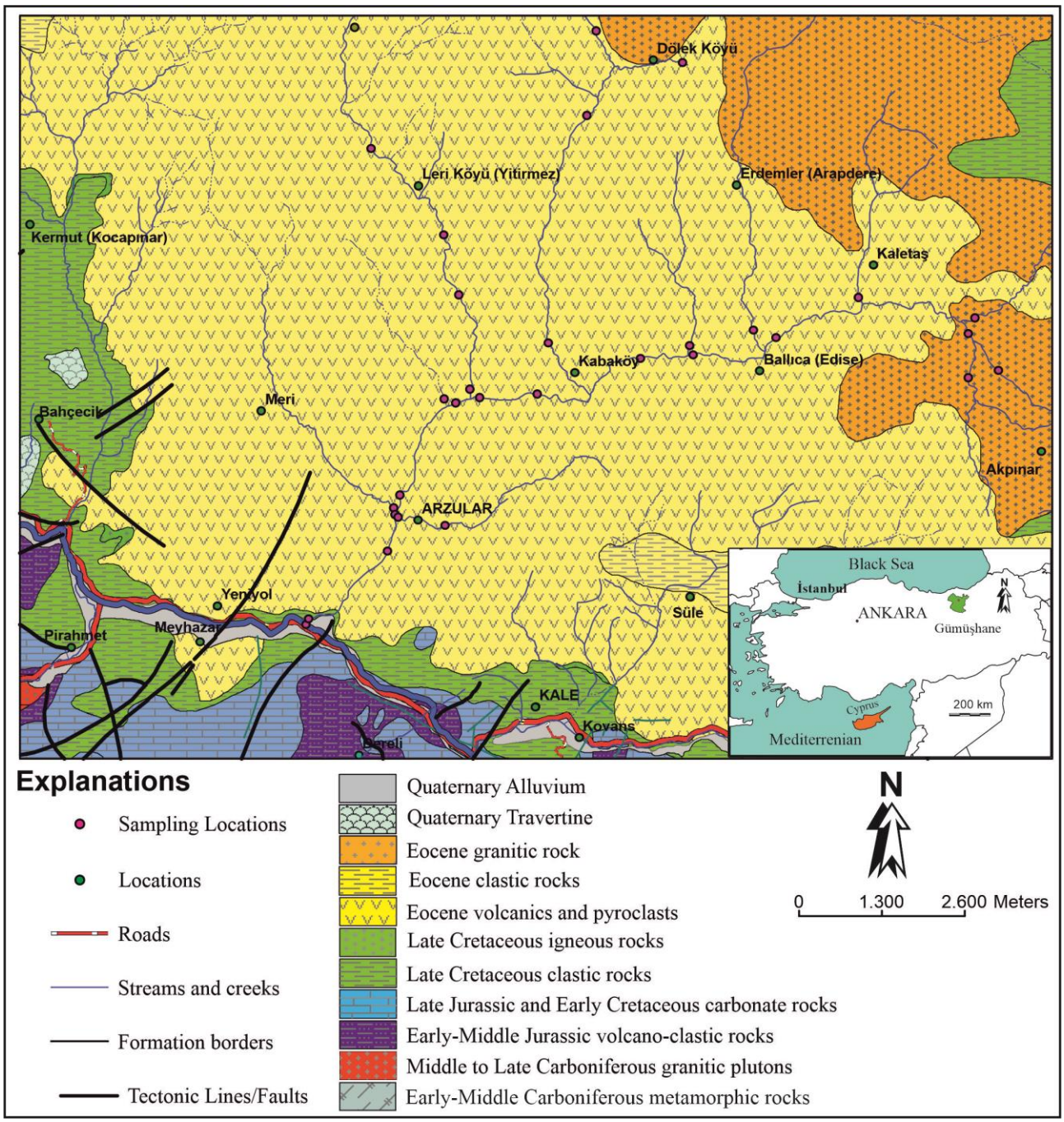


Figure 1. Geological and location map of the study area (After [9])

2. Material and Method

2.1. Geological and Hydrological/Hydrogeological Features of The Study Area

The study area covering Arzular-Kabaköy, Kaletaş, Dölek and its vicinity is located in the southern zone of the Black Sea Tectonic Unit (Fig. 1). In the Gümüşhane province, it can be observed all stratigraphic units of the region. If the general geology of the site is mentioned, the basement geological unit of the region consists of pre-Jurassic (Early-Middle Carboniferous) metamorphic rocks [10] and middle-late Carboniferous unmetamorphosed granitoids [11,12]. These basement units are unconformably overlain by an early to middle Jurassic volcani-clastic unit which passes upward to the late Jurassic and early Cretaceous carbonate rocks. In the area, late Cretaceous clastic unit begins with sandy limestones at the bottom and grades upward to red pelagic limestone and then to a turbiditic series which conformably overlie late Jurassic and early Cretaceous carbonates. All these units -outside of the study area, in Torul/Gümüşhane and its near vicinity- were cross cut by late Cretaceous intrusions [13] and were cut and cover by late Cretaceous volcanic rocks [14]. The Eocene volcanic and volcani-clastic rocks overlie the late Cretaceous volcanic and/or sedimentary rocks with an angular unconformity [15], and intruded by calc-

alkaline granitoids of same age [16]. The youngest units in the region are actual travertine occurrences [17], talus and alluvium.

All lithological units belonging to the geology of the region are observed in the area that is the subject of this study on the risks of heavy metal and/or minor elements on surface waters (Fig. 1). Pre-Jurassic metamorphic rocks are observed in the southern parts of the study area, while unmetamorphosed granitoids crop out in the southwestern part of the study area. The Early to middle Jurassic volcani-clastic unit, late Jurassic and early Cretaceous carbonate rocks, and late Cretaceous clastic crop out to the west and southwest of the study area.

In the study area, dominantly Eocene volcani-clastic rocks outcrop in the vast majority of the area and these rocks undergone intense hydrothermal alteration and weathering and host some epithermal ore deposits such as Arzular gold deposits, Kaletaş gold mineralizations. Actual travertine occurrences are also observed around Bahçecik and Tekke villages in the area [17], and there are alluviums in the stream banks.

2.2. Sampling, Reagents-Solutions and Analysis

For elucidating the potential environmental risks associated with mineralization and alteration areas with buried ore deposits potential by quantifying the heavy metal contamination in stream waters in the areas, 23 surface water samples were collected from the streams (Fig. 1) in September 2016, and analyzed for some heavy metals and minor elements by Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) (Agilent 7700e model ICP-MS, Santa Clara, US) at Gümüşhane University Central Research Laboratory, Gümüşhane, Turkey. Water pH measurements were also performed in the area during sampling.

Surface water samples were taken with 500 mL PP bottles. Prior to sampling, the bottles were thoroughly rinsed with tap water, washed with 1:10 dilute nitric acid, then again well washed with tap water, then ultrapure water, and finally labeled. As soon as water samples were taken for metal analysis, they were acidified with 0.5 mL of ultrapure concentrated HNO_3. After the water samples were brought to the laboratory during the day, they were filtered through a vacuum filtration device with a 0.45 µm nitrocellulose membrane. The clear water samples were then analyzed in terms of some heavy metal/minor element contents using the ICP-MS.

2.3. Quality Assurance

Both accuracy tests were made with some parameters of the measurements made in ICP-MS and the performance of the system was tested. First, the calibration method used for measurements on the device was the internal standard method to eliminate possible interference [18]. ^{45}Sc, ^{89}Y, ^{185}Re, ^{209}Bi isotopes were used as the internal standards.

Table 1 reflects some analytical performance criteria (RSD, LOD and LOQ) of the ICP-MS instrument used in metal measurements. The limit of detection (LOD), expressed as the smallest signal (or concentration) that the relevant method/device can detect for each metal, was calculated by taking 3 times the standard deviation of the results obtained from the measurement of a series of blank solutions (approximately 20). Similarly, the limit of quantification (LOQ), which is the lowest signal or concentration quantitatively sensed by the instrument/method, is 10 times the standard deviation of the results from blank solutions. Relative standard deviation (RDS), the most common indicator of the repeatability, was determined by dividing the standard deviation of the results obtained from 20 repeated measurements of a series of multi-element standard solutions at very low concentration by the mean value. This ratio was multiplied by 100 and the results were converted to the percent relative standard deviation (RSD%). The accuracy of the method has been proven in two ways. The first is the spiked/recovery tests, and the other is the certified reference material (CRM-TMDW-A Trace Metals in Drinking Water) analysis. As a result of both tests, very satisfactory results were obtained with recovery values between 95–105% [19,20].

Table 1. Analytical performance of the device, Agilent 7700x ICP-MS

	1	2	3	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Parameter	**pH**	**Ca**	**Mg**	**Na**	**K**	**Sr**	**Fe**	**Ba**	**Zn**	**Cu**	**Al**	**Mn**	**Cd**	**Co**	**Ni**	**Pb**	**Cr**	**V**	**As**	**Rb**
RSD (%)	0.02	3.1	2.8	3.0	3.6	1.2	3.4	2.6	2.0	2.6	4.1	2.2	3.0	3.4	2.3	3.3	3.8	4.2	3.8	2.8
LOD (µg/L)	–	2.2	1.4	3.2	2.2	2.9	3.1	2.5	2.7	2.5	3.5	3.1	2.0	3.0	3.1	3.0	3.0	2.7	2.1	2.7
LOQ (µg/L)	–	7.3	4.7	10.7	7.3	9.7	10.3	8.3	9.0	8.3	11.7	10.3	6.7	10.0	10.3	10.0	10.0	9.0	7.0	9.0

3. Result and Discussion

Descriptive statistics parameters for pH values and some minor-trace elements and heavy metals contents of the streams in the study area are given in Table 2. Figure 2 and Figure 3 also reflect the distributions of these contents of a total of 23 water samples. In addition, dot distribution maps plotted for the same parameters can be seen in Figure 4. Since some elements are detected only in limited locations above the detection limit (such as, Cr, Co, Cd), and some (Na, Mg etc.) are abundant in nature under normal conditions, their point distribution maps have also been created (Figure 4).

Interestingly, the Ca, Mg and Na contents of five samples between w-28 and w-32 coded samples are at very low levels (Figures 2 and 4). These samples were taken from the small branches of the main stream and were probably fed by rain water. Therefore, their effects on distribution are also great. Since the levels of some metals were below the LOQ value, these metals could be detected in a limited number of samples. In this respect, when Table 2 is examined, Cr could be determined in only two samples, and Cd and Co in three samples. The fact that domestic wastes (batteries, parts of some household appliances and electronic devices, etc.) were occasionally encountered at the points where these samples were taken indicate anthropogenic pollution at these points.

It has been found that the pH values of the waters vary between 7.02 to 8.68 (Table 1 and Figures 2 and 4). According to these values, waters are between slightly neutral and generally basic but some of them are near neutral to acidic in especially intense altered areas such as Leriköy, northwest of Kabaköy, northwest of Akpınar etc. (Figure 1). It has been observed that pH values are more alkaline, especially in areas exposed to anthropogenic effects. Probably due to detergent-based household waste.

It has been observed that Mn and Fe concentrations are high especially in two samples coded w-34 and w-35 in the areas close to granitic intrusions in the northeast of the field (Figure 2). Considering that the mobilities of Mn and Fe are close to each other, the observed results met the expectations (Figure 4). Cu concentrations were observed at high values around Dölek village, around Arzular village and northwest of Akpınar village, where intense alteration development occurs. Zn values also show a similar pattern with Cu because they have similar physicochemical properties. While the Al values of the samples are generally below 40 µg/L, the three samples coded w-35, w-36 and w-36 have high contents with values of 109.4 µg/L, 160.0 µg/L and 117.4 µg/L, respectively. Vanadium (V) could be determined quantitatively in 14 samples. While the results were below 20 mg in most samples, 115.4 mg V was detected in the w-44 coded sample, 63.4 in the w-50 coded sample, and 47.5 mg V in the w-45 coded sample. Although there is no remarkable situation in other metal contents, the results show a very heterogeneous distribution. It is expected that the clarification of the origins of these anomalies will be clarified with the completion of the other parts of the study (periodical water results, stream sediment and soil data)

The metal contents of the studied water samples were compared with the EPA's guidelines for surface waters. Table 3 summarizes these guideline values. The red dots on the graphs represent samples that exceed the EPA-reported limit values (Figures 2 and 3). The surface water's pH values are within the standards but some HMTEs were above the EPA's standards. Mn values were found to be higher than EPA standards, especially at Dölek village alteration zone. In addition to the two samples with high Mn content, w-34 and w-35, the other two samples, w-25 and w-36, are slightly above the guideline value. As to Fe, those values are generally high in whole area. In terms of Fe content, while many samples are outside of the A1 class surface water quality, two samples coded w-34 and w-35 have extremely high Fe content with 13.15 mg/L and 13.14 mg/L contents, respectively. High Mn and Fe values are associated with the intense hydrothermal alteration of the region, and since Mn and Fe are mobile elements, they can easily pass into water at surface conditions and at appropriate pH values. The stream from which these samples were taken should be monitored especially in terms of Mn and Fe.

w-34 and w-35 coded samples have high values for Co metal, which can be detected only in three samples. While Ni can be detected at four sample points similar to Co, the w-34 and w-35 coded samples have the highest value here too. These values slightly exceed the EPA's guidelines for drinking water. The fact that household wastes (battery, electronic device fragments and small household appliances, etc.) are encountered at sampling points suggest that high Ni and Co values are probably anthropogenic. EPA reports the limit value for Ba content of A1 class surface waters as 0.1 mg/L. The Ba contents determined in twenty samples have a very heterogeneous distribution as seen in Figure 2 and Figure 4, but no sample exceeds the guide value.

Cd values were also determined at only three points. These values were close to the upper limit of the standard values.

Perhaps the most critical values in this study are Pb values. Pb values in many samples were above the LOQ, but the content of eight samples exceeded the EPA-reported threshold of 0.05 mg/L. Wastes from various factories, exhaust gases, pollution caused by mining operations and of course Pb pollution caused by natural geological formations cause significant health problems in waters.

Arsenic detected at seven points in total is generally below 20 mg/L. However, the w-28 coded sample with an As value of 137.9 mg/L far exceeds the 50 mg/L limit value reported by the EPA. Known as the "king of poisons", arsenic enters the food chain through many activities. As can enter the waters with many anthropogenic activities, geological formation and mining activities also trigger the As level in the water. It is seen that high As values are mostly close to the region where volcanic rocks and granitic rocks cut them. It is thought that the hydrothermal alterations developed due to granitic intrusions in these regions increase the As concentration in these hydrothermal alteration zones, and therefore the As values in the waters interacting with these units also increases (Figure 4).

Minor elements such as Ca, Mg, Na and K, which are more abundant in waters (Figures 2 and 4) and do not pose much threat to health compared to heavy metals and trace elements, are also extremely important especially for the sustainability of aquatic life. However, although they are important for health, very high amounts of them are still undesirable. In this respect, the determination of minor elements, in other words mineral elements, is also very important. The content of minor elements in water were determined within the limit values of EPA's surface water standards.

Table 2. Descriptive statistics of the surface waters and analytical performance criteria of the analysis method

	1	2	3	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
		mg/L						µg/L												
Parameter	**pH**	**Ca**	**Mg**	**Na**	**K**	**Sr**	**Fe**	**Ba**	**Zn**	**Cu**	**Al**	**Mn**	**Cd**	**Co**	**Ni**	**Pb**	**Cr**	**V**	**As**	**Rb**
Mean	8.02	54.1	7.0	7.4	10.1	1.2	1.6	30.5	65.5	43.0	47.2	64.9	6.6	43.1	17.3	50.2	20.7	26.4	27.8	12.1
Median	8.07	57.4	6.9	6.9	12.1	1.1	0.4	26.2	50.1	32.5	25.3	24.0	6.6	57.4	17.4	45.4	20.7	13.5	7.8	11.4
Std deviation	0.43	33.5	4.9	4.0	4.9	0.5	3.7	18.1	34.6	24.0	48.3	99.7	0.0	25.5	17.3	32.8	13.4	30.1	48.6	2.1
Min.	7.02	1.4	0.1	0.1	0.4	0.7	0.02	8.3	22.6	13.9	12.2	10.9	6.6	13.6	10.2	10.1	11.2	9.3	7.2	9.9
Max.	8.68	101.5	15.0	16.0	20.3	2.4	13.1	77.0	164.4	100.2	160.0	288.6	6.7	58.2	24.4	96.7	30.2	115.4	137.9	16.1
Number	20	23	23	20	23	23	23	20	23	23	13	13	3	3	4	16	2	14	7	7

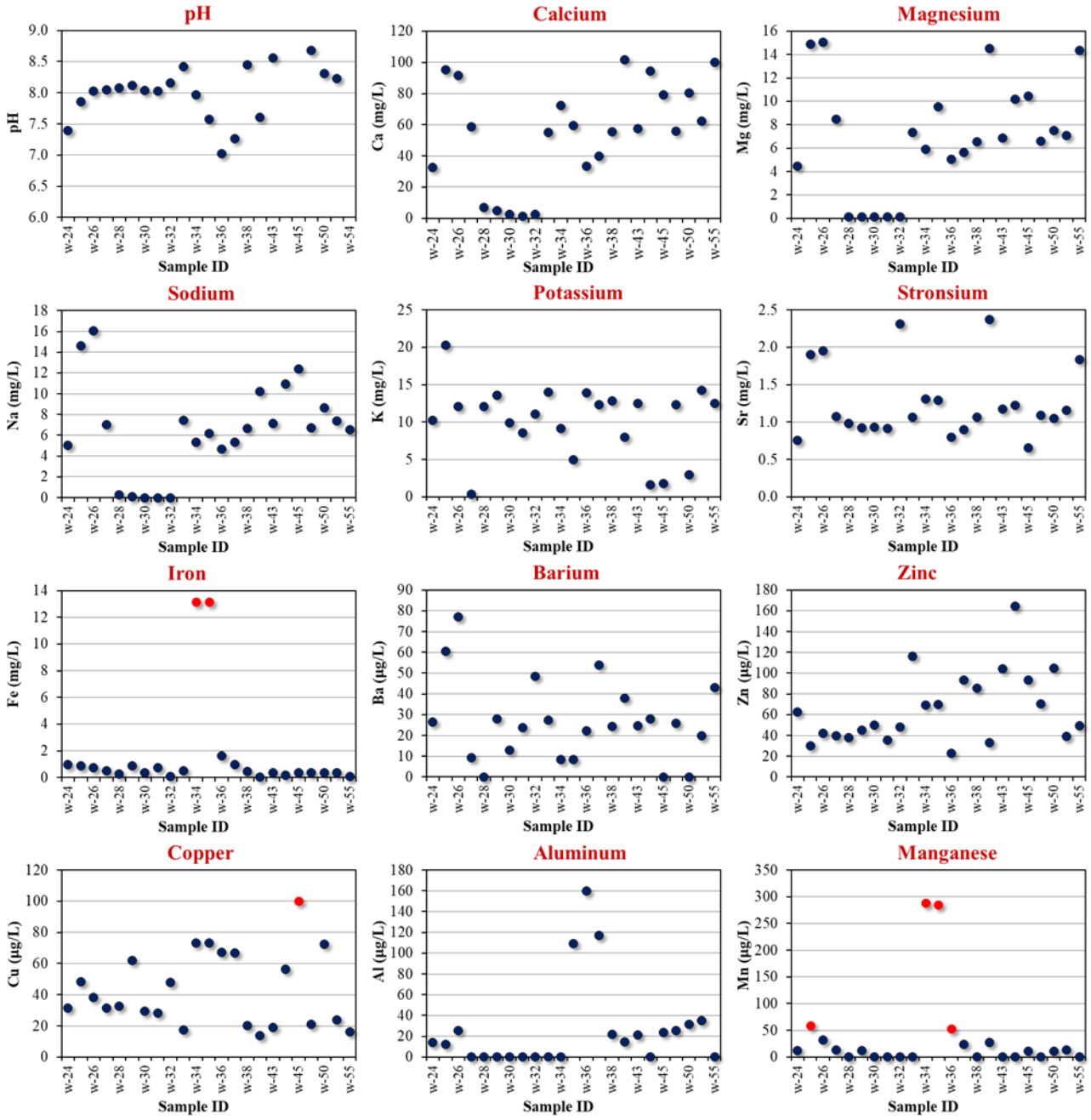


Figure 2. Distributions of pH and metal contents of the water samples

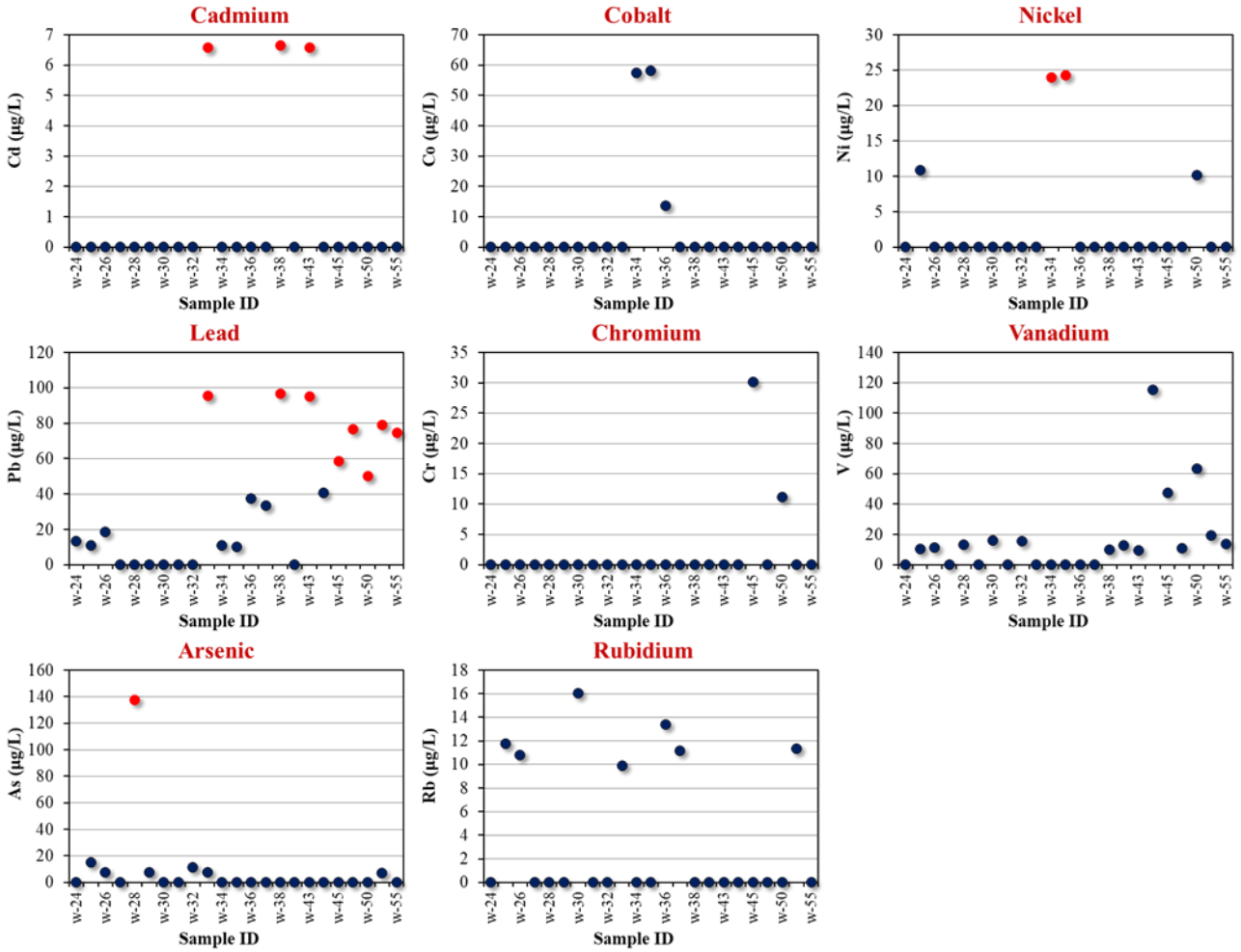


Figure 3. Distributions of metal contents of the water samples

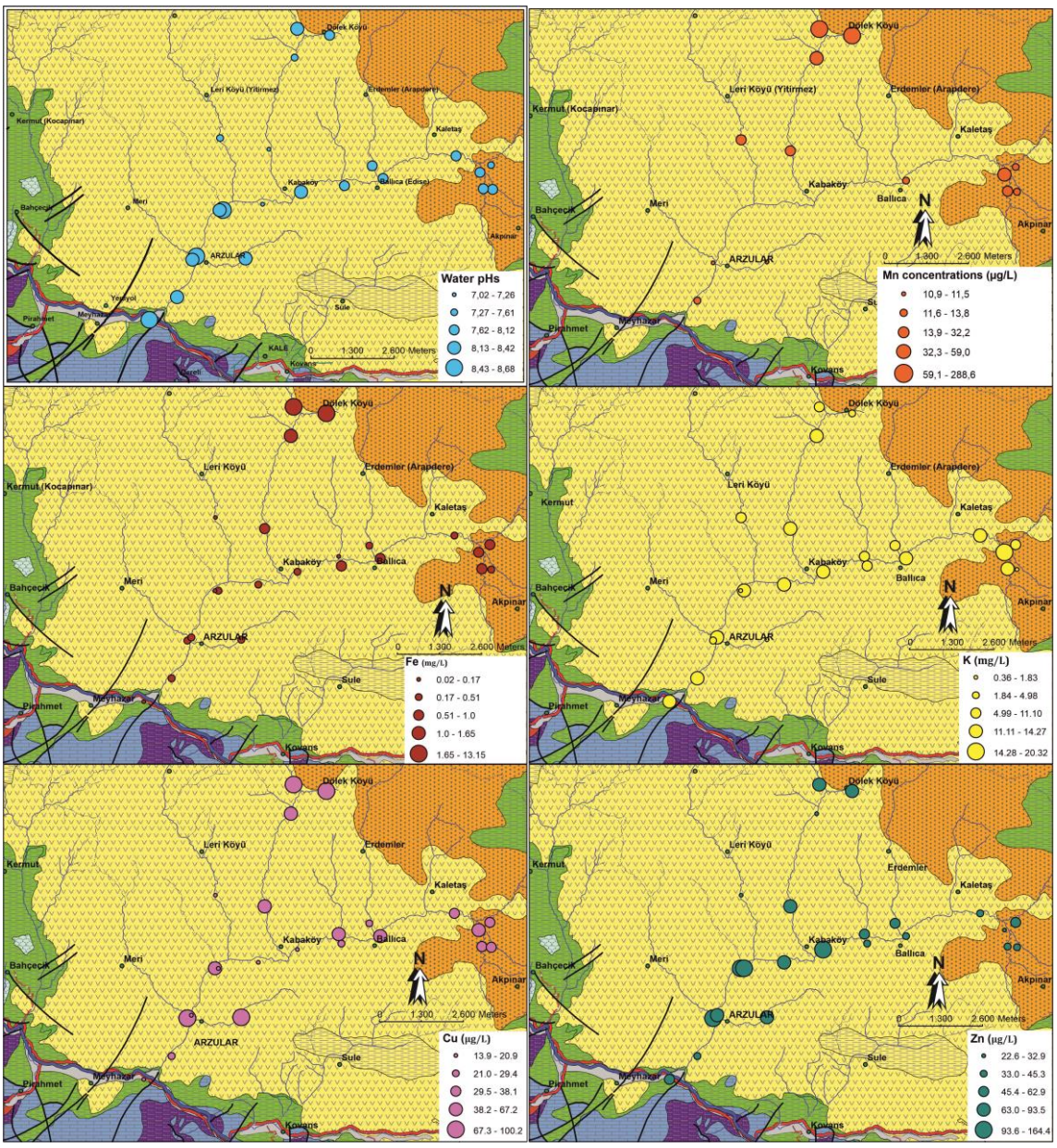


Figure 4. Dot distribution map of HMTEs, minor elements and pH values of the study area streams (Explanations of the geological map are given in Fig. 1)

Figure 4 (Continue) (Explanations of the geological map are given in Fig. 1)

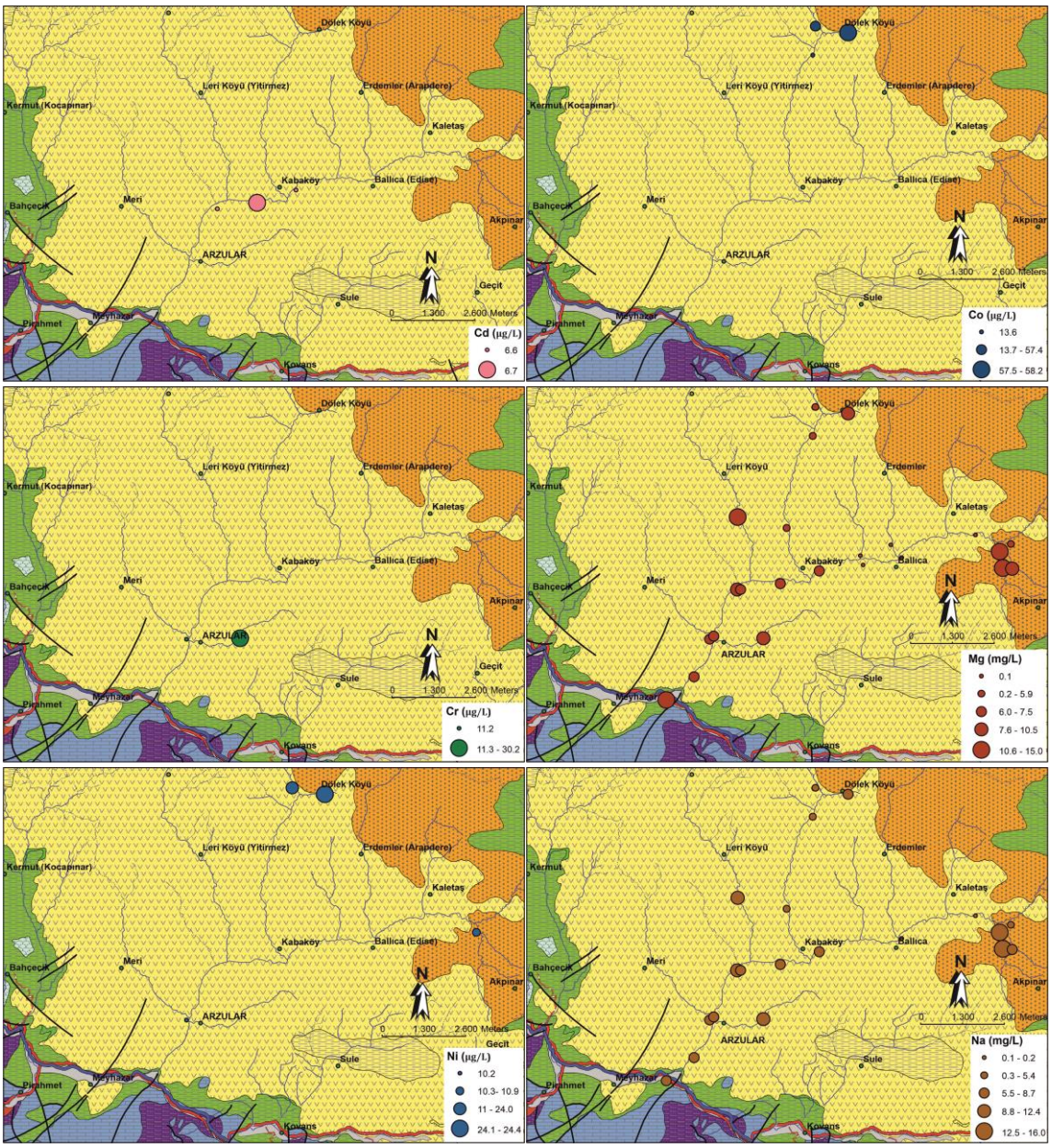


Figure 4 (Continue) (Explanations of the geological map are given in Fig. 1)

Table 3. EPA Water Quality Standards for some elements

			Surface water		
Element	**Unit**	**Drinking water**	**A1 class**	**A2 class**	**A3 class**
pH		6.5 – 9.0	5.5 – 8.5	5.5 – 9.0	5.5 – 9.0
Al	µg/L	200	-	-	
As	mg/L		0.05	0.05	0.1
Ba	mg/L		0.1	1	1
Cd	mg/L		0.005	0.005	0.005
Ca	mg/L		-	-	-
Cr	mg/L		0.05	0.05	0.05
Co	mg/L		-	-	-
Cu	mg/L		0.05	0.1	1
Fe	mg/L	0.2	0.2	2	2
Pb	mg/L		0.05	0.05	0.05
Mg	mg/L		-	-	-
Mn	mg/L	0.05	0.05	0.3	1
Hg	µg/L	1	1	1	1

Mo	µg/L	-	-	-	-
Ni	µg/L	20	-	-	-
K	mg/L	-	-	-	-
Se	mg/L	0.01	0.01	0.01	0.01
Na	mg/L	200	-	-	-
Sn	mg/L	-	-	-	-
Zn	mg/L	-	3	5	5

As a result, these findings point out that the surface waters in the Arzular-Kabaköy, Kaletaş, Dölek and its surrounding areas were affected by some HMTEs such as Mn, Fe, Cd, Co, As and Pb. Although there is an anthropogenic effect especially for Ni, Cr, Co, Cd at some points, ore deposit sites and related hydrothermal alteration areas with buried ore deposits potential have been found to cause heavy metal contamination in the surface waters around these environments, therefore, more detailed studies should be done on such fields and necessary precautions should be taken in order to avert the dispersal of contamination sourced those areas.

Studies on water quality parameters and other physical properties in the region are continuing. These data are the first to be obtained, when other data are obtained then they will be evaluated separately in different articles.

Acknowledgement

This study was supported by TUBITAK (Grant Number: 115Y146).

Author Contribution

Corresponding author Alaaddin Vural conducted field studies during the data collection process, took part in the analysis and interpretation of the data, and undertook the writing of the article. Ali Gündoğdu both contributed to the field studies and took part in the analysis. Fatih Saka contributed to the field studies. Volkan Numan Bulut took part in the analysis. Selçuk Alemdağ contributed partially to the collection of the samples used in this article from the field. Mustafa Soylak contributed to the evaluation of the analysis data obtained.

Conflict of Interest

The authors confirm that there is no known conflict of interest or common interest with any institution/organization or person.

References

[1] Hölting, B. and Coldewey, W.G. (2019) Hydrogeology. Springer, .
[2] Bird, G., Brewer, P.A., Macklin, M.G., Serban, M., Balteanu, D., and Driga, B. (2005) Heavy metal contamination in the Arieş river catchment, western Romania: Implications for development of the Roşia Montana gold deposit. *Journal of Geochemical Exploration*. 86 (1), 26–48.
[3] Karataş, M. and Çevik, S. (2010) Stratejik Doğal Kaynak Olarak Su ve Türkiye'nin Konumunun Değerlendirilmesi. *Akademik Araştırmalar Dergisi*. 45 1–29.
[4] Turan, E. and Bayrakdar, E. (2020) Türkiye'nin Su Yönetim Politikaları: Ulusal Güvenlik Açısından Bir Değerlendirme. *International Journal of Political Studies*. 6 (2), 1–19.
[5] Vural, A., Kaya, S., Başaran, N., and Songören, O.T. (2009) Anadolu Madenciliğinde İlk Adımlar. Maden Tetkik ve Arama Genel Müdürlüğü, MTA Kültür Serisi-3, Ankara, Türkiye.
[6] Vural, A. and Ersen, F. (2019) Geology, mineralogy and geochemistry of manganese mineralization in Gumushane, Turkey. *Journal of Engineering Research and Applied Science*. 8 (June), 1051–1059.
[7] Sungur, A., Vural, A., Gundogdu, A., and Soylak, M. (2020) Effect of antimonite mineralization area on heavy metal contents and geochemical fractions of agricultural soils in Gümüşhane Province, Turkey. *Catena*. 184 (January 2019), 104255.
[8] Sungur, A., Soylak, M., and Ozcan, H. (2019) Fractionation, Source Identification and Risk Assessments for Heavy Metals in Soils near a Small-Scale Industrial Area (Çanakkale-Turkey). *Soil and Sediment Contamination*. 28 (2), 213–227.
[9] Güven, İ.H. (1993) Doğu Pontidlerin 1/100.000 Ölçekli Kompilasyonu. MTA Genel Müdürlüğü, Ankara.
[10] Topuz, G., Altherr, R., Schwarz, W.H., Dokuz, A., and Meyer, H.P. (2007) Variscan amphibolite-facies rocks

from the Kurtoğlu metamorphic complex (Gümüşhane area, Eastern Pontides, Turkey). *International Journal of Earth Sciences*. 96 (5), 861–873.

[11] Yılmaz, Y. (1972) Petrology and structure of the Gümüşhane granite and surrounding rocks, NE Anatolia. .

[12] Topuz, G., Altherr, R., Siebel, W., Schwarz, W.H., Zack, T., Hasözbek, A., et al. (2010) Carboniferous high-potassium I-type granitoid magmatism in the Eastern Pontides: The Gümüşhane pluton (NE Turkey). *Lithos*. 116 (1–2), 92–110.

[13] Kaygusuz, A., Siebel, W., Şen, C., and Satir, M. (2008) Petrochemistry and petrology of I-type granitoids in an arc setting: the composite Torul pluton, Eastern Pontides, NE Turkey. *International Journal of Earth Sciences*. 97 (4), 739–764.

[14] Vural, A., Akpınar, İ., and Kaygusuz, A. (2021) Petrological characteristics of Cretaceous volcanic rocks of Demirören (Gümüşhane, NE Turkey) region. *Journal of Engineering Research and Applied Science*. 10 (2), 1828–1842.

[15] Arslan, M. and Aliyazicioglu, I. (2001) Geochemical and petrological characteristics of the Kale (Gümüşhane) volcanic rocks: Implications for the Eocene evolution of eastern Pontide arc volcanism, northeast Turkey. *International Geology Review*. 43 (7), 595–610.

[16] Karslı, O., Chen, B., Aydin, F., and Şen, C. (2007) Geochemical and Sr-Nd-Pb isotopic compositions of the Eocene Dölek and Sariçiçek Plutons, Eastern Turkey: Implications for magma interaction in the genesis of high-K calc-alkaline granitoids in a post-collision extensional setting. *Lithos*. 98 (1–4), 67–96.

[17] Vural, A. and Külekçi, G. (2021) Zenginleştirilmiş Jeoturizm Güzergahı:Gümüşhane-Bahçecik Köyü (Enriched Geotourism Route: Gümüşhane-Bahçecik Village). *Euroasia Journal of Mathematics, Engineering, Natural & Medical Sciences*. 8 (19), 1–23.

[18] Yuksel, B. and Arica, E. (2018) Assessment of Toxic, Essential, and Other Metal Levels by ICP-MS in Lake Eymir and Mogan in Ankara, Turkey: An Environmental Application. *Atomic Spectroscopy*. 39 (5), 179–184.

[19] Khan, N., Jeong, I.S., Hwang, I.M., Kim, J.S., Choi, S.H., Nho, E.Y., et al. (2013) Method validation for simultaneous determination of chromium, molybdenum and selenium in infant formulas by ICP-OES and ICP-MS. *Food Chemistry*. 141 (4), 3566–3570.

[20] Voica, C., Dehelean, A., Iordache, A., and Geana, I. (2012) Method Validation for Determination of Metals in Soils by ICP-MS. *Romanian Reports in Physics*. 64 (1), 221–231.

Journal of Investigations on Engineering & Technology
Volume 5, Issue 1, 34-46, 2022

ISSN: 2687-3052

Research Article

Determination of Fair Usage Rate and Sizing Optimization for a Site Houses Photovoltaic-Battery Energy Source

Burhan BARAN[1]

(Alınış / Received: 30.12.2021, Kabul / Accepted: 12.03.2022, Online Yayınlanma / Published Online: 30.06.2022)

Keywords
Fair usage
Optimization
Particle swarm optimization
Renewable energy
Standard deviation

Abstract: With this study, it is aimed to meet a part of the total electricity need in a site consisting of 4 houses with the photovoltaic-battery system. Partial energy to be covered by the photovoltaic-battery system has been focused on equitably distributed among these 4 houses. Standard deviation method was used for fair distribution. Standard deviation calculations were made with the conditional flow algorithm, random algorithm and excel solver. Sizing optimization studies were carried out according to this standard deviation value. Conditional flow algorithm and particle swarm optimization algorithms were used for sizing optimization. At the end of the study, the minimum standard deviation value was calculated as 0.219779. Fair usage percentages were calculated as 1.1061%, 9.1814%, 32.3474% and 57.3650% for the minimum standard deviation for each house, respectively. According to the minimum standard deviation value, the optimum result was obtained with conditional flow algorithm.

1. Introduction

In today's society, electrical energy plays an important role in human life. Electrical energy can be generated in many ways. One of these generation methods is electricity generation from fossil fuels. However, this method causes many environmental problems. As a result of the burning of coal and petroleum fuels like fossil fuels, global and local pollution problems occur. Both living lives are threatened and ecological balance is disrupted. In addition, the high cost of fossil fuels is another problem with this method. Due to these problems, it has become necessary to find new energy sources [1-2]. In this context, electricity generation based on renewable energy sources is seen as a promising solution. Solar and wind power systems in particular have become popular choices [3-4]. However, the discontinuity of their energy and their high costs are the downsides of these systems. They must be optimally sized in order to provide continuous energy and minimize costs. For this, algorithms are developed for sizing optimization in hybrid renewable energy systems [5]. The variation of solar radiation values requires PV systems to work with a storage unit. The use of renewable energy systems together with energy storage systems helps to increase the reliability of these systems and to reduce the mismatch between energy consumption and production profiles [6-7]. The optimum dimensioning of BT energy storage systems is an important element in the planning and design of microgrids [8-10]. Choosing battery numbers in a random or non-optimal size can increase the cost and system losses [11]. In addition, a fair distribution should be ensured in case of joint use of renewable energy resources. Here, what is meant by fair usage is to determine the percentage of energy to be drawn from the source according to the average consumption amount, taking into account all consumption amounts at different consumption times. It is assumed that the energy demand of the site considered in the study will primarily be taken from the renewable energy source and then from the main grid in case of the remaining energy demand. For fair usage purpose, methods such as standard deviation can be used. Two main issues should be taken into account while designing the PV-BT system. The first is the loss of power supply probability (LPSP) of the system, and the second is the system annual cost (SAC) that arises due to this possibility. To achieve this, an optimization method have been used [12].

[1] Electrical and Electronic Engineering Department, Inonu University, 44280, Malatya, Turkey

Corresponding Author: Burhan BARAN, burhanbaran@gmail.com

Ghaffari et al. used a modified crow search algorithm for sizing optimization of a hybrid system consisting of a photovoltaic panel, diesel generator and fuel cell. Total cost and LPSP criteria were taken into account. They concluded that the proposed method had a great effect on the LPSP value and gave better results than the genetic algorithm and PSO [13]. Makhdoomi et al. proposed an algorithm for the optimization of a photovoltaic, diesel generator and pumped hydro storage system. The crow search algorithm produced better results than the genetic algorithm, PSO and the original crow search algorithm [9]. Hemeida et al. worked on the optimization of wind-solar-battery systems. They concluded that it was cheaper to install the hybrid system than to install it separately [14]. Kaabeche et al. aimed to minimize the cost and sizing of a solar-wind system with their study. Ant lion optimizer, gray wolf optimizer, krill herd and jaya algorithms were used in optimization. They had obtained the most suitable solutions with three battery technologies with the jaya algorithm [15]. In the study by Mahmoud et al., optimization was carried out for batteries in a micro grid with solar-wind-battery. PSO, genetic algorithm and flower pollination algorithm methods were used. They concluded that using smart batteries could reduce the annual generation cost of microgrids [10]. In the study by Bukar et al., LPSP and cost optimization were performed for a micro-grid. Grasshopper optimization algorithm was used. It was concluded that the grasshopper optimization algorithm achieved results close to the results of the PSO and cuckoo search algorithms, and reduced the system capital cost by 14% and 19.3%, respectively [16]. In the study by Jamshidi et al., a hybrid system consisting of photovoltaic, fuel cell and diesel generator was designed. It was aimed to minimize the total cost and LPSP. Multi-objective crow search algorithm had been utilized for optimization [17]. In the study by Zhang et al., they analyzed the optimum sizing of a hybrid solar-wind system based on chaotic search, harmony search and simulated annealing algorithms. The reliability of the system was evaluated by LPSP [1]. Zhang et al. optimized a hybrid renewable system that includes hydrogen and battery storage units. Modified simulated annealing algorithm-based chaotic search and harmony search were used. As a result, it was determined that a wind-solar energy-based hybrid system with electrochemical storage offers the most cost-effective and reliable energy [6]. Astaneh et al. proposed a method to find the minimum cost structure of a lithium-ion battery-based and off-grid renewable energy system. All possible combinations were evaluated in the proposed method. It was seen that the cost in the scenario optimized according to the results decreased by 9.7% compared to the basic scenario cost [18]. Mohamed et al. aimed at the cost optimization of the grid-connected hybrid photovoltaic wind energy system with their study. They used the PSO algorithm. It was seen that the proposed algorithm provides a reliable sizing solution and responds well to changes in system parameters [4]. Chauhan et al. focused on the optimal sizing of a stand-alone integrated renewable energy system which comprises the resources of micro hydropower, biogas, biomass, solar and wind energy. A demand response strategy based on the energy consumption planning of the devices was proposed. With this strategy, it was aimed to minimize the highest hourly energy consumption in the study area. They concluded that significant savings were achieved in system size and cost [2]. In the study by Fetanat et al., a sizing optimization was made for the hybrid photovoltaic-wind energy system. Ant colony optimization for continuous domains based integer programming technique was used for optimization. It was observed that this method performs better than the others in terms of reaching the optimum solution and speed [19]. In the study by Bukar et al., an optimization study was carried out for the photovoltaic and wind energy system integrated with a fuel cell. It was aimed to create a cost-effective system with optimization [20]. Al-falahi et al. presented a comparison of single algorithms, hybrid algorithms, and software tools used to determine the number of variables in an independent solar and wind hybrid system. The evaluation parameters of all possible combinations in economic, safety, environmental and social aspects were taken into account [5]. Kerdphol et al. made a sizing optimization using frequency control based on the PSO method in order to prevent instability of the microgrid after power failure and to minimize the total cost of the system. The results showed that the optimum size of the battery energy storage system based PSO method could achieve higher dynamic performance than its traditional size [11]. conducted by Lian, it was aimed to find an equal and lowest-demand profile among all combinations in terms of energy supply. It was aimed to guide the architects to find the most appropriate consumption among different combinations of buildings with various functions and how many buildings should be in a project area. To achieve this, the standard deviation method was used. With this case study, a general methodology was developed that facilitates the work of urban designers [21].

This study, it was carried out to ensure fair usage of energy from a renewable energy source among loads. To ensure this fair usage, the standard deviation method was used. Three different (CFA, RA, ES) methods were used for calculating the standard deviation. Studies were also made for optimal sizing of the PV-BT system. For this, CFA and PSO methods were used. In the methods, technical LPSP and economically SAC research were done and the number of components was calculated depending on these concepts. In the optimization process, the first mathematical model of the PV-BT system was derived. Then, the system reliability model based on the LPSP technique and economic models based on the SAC concept presented. The optimization process aimed to find a compromise between these two goals. The decision variables were PV and BT numbers. In this respect, it is thought to be a guiding study towards ensuring the fair usage of energy obtained from a renewable energy source among loads.

The next part of the study after this section consists of the following sections. The second section is the materials section where photovoltaic, battery, inverter and load models are defined. The third section is the method section in which the standard deviation calculations performed with CFA, RA, solver and the criteria used in optimization are explained. The fourth section is the results and conclusions section, where results are obtained and discussed. Finally, the results are given in the fifth section.

2. Materials

2.1. Photovoltaic Panel

The solar radiation data of the province of Malatya used in this study were taken from the General Directorate of Meteorology [22]. These were the hourly data of 720 hours for June 2010.

PV panel values used in this study are shown in Table 1.

Table 1. PV panel technical specifications [23]

V_{oc} (V)	I_{sc} (A)	V_{max} (V)	I_{max} (A)	P_{max} (W)	Area (m^2)	Lifespan
45.	5.9	36.6	5.47	200	1.3	25

2.2. Battery

Batteries are electrochemical elements that store energy in chemical form. The amount of energy in the battery is related to the state that the battery changes from t-1 hour to t hour. This can happen in charge, discharge or neutral situations.

Technical specifications of the batteries used in this study were as in Table 2.

Table 2. Battery technical specifications [24]

Capacity (W)	Voltage (V)	Current (A)	Battery efficiency (%)	Lifespan
900	12	75	87	4

Batteries had 12 V/75 Ah characteristics and their lifespan was assumed to be 4 years. The design of the battery system to be installed was planned in a way that it could feed the total electricity for 1 hour. Based on the 720-hour consumption values of four houses determined as the load, hourly average of the demanded load was determined as 1600 Wh. The batteries used in the study were 12 V/75 Ah and the batteries with a maximum charge depth of 87%. Accordingly, the minimum number of batteries that could feed the load for 1 hour:
1600 (Wh) ×1 / 12 (V) = 133,3 Ah,
It is determined as (133,3 Ah) / (0,87×75 Ah) = 2,04.
Accordingly, in cases where energy can't be generated from the main source, 2×(12 V/75 Ah) serial batteries must be used in order for the batteries to feed the load independently for 1 hour. In this case, the string would consist of 2 batteries, and the number of batteries would increase by two while scanning the most appropriate battery number in both CFA and PSO algorithms.

2.3. Inverter Model

It was assumed that the devices used in 4 houses, which were determined as load, operate with 220 V and 50 Hz alternative current in this study. For this reason, direct current generated from PV and batteries must be inverted into alternative current. For this, an inverter was needed. The power value to be taken from the inverter output was obtained by multiplying the power coming to the input of the inverter with the efficiency of the inverter. This situation was shown in equation 1.

In this study, inverter loss was accepted as 8%.

$$\text{Inverter Capacity} = \left(\frac{\text{hourly energy requirement} \times \text{inverter loss compensation}}{\text{Base time (1 hour)}}\right) \quad (1)$$

Accordingly, the inverter capacity that should be in the system was:

$$\text{Inverter Capacity} = \left(\frac{1600\ \text{W} \times 1.08}{(1\ \text{hour})}\right) = 1728\ \text{W}$$

2.4. Load Model

The load is also called the demanded energy. It is waived in kW. The load profile helps to see how the demanded energy is used over time [25]. The data set was collected by Precon [26] with smart meters. It contains data from users of different demographic features and different social and financial backgrounds. The properties of four different houses used in this study are as in Table 3.

Table 3. The properties of four different houses [26]

House	Building Year	Ceiling Insulation	Children	Senior	Refrigerator	Washing Machine	Electric Heater	Iron
1	1992	Yes	3	2	2	1	2	1
2	1985	No	0	1	2	1	1	1
3	2010	No	0	0	2	1	0	1
4	1990	No	1	0	1	1	0	1

Figure 1 shows 720 hours consumption data for four houses.

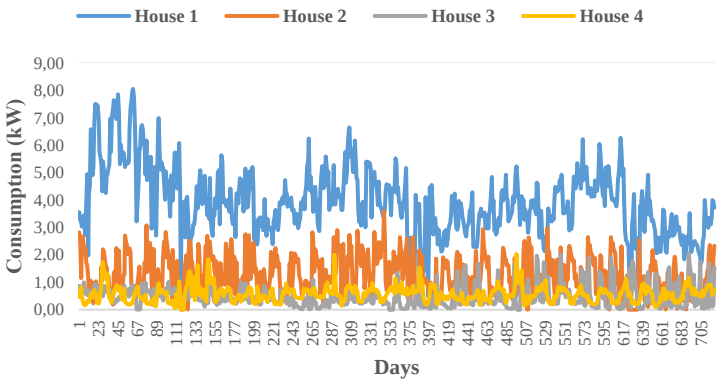


Figure 1. 720 hours consumption data of four houses

3. Method

The aim of this study was to ensure that part of the total electricity need in a site consisting of four houses was met by the PV-BT system with the minimum cost within the framework of fair usage. The standard deviation method was used for this. The standard deviation value was obtained by finding the minimum and maximum values from the combinations of the 720×1 matrix, which was the result of multiplying the values of 720-hour electricity consumption of four houses by four different percentages. Table 4 shows the unknowns in the study and the limitations of these unknowns.

Table 4. Unknowns and restrictions

Unknown values	H_1	H_2	H_3	H_4
Restrictions	$0 <= H_n <= 100$,			
	$H_1+H_2+H_3+H_4=100\%$			

There were 720 hours of data for four houses separately. These data form a 720×4 matrix consisting of 720 rows and 4 columns. H_1, H_2, H_3 and H_4 were the parameters that determine the percentage ratios of the electrical energy obtained from the PV-BT system used by the houses. These parameters form a 4×1 matrix consisting of 4 rows and 1 column. As a result of the multiplication of these two matrices, a 720×1 matrix consisting of 720 rows and 1 column was formed. This matrix gives total load amounts with minimum and maximum standard deviation. It was ensured that these load amounts were met fairly by the PV-BT system.

Standard deviation is defined as the standardised measure of the distance of each data from the mean in a data set. It is a statistic calculated based on the average. Its value is desired to be the smallest. It is calculated as in equation 2 [27].

$$A = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}} \quad (2)$$

Flowchart of the study is as in Figure 2.

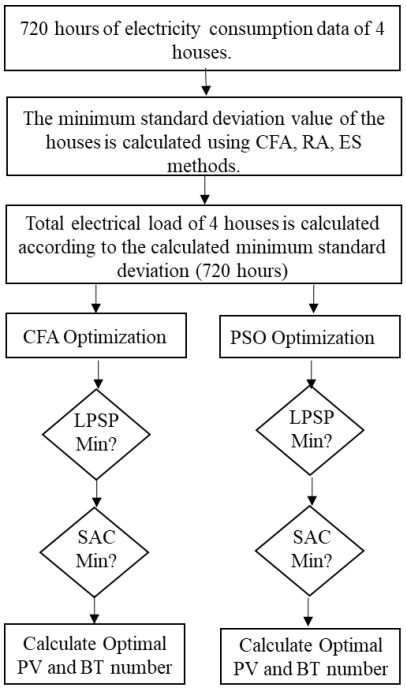


Figure 2. Flowchart of the study

3.1. Minimum Standard Deviation Calculation with CFA

In the conditional flow algorithm, there is no need for any extra information about the objective function. It is a method used in solving optimisation problems. Unlike traditional optimisation methods, it searches for a point lower than the value of the objective function for the moment in the set of points [28]. It is a method that guarantees to reach the target result. It works on the principle of scanning all values in the range that must be scanned to achieve the target result. The ability to scan only two parameters, such as solar panels and battery numbers and to guarantee to find the global minimum are the features that make this algorithm superior. At this stage, the result obtained with CFA is the actual minimum deviation value result. This is due to the fact that it was obtained as a result of scanning the entire range. While making standard deviation calculations with CFA, the variables H_1, H_2, H_3 and H_4 were scanned between 0 and 100 with 1 number increment. Scanning ranges were as follows:

H_1=0:1:100
H_2=0:1:100
H_3=0:1:100
H_4=0:1:100

In this direction, the result obtained with CFA was as in Table 5.

Table 5. Minimum standard deviation and usage percentages with CFA (Integer)

H_1 (%)	H_2 (%)	H_3 (%)	H_4 (%)	Min. Std. Dev.
1	9	32	58	0.219799

3.2. Minimum Standard Deviation Calculation with RA

The calculation was made by assigning random values to the relevant variables in the specified range in the random algorithm. The results were obtained by considering the restrictions in Table 4 and assigning random values to each variable. Scanning ranges for variables were done as follows:

H_1=0:1:100
H_2=0:1:100
H_3=0:1:100
H_4=0:1:100

The results were obtained by running the program 10 times. All results are shown in Table 6.

Table 6. Minimum standard deviation and usage percentages with RA (Integer)

Run	H_1 (%)	H_2 (%)	H_3 (%)	H_4 (%)	Min. Std. Dev.
1	1	20	25	54	0.234196
2	3	23	5	69	0.263518
3	9	3	40	48	0.245961
4	4	1	38	57	0.230273
5	8	11	34	47	0.237975
6	1	31	23	45	0.272449
7	5	8	48	39	0.238395
8	***2***	***14***	***26***	***58***	***0.223898***
9	6	1	23	70	0.239731
10	13	20	19	48	0.281758

The best result was obtained with the eighth run. The minimum standard deviation was achieved as 0.223898, while the percentages were 2%, 14%, 26% and 58%, respectively. When the results obtained with the conditional flow algorithm were compared with the results obtained by assigning a random value, it was seen that the random algorithm also obtained a value close to the minimum result.

3.3. Minimum Standard Deviation Calculation with Solver

The result obtained in the study with ES was the same the result obtained with CFA. Considering that CFA calculates the global minimum standard deviation by calculating among all combinations, ES's reaching the same result showed that it was also a successful method. The result obtained with ES is as in Table 7.

Table 7. Minimum standard deviation and usage percentages with ES (Integer)

H_1 (%)	H_2 (%)	H_3 (%)	H_4 (%)	Min. Std. Dev.
1	9	32	58	0.219799

In the studies up to this stage, the results were obtained by scanning with integer values. After this stage, the study was carried out according to the decimal status of the variables. Accordingly, all the minimum standard deviation results obtained with CFA, RA and ES are as in Table 8.

Table 8. Minimum standard deviation with all methods (Integer and Decimally)

Method	H_1	H_2	H_3	H_4	Minimum Standard Deviation
CFA (Integer)	1	9	32	58	0.219799
CFA (Decimally) (From 0 to 100 with 0.50 steps)	1.0000	9.0000	32.5000	57.5000	0.219788
ES (Integer)	1	9	32	58	0.219799
ES (Decimally)	***1.1061***	***9.1814***	***32.3474***	***57.3650***	***0.219779***
RA (Integer)	2	14	26	58	0.223898
RA (Decimally) (The lowest after 10 times of running)	0.6643	15.5541	16.7632	67.0172	0.233042

When Table 8 is examined, it is seen that the lowest value among the minimum standard deviation results in the use of decimals scenario was produced by ES. The minimum standard deviation value obtained as a result of this

scenario was 0.219779, and fair usage percentages for each house were 1.1061%, 9.1814%, 32.3474% and 57.3650%, respectively. In other words, if 1.1061% of the energy received from the PV-BT system were used by house 1, 9.1814% by house 2, 32.3474% by house 3 and 57.3650% by house 4, the energy to be obtained from the PV-BT system would be used fairly.

3.4. Maximum Standard Deviation Calculation

While calculating the minimum standard deviation, since the best result was obtained with decimal numbers in the ES application, only ES application and decimal numbers were studied in the maximum standard deviation calculation. Accordingly, the result obtained by ES was as in Table 9.

Table 9. Maximum standard deviation and usage percentages with ES (Decimally)

H_1 (%)	H_2 (%)	H_3 (%)	H_4 (%)	Max. Std. Dev.
100	0	0	0	1.240993

Considering these results, the graph of the total load values required to be fed by the PV-BT system at the minimum and maximum standard deviation values is as in Figure 3.

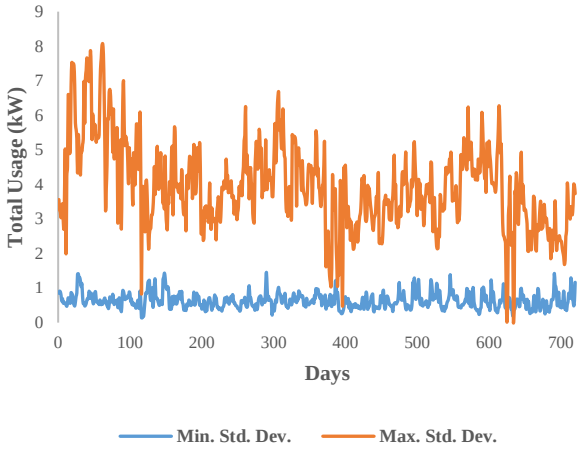


Figure 3. Total consumed load values according to the minimum and maximum standard deviation values

3.5. Technical Criteria

The technical criteria can also be called LPSP. The system that provides reliable electrical power means having enough power to supply the power demanded by the load in each period. In other words, it is the case where the LPSP value is zero. When the power system cannot meet the energy demanded by the load, the probability of the load being de-energized is calculated. If the value of the variable used in the calculation does not change, it means that the power demanded by the load is provided by the power system. An increase in its value means that the energy generated by the power system at that hour is not sufficient to feed the load. LPSP is a statistical parameter. The calculation of LPSP value is not only related to the solar radiation value. It is also related to the condition of the batteries at that time. For this reason, the system is largely powerless at times when these data are bad. In this study, LPSP, one of the target functions, is calculated as a percentage and determined according to equation 3.

$$\text{LPSP}\ (\%) = (A/720) \times 100 \tag{3}$$

Variable A indicated the total number of hours in which the energy supplied by the PV panel and battery string could not meet the demanded energy. The load remain powerless during these hours. The number 720 refers to the total hours from 00:00 on June 1, 2018 until 23:00 on June 30, 2018.

The power required by the load is calculated as in equation 4:

$$P_{iht}(t) = P_{ACload}(t)/\eta_{inv}(t) \tag{4}$$

For the specified hour, if the energy generated by the PV panels was more than the load, the excess energy was stored in the batteries and the new state of the battery was determined according to equation 3. The remaining energy was not used. The LPSP value remains at its final value. (LPSP=LPSP).

If the energy demanded by the load were more than the energy generated by the PV panels, the batteries would be activated to meet the energy demanded by the load. The batteries would be discharged according to equation 3 and would feed the load. If the total amount of energy in the batteries were enough to feed the load, the LPSP value would remain in its final state. If the energy in the batteries were insufficient, the load would be weak, and the LPSP value would be increased by 1. (LPSP=LPSP+1).

3.5. Economic Criteria

It is the economic criterion used in optimisation in this study. The aim was to minimise the cost based on the SAC concept. CFA and PSO algorithms were used to minimize the SAC. The total cost of the system consisted of the initial main costs, replacement costs and maintenance costs. While calculating these costs, PV, BT, inverter (INV), charge control unit (CCU), panel and installation costs were taken into consideration. Accordingly, the total cost is as in equation 5.

$$\text{Total Cost (for 25 years)} = \text{Initial Main Costs} + \text{Replacement Costs} + \text{Maintenance Costs} \quad (5)$$

PV panel, battery group, inverter, charge control unit, panel and installation costs were taken into account in the initial cost of each component. This is as in equation 6.

$$\text{Initial Main Costs} = (\text{PVinitialcost}\times\text{PV})+(\text{BTinitialcost}\times\text{BT})+(\text{INVinitialcost}\times 1)+(\text{CCUinitialcost}\times 1)+(\text{PANELinitialcost}\times 1)+\text{Installation} \quad (6)$$

Replacement costs are calculated by considering equation 7.

$$\text{Replacement Costs} = (\text{BTinitialcost} \times \text{BT}\times \text{number of replacement}) + (\text{INVinitialcost} \times 1\times \text{number of replacement}) \quad (7)$$

Considering that the life span of PV is 25 years, the number of changes was taken as 5 for BT (since its life span is 4). For INV (since its life span is 10), the number of changes was taken as 1.5. Considering the life span of the components, only the battery and the inverter need to be changed periodically during the life of the system.

Maintenance costs (25 years) are calculated by considering equation 18.

$$\text{Maintenance Costs (25 years)} = (\text{PVannualmaintenance}\times\text{PV}\times \text{number of maintenance})+(\text{BTannualmaintenance}\times\text{BT}\times \text{number of maintenance}) \quad (8)$$

Since the life span of PV is 25, the number of maintenance for PV is 24. It will complete its life without maintenance in the last year. Since the life span of BT was 4 years, 5 times change was required for BT in 25 years. There will be no maintenance during these replacement years. For this reason, there will only be BT maintenance in the intervening 3 years. Accordingly, the string of the battery feeding the load will change 6 times. In this case, the number of maintenance for BT will be 18 (6×3).

Unit price, full price, maintenance cost and lifespan of PV panel, battery and other devices in this study are shown in Table 10.

Table 10. All cost of the system component [23-24]

Component	Unit Price ($)	Price ($)	Maintenance Cost ($)	Lifespan (year)
PV (200 W)	0.85	85	0.85 (1%)	25
BT	-	115	1.15 (1%)	4
Inverter	-	700	-	10
Charge Control Unit	-	1000	-	-
Panel (Electric Panel)	-	300	-	-
Installation Cost (cable etc.)	-	1500	-	-

Accordingly, the final version of the above cost equations was as follows:

Initial Main Costs ($) = (85×PV) + (115×BT) + (700×1) + (1000×1) + (300×1) + 1500

Replacement Costs ($) = (115×BT×5) + (700×1×1.5)

Maintenance Costs ($) = (0.85×PV×24) + (1.15×BT×18)

Accordingly SAC value was:

SAC = (Initial Main Costs + Replacement Costs + Maintenance Costs) / (25 year);

SAC = ([(85×PV) + (115×BT) + (700×1) + (1000×1) + (300×1) + 1500] + [(115×BT×5) + (700×1×1.5)] + [(0.85×PV×24) + (1.15×BT×18)])/25;

SAC= (105.4×PV + 710.7×BT + 4550)/25

For determining the number of PV and BT that could feed load value at an optimum level, minimum LPSP and min SAC values will be taken into account. Two different algorithms were used for optimisation. These were CFA and PSO. CFA was described in section 3.1. PSO is a meta-heuristic algorithm often used in discrete, continuous and combinatorial optimisation problems. It was developed in 2001 by Kennedy and Eberhart [29]. It imitates the social behaviour of food-seeking bird flocks. The only solution is called a particle. The sum of all solutions is called a herd. The algorithm is started with randomly generated particles looking for the most suitable solution. Every particle has a position and velocity. The velocity of each particle is updated according to the following equation 9 [30-31]:

$$v_{i,j}(t+1) = w \times v_{i,j}(t) + c_p \times r_p \times \left(pBest_{i,j} - x_{i,j}(t)\right) + c_g \times r_g \times \left(gBest_j - x_{i,j}(t)\right) \quad (9)$$

v_{ij} is the velocity of i th particle in the j th dimension. x is the current particle position, and w is the constant called momentum, which sets from the previous time step how much the velocity will affect the velocity in the current time step. c_p and c_g are predefined constants. r_p and r_g are random numbers in the range (0, 1). The position of particle i in the j_{th} dimension is updated as in equation 10:

$$x_{i,j}(t+1) = x_{i,j}(t) + v_{i,j}(t+1) \quad (10)$$

Excel solver (ES) is an excel application that can be used for simulation analysis. The aim **is** to find the optimum value for a formula in the cell. ES works with a group of cells used in calculation formulas in objective and constraint cells. Values in the decision variable cells are set to accommodate the constraints in the constraint cells and create the desired value for the objective cell [32].

4. Results and Discussions

4.1. Minimum SAC with Minimum Standard Deviation

While scanning with CFA to calculate the minimum cost in the minimum standard deviation, the number of PV was obtained with 1 increment in the range of 0-200. The number of BT was obtained by scanning with two increments between 2-50. Considering these range, a total of 306 LPSP values out of 500 (200×25) probabilities were calculated as 0%, and among these values, the value of 1850.576 $ in the 97th rank was calculated as the minimum SAC value. This situation is seen in Figure 4 (a). According to the cost of 1850.576 $, the number of PV and BT was calculated as 180 and 32, respectively. In order to calculate the minimum cost with PSO according to the minimum standard deviation, the number of PV was scanned in the range of 0-200, and the number of BT was scanned in the range of 0-100. The SAC value was obtained as 1759.992 $. This value could only be obtained at a minimum value of 0.83% LPSP. As can be seen in Figure 4 (b), the minimum cost had calculated by PSO from the 63rd iteration out of 100 iterations in total. According to the cost of 1759.992 $, the number of PV and BT was calculated as 172 and 30, respectively.

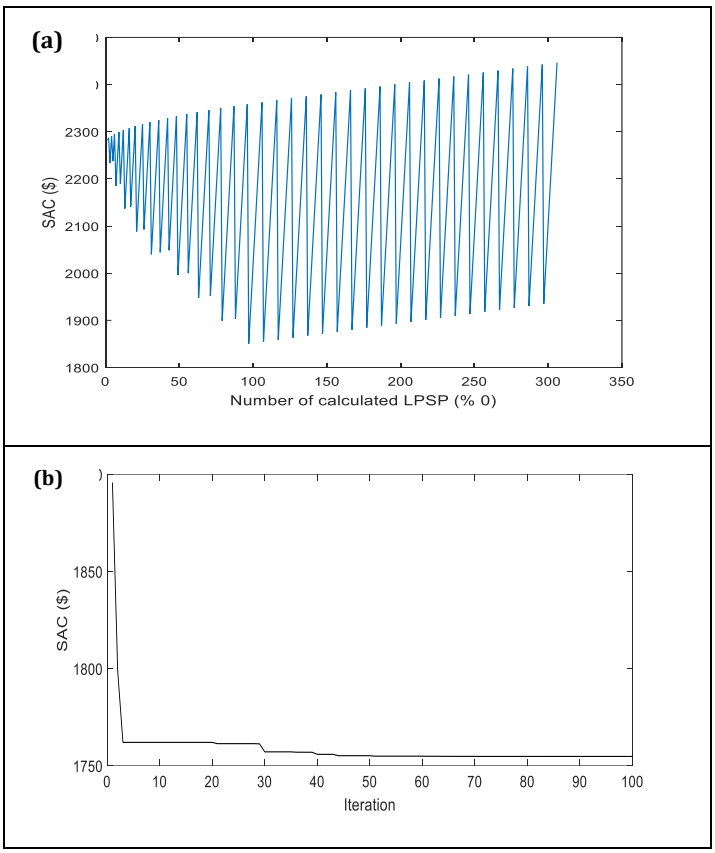
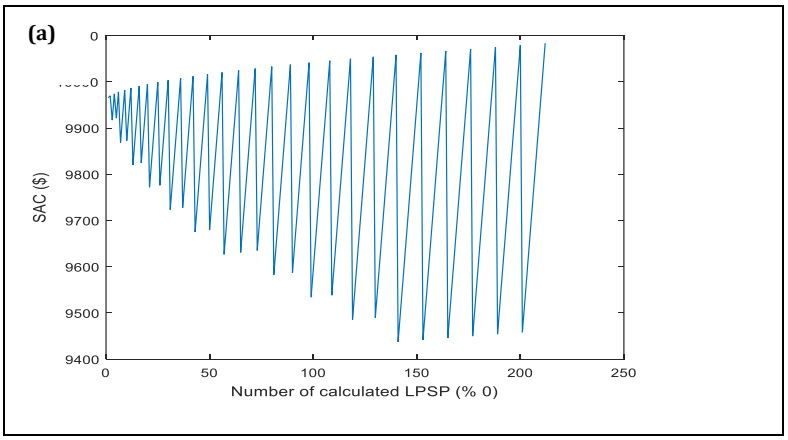


Figure 4. (a) Graph of calculating the minimum SAC value with CFA
(b) Graph of calculating the minimum SAC value with PSO

4.2. Minimum SAC with Maximum Standard Deviation

In order to calculate the minimum cost according to the maximum standard deviation with CFA, the number of PV was obtained by scanning with one increment in the range of 0-1000, and the number of BT was obtained by scanning with two increments in the range of 2-200. Considering these ranges, a total of 212 LPSP values among 1000 × 100 = 100.000 probabilities were calculated as 0%, and among these values, the value of 9437.104 $ in the 141st row was calculated as the minimum SAC value. This situation is seen in Figure 5 (a). According to the cost of 9437.104 $, the number of PV and BT was calculated as 995 and 178, respectively. With PSO, according to the maximum standard deviation, the number of PV was scanned between 0-1000 and the number of BT between 0-200, and the minimum cost was obtained as 8918.896 $. This value could be obtained at 0.83% LPSP value. As can be seen in Figure 5 (b), the minimum cost calculation had been made by PSO from the 59th iteration out of 100 iterations in total. According to the cost of 8918.896 $, the number of PV and BT was calculated as 953 and 166, respectively.



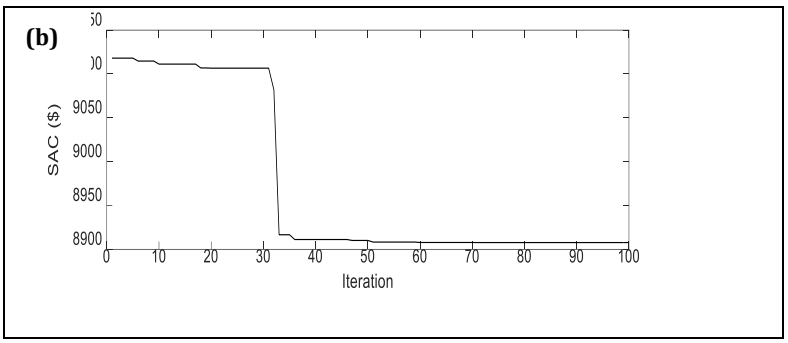


Figure 5. (a) Graph of calculating the maximum SAC value with CFA
(b) Graph of calculating the maximum SAC value with PSO

The results obtained for both cases (case 1 and case 2) are as in Table 11.

When the studies in the literature were reviewed, it wasn't possible to see studies regarding the fair usage of energy resources by loads. In this study, unlike the studies in the literature, it was aimed to ensure fair usage of the power obtained from the renewable energy source among the loads. The standard deviation method was used for this. CFA, RA and ES methods were used to calculate the standard deviation. Ghaffari et al. [13], Bukar et al. [16] and Astaneh et al. [18] achieved their goals by using different algorithms for LPSP and minimum cost calculations in their studies. In this study, after calculating the minimum standard deviation, LPSP and SAC research was done with CFA and PSO for optimum sizing of the PV-BT system.

Table 11. Results of case 1 and 2 (LPSP= % 0 – 1)

Case	Standard Deviation Method	H_1 (%)	H_2 (%)	H_3 (%)	H_4 (%)	Standard Deviation	Optimization Method	PV (200W)	BT (12 V 75 Ah)	Min. SAC ($)	LPSP (%)	Load (kW) (720 hours)
1	Excel Solver (Decimally)	1.1061	9.1814	32.3474	57.3650	0.219779 (min)	CFA	180	32	1850.576	0.00	456.58
							PSO	172	30	1759.992	0.83	
2		100.00	0.00	0.00	0.00	1.240993 (max)	CFA	995	178	9437.104	0.00	2836.23
							PSO	953	166	8918.896	0.83	

This study, unlike the studies in the literature, it was aimed to ensure fair usage of the power obtained from the renewable energy source among the loads. Standard deviation method was used for this. CFA, RA and ES methods were used to calculate the standard deviation. Ghaffari et al. [13], Bukar et al. [16] and Astaneh et al. [18] achieved their goals by using different algorithms for LPSP and minimum cost calculations in their studies. In this study, after calculating the minimum standard deviation, LPSP and SAC research was done with CFA and PSO for optimum sizing of the PV-BT system.

5. Conclusions

In this study, in order to benefit from a renewable energy system consisting of PV and BT both fairly and economically, an optimum design was made according to the electrical energy data of a site consisting of 4 houses. Seven different standard deviations were calculated. Among these, the minimum standard deviation value obtained in the study of ES with decimal numbers had the lowest value. This value was calculated as 0.219779. The fair usage percentages of the energy to be obtained from the renewable energy source on the basis of houses were calculated as 1.1061%, 9.1814%, 32.3474% and 57.3650%, respectively. The maximum value of the standard deviation was 1.240993, with fair usage percentages being 100%, 0%, 0% and 0%. Optimisation studies were carried out considering the partial consumption amount obtained according to these standard deviation values. Results were obtained with both CFA and PSO to calculate the number of PV and BT where both criteria were minimum. Considering the minimum standard deviation value, in the optimization study with CFA, 306 combinations with an LPSP value of 0% among 500 probabilities were calculated. Among these, the minimum SAC value was calculated as 1850.576 $. The numbers of PV and BT calculated for this cost were 180 and 32, respectively. Considering the minimum standard deviation value, the LPSP value was calculated as 0.83% in the study with PSO. The minimum SAC value was calculated as 1759.992 $. The numbers of PV and BT calculated for this cost were 172 and 30, respectively. The same studies were also carried out on the basis of the maximum standard deviation. According to the optimisation with CFA, the minimum SAC value

was calculated as 9437,104 $. The numbers of PV and BT corresponding to this SAC value were calculated as 995 and 178, respectively. In case study, the standard deviation was maximum, and the minimum SAC value was obtained as 8918.896 $ in the optimisation with PSO, while this value was obtained at 0.83 % LPSP value. The calculated numbers of PV and BT were 953 and 166. Considering the results obtained, although the CFA algorithm obtained the global minimum for the minimum LPSP value, a minimum LPSP value above this value was obtained by the PSO algorithm. The algorithm's inability to catch the global minimum can be seen as a shortcoming. However, the difference of only 0.83% ensures that both the number of PV and BT are lower and, accordingly, the total cost is lower. As a consequence, it is thought that this study will be a guide in ensuring fair usage and sizing optimisation of renewable energy in places where there is collective energy usage.

Conflict of Interest

Author has declared no conflict of interest.

References

[1] Zhang, W., Maleki, A., Rosen, M.A. and Liu, J. (2019). Sizing a stand-alone solar-wind-hydrogen energy system using weather forecasting and a hybrid search optimization algorithm. *Energy Conversion and Management,* 180, 609-621.

[2] Chauhan, A. and Saini, R.P. (2017). Size optimization and demand response of a stand-alone integrated renewable energy system. *Energy,* 124, 59-73.

[3] Sobek, W. (2018). Building as renewable power plants: Active houses for the electric City, Urban Energy Transition (Second Edition). *Renewable Strategies for Cities and Regions,* 131-138.

[4] Mohamed, M.A., Eltamaly, A.M. and Alolah,A.I. (2017). Swarm intelligence-based optimization of grid-dependent hybrid renewable energy systems. *Renewable and Sustainable Energy Reviews,* 77, 515-524.

[5] Al-falahi, M.D.A., Jayasinghe, S.D.G. and Enshaei, H. (2017). A review on recent size optimization methodologies for standalone solar and wind hybrid renewable energy system. *Energy Conversion and Management,* 143, 252-274.

[6] Zhang, W., Maleki, A., Rosen, M.A. and Liu, J. (2018). Optimization with a simulated annealing algorithm of a hybrid system for renewable energy including battery and hydrogen storage. *Energy,* 163, 191-207.

[7] Al-Ghussain, L., Samu, R., Taylan, O. and Fahrioglu, M. (2020). Sizing renewable energy systems with energy storage systems in microgrids for maximum cost-efficient utilization of renewable energy resources. *Sustainable Cities and Society,* 55, 102059.

[8] Mehrabankhomartash, M., Rayati, M., Sheikhi, A. and Ranjbar, A.M. (2017). Practical battery size optimization of a PV system by considering individual customer damage function. *Renewable and Sustainable Energy Reviews,* 67, 36-50.

[9] Makhdoomi, S. and Askarzadeh, A. (2020). Optimizing operation of a photovoltaic/diesel generator hybrid energy system with pumped hydro storage by a modified crow search algorithm. *Journal of Energy Storage,* 27, 101040.

[10] Mahmoud, T.S., Ahmed, B.S. and Hassan, M.Y. (2019). The role of intelligent generation control algorithms in optimizing battery energy storage systems size in microgrids: A case study from Western Australia. *Energy Conversion and Management,* 196, 1335-1352.

[11] Kerdphol, T., Fuji, K., Mitani, Y., Watanabe, M. and Qudaih, Y. (2016). Optimization of a battery energy storage system using particle swarm optimization for stand-alone microgrids. *International Journal of Electrical Power&Energy Systems,* 81, October, 32-39.

[12] Baran, B. (2012). *Cost optimization of solar - wind hybrid systems,* Inonu University Electrical and Electronic Engineering Department. Master of Science Thesis.

[13] Ghaffari, A. (2020). Design optimization of a hybrid system subject to reliability level and renewable energy penetration. *Energy,* 193, 116754.

[14] Hemeida, A.M., El-Ahmar, M.H., El-Sayed, A.M., Hasanien, H.M., Alkhalaf, S., Esmail, M.F.C. and Senjyu, T. (2020). Optimum design of hybrid wind/PV energy system for remote area. *Ain Shams Engineering Journal,* 11:1, 11-23.

[15] Kaabeche, A. and Bakelli, Y. (2019). Renewable hybrid system size optimization considering various electrochemical energy storage Technologies. *Energy Conversion and Management,* 193, 162-175.

[16] Bukar, A.L., Tan, C.W. and Lau, K.Y. (2019). Optimal sizing of an autonomous photovoltaic/wind/battery/diesel generator microgrid using grasshopper optimization algorithm. *Solar Energy,* 188, 685-696.

[17] Jamshidi, M. and Askarzadeh, A. (2019). Techno-economic analysis and size optimization of an off-grid hybrid photovoltaic, fuel cell and diesel generator system. *Sustainable Cities and Society,* 44, 310-320.

[18] Astaneh, M., Roshandel, R., Dufo-López, R. and Bernal-Agustín, J.L. (2018). A novel framework for optimization of size and control strategy of lithium-ion battery based off-grid renewable energy systems. *Energy Conversion and Management,* 175, 99-111.

[19] Fetanat, A. and Khorasaninejad, E. (2015). Size optimization for hybrid photovoltaic–wind energy system using ant colony optimization for continuous domains based integer programming. *Applied Soft Computing,* 31, 196-209.

[20] Bukar, A.L. and Tan, C.W. (2019). A review on stand-alone photovoltaic-wind energy system with fuel cell: System optimization and energy management strategy. *Journal of Cleaner Production.,* 221, 73-88.

[21] Lian, Y (2020). Energy optimization potential for interconnected buildings in a new urban development project. Master of Science Thesis TRITA-ITM-EX 2020:4, *KTH Industrial Engineering and Management,* Sustainable Energy Engineering.

[22] Turkish State. 2020. Meteorological Service, Solar radiation data of Malatya. www.mgm.gov.tr, (Accessed Date: 30 June 2020).

[23] PV Panel. 2020. https://www.terasolarsatis.com/200w-polikristal-fotovoltaik-panel (Accessed Date: 07 July 2020).

[24] Battery. 2020. Solar Systems, 100Ah Full Energy accumulator. https://www.mutlu.com.tr/full-enerji-akuler (Accessed Date: 07 July 2020).

[25] Garson, S. What is a Load Profile and why is it Important?. Retrieved from: http://bettercostcontrol.com/what-is-a-load-profile-andwhy-is-it-important, (Accessed Date: 27 July 2020).

[26] Precon. 2020. Pakistan Residential Electricity Consumption Dataset. https://opendata.com.pk/dataset/precon-pakistan-residential-electricity-consumption-dataset, (Accessed Date: 30 March 2020).

[27] Standard Deviation. 2020. https://acikders.ankara.edu.tr/pluginfile.php/1379/mod_resource/content/2/B6_Merkezden%20Da%C4%9F%C4%B1lma%20%C3%96l%C3%A7%C3%BCleri.pdf, (Accessed Date: 16 May 2020).

[28] Yang, H., Zhou, W., Lu, L. and Fang, Z. (2008). Optimal sizing method for stand-alone hybrid solar–wind system with LPSP technology by using genetic algorithm. *Solar Energy,* 82, 354–367.

[29] Kennedy, J. and Eberhart, R.C. (2001). Swarm Intelligence. Academic Press, San Diego, CA.

[30] Junior, F.E.F. and Yen, G.G. (2019). Particle Swarm Optimization of Deep Neural Networks Architectures for Image Classification. *Swarm and Evolutionary Computation,* 49, 62-74.

[31] Shamshirband, S., Mosavi, A. and Rabczuk, T. (2020). Particle swarm optimization model to predict scour depth around a bridge pier. *Frontiers of Structural and Civil Engineering,* 1-12.

[32] Excel Solver. 2020. https://support.microsoft.com/tr-tr/office/%C3%A7%C3%B6z%C3%BCc%C3%BC-y%C3%BC-kullanarak-bir-sorunu-tan%C4%B1mlama-ve-%C3%A7%C3%B6zme-5d1a388f-079d-43ac-a7eb-f63e45925040, (Accessed Date: 22 May 2020).

Journal of Investigations on Engineering & Technology
Volume 5, Issue 1, 47-62, 2022

ISSN: 2687-3052

Research/Review Article

Investigation of the eczema and skin cancer disease diagnosis by using image processing techniques

Yusuf Süer Erdem [1], Özhan Özkan [2]

(Alınış / Received: 23.05.2022, Kabul / Accepted: 30.06.2022, Online Yayınlanma / Published Online: 30.06.2022)

Keywords
Eczema
Matlab
Image processing
Skin Diseases
Skin Cancer

Abstract: It is seen that many diseases, especially dermatological diseases, arise due to bad weather conditions such as high temperature, dust, smoke, and sun in the environment. The most common diseases are eczema caused by malnutrition, soil, bacteria, bad food, and other factors, and risky moles, which are usually caused by excessive sunlight or during childbirth. Due to all these environmental, physiological, and chemical factors, it is important to quickly detect all existing skin diseases, especially eczema and risky moles, and it has become inevitable to establish a less costly diagnostic system with the help of doctors to prevent the aggravation of the diseases. If eczema and risky skin problems progress, they will be difficult to treat and take a long time. For this reason, the research aims to take an image from the infection site and then process this image in many ways in a MATLAB environment to obtain an output that can help doctors in their work. Decision trees supported the currently used image processing and classification method, and the results of both methods are also compared with this method. According to the results obtained, the accuracy, sensitivity, and selectivity ratios of decision trees are obtained compared to image processing. The software used gives a warning when the image processing and decision tree methods give conflicting results. If this occurs, it is necessary to stick to the doctor's data. The system in this study aims to improve human life and make it safe by recognizing eczema and risky moles. In this study, samples were selected from various layers of the body. In addition, a different interpretation can be made in the normal situation. When this approach technique is applied, more appropriate results have emerged in the process of detecting eczema and risky moles. In addition, normal skin is also involved in the process. Being able to define the normal state has been a contribution to science and it is aimed in this study to facilitate the work of medical personnel.

[1] Author 1 Yusuf Süer Erdem (Department of Electrical and Electronics Engineering, Sakarya University, 54050 Sakarya, Turkey)
[2] Author 2 Özhan Özkan (Department of Electrical and Electronics Engineering, Sakarya University, 54050 Sakarya, Turkey)

Corresponding Author and e–mail address: Yusuf Süer Erdem suer.erdem@ogr.sakarya.edu.tr

1. Introduction

When skin diseases are examined, it is seen that there are many causes. Examples of physical skin diseases are wounds in the human body, burns, and cracks caused by cold. In addition, there are also skin diseases caused by parasites. In addition, another type of skin disease is seen as allergic skin disease [1].

Dermatological diseases are frequently seen in the community as diseases with low mortality. These diseases are thought to be important diseases due to their high morbidity, but it is seen that the number of patients with skin diseases who apply for medical help is low. As the years progressed, it has been determined that there has been an increase in the applications of patients with dermatological complaints to health institutions, of which society has become conscious. Since dermatological diseases are long-term diseases and these diseases are costly diseases, their importance is increasing [2].

When the studies on the subject are examined, it has been determined that the use of artificial intelligence and data mining applications is used effectively in the field of health. At this point, in many areas; It is seen that alf activity takes place in many fields such as medical disease diagnosis, follow-up, cost estimation, imaging analysis, resource planning, emergency management, processing of unstructured data [3, 4, 5].

Artificial intelligence models, which are used as a result of bringing patient data in high dimensions to the functional point, play an active role in increasing data reliability and quality [6, 7]. Regarding the use of artificial intelligence in the field of health, legal and ethical processes are used effectively regarding the accuracy of clinical data, data management, and data protection [8]. The processes regarding this issue differ from country to country [9].

The use of artificial intelligence is especially used in the detection of diabetes. Different applications are used concerning artificial intelligence in the detection of diabetes. In the study of Mujumdar and Vaidedi in 2019, large data sources in the health management system are used to increase the accuracy rate in the detection of diabetes [10].

A decision tree is defined as a model used in data mining to show classifiers and regressions. The decision tree consists of some nodes and branches. In the classification carried out in the decision tree, leaves and classes are specified. It has been determined that the decision is made according to one or more attributes in another node. A decision tree is used extensively in data mining because it is simple and understandable. To interpret the output in the decision tree, the entire content is explained one by one, independently of an expert. A graphical method is applied in the decision tree. Since this method is used in the decision tree, it is seen that it is simpler than other classification methods. Due to a large number of nodes in the decision tree, it can be difficult to display the graph [11].

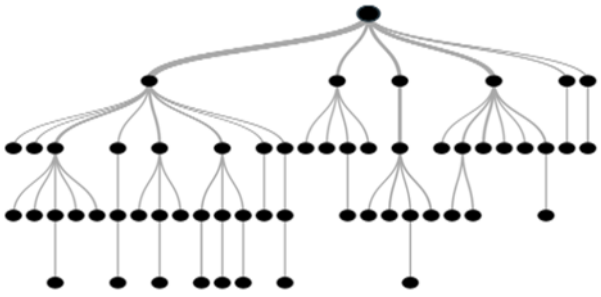


Figure 1. Example of Decision Tree

When the decision tree model was examined, it was determined that the model consisted of several parts. These parts:

- Creating node definitions to assign each record of the data set in the decision tree to a leaf node,
- Rules for assigning each record of the dataset to a leaf node,
- Revealing the posterior probabilities of the nodes of each leaf,
- A target level is assigned to each leaf node in the decision tree.

In this part, it was determined that node definitions were developed using the training data set. And it is specified in terms of input ranges. Using the training dataset, posterior probabilities are calculated for each node. Finally,

the target level found is assigned to each node and then the process is carried out using the training datasets within the training phase [12].

Many methods are used in the image analysis process. These methods are morphologically processing the image, extracting the image features, realizing the separation of the image into layers, identifying the patterns, and edge detection methods.

When morphological image processing is examined, it has been determined that it is a type of image processing in which the structure of objects in an image is changed. It is seen that three basic morphological processes, expansion, erosion, and skeletonization method are used in this species.

In enlargement, it is ensured that an object is reduced uniformly by erosion as a result of uniform enlargement within the spatial scale. In the skeleton method, especially an object is represented in the form of sticks [13].

It has been determined that there are activities to extract information such as high-level color and shape from images to obtain the characteristics of an image. For this reason, when the applied processes are examined,shadow detection, edge detection, contrast detection, motion detection, slope detection, and determination of the sequence of movements are made to examine the image properties caused by the density difference. In the operations applied in this framework, high- and low-level feature extraction operations are carried out. The basis of these processes is based on the principle of shadow, edge, contrast determination, detection of motion, and slope and determination of motion sequences depending on the image properties caused by the density difference [14].

It has been determined that especially machine learning increases the accuracy rate in the determination of diseases. When the studies on this subject are examined, a study has been made on the determination of skin diseases by using supervised machine learning algorithms because of the use of the data set on the patients, especially in the Indian sample [15].

When the health management processes in which machine learning methods are used are examined, a study related to the management of diabetes has been made. As a result of this examination, the prediction and determination of the disease, the complications, the genetic history of the patients, management, and treatment processes were examined. In this study, it is seen that 85% of classification and supervised algorithms are used in prediction regarding the machine learning process. In this study, the most preferred algorithm is the Support Vector Machine.

The reasons for the widespread use of machine learning techniques in the field of health sciences are listed as follows. These are:

- Estimation,
- Diagnosis and determination of post-disease complications,
- It is determined as providing quality service to patients by saving time and workload [16].

It is seen that a patient-oriented and artificial intelligence-based patient follow-up system is recommended for providing comfort for patients with diabetes and improving treatment processes within the health management processes [17].

In another study investigating the opportunities of machine learning in healthcare, a model classified according to the risk groups of patients using hospital data was proposed, and it was determined that there was an improvement in patient follow-up and related actions [18].

In another study, it is seen that an artificial intelligence model was created that identifies high-cost and risky patients to determine the necessary treatments and steps in the treatment of diabetes [19].

In the study of Mercaldo et al. in 2017, it was determined that the Hoeffding Tree algorithm was used to diagnose and classify diabetes [20].

Despite the availability of advancing technology, precisely diagnosing the symptoms of any skin disease is still a major challenge for physicians. Many people suffer from serious skin conditions that require them to go to hospitals and undergo expensive medical examinations that take days. In this study, a method that solves skin diseases based on image processing and data science is used. Codes prepared on MATLAB help the above problem to a certain extent, and can also identify two different skin diseases, eczema, and skin cancer. In addition to previous studies, it can also provide an interpretation for normal skin. The study detects a range of symptoms in

a matter of seconds and acts as a switch. It also makes the diagnosis more intuitive and realistic. This study aims to classify different diseases and normal conditions based on images given as input. The study uses the MATLAB software program. Images, Dermnet and DermWeb etc. It is collected from various publicly available databases such as First, sample images of the two skin diseases and the normal condition need to be processed. Second, the diseased parts of the images are identified, and the corresponding mathematical operations are performed. Based on these, the characteristics of the two types of skin diseases and normal conditions are extracted, and the associated feature tissue parameters of these diseases and pixels of the lesion areas are collected through image segmentation. Finally, eczema and skin cancer symptoms are defined using the multiple support vector machine (MultiSVM) method to increase the identification accuracy and provide an interface to the physicians, and the normal state is not neglected either. It was created with 25 of 72 images in the Multi SVM image folder. However, the system gives consistent results when any new image is introduced that does not form the database. Clinical data of images were also collected from public databases such as Dermnet and DermWeb. Adding the decision tree system to the study helps the other party in the accuracy analysis part. Decision trees use the clinical data of the images to indicate that if the image is sick, it is sick or if it is normal, it is normal. The results of image processing and decision trees are compared with each other and in case of conflict, the system warns the system administrator.

1.1.Description of Diases

Eczema is in the form of red, inflamed, and itchy patches that appear during the heat increase in the skin. Skin cancer, which occurs in the form of malignant moles, can suddenly appear on normal skin without any warning, or it can develop on a pre-existing mole. Source of vitally serious skin cancer; cells called melanocytes. Melanocytes produce the pigment melanin, which gives our skin its color and enables us to tan. Skin cancer occurs because of the abnormal and uncontrolled proliferation of melanocytes.

1.1.1.Eczema

The best definition for eczema is an inflammatory skin reaction to various exogenous and/or endogenous stimuli [21].

The main manifestations and symptoms of eczema in infants are itching, rash and extensors, and in older children, skin deformities may occur. The occurrence of symptoms depends on the type of chronic recurrent eczema and whether there is another atopic disease in the family, and dryness is another manifestation of this disease. Facial pallor, various allergic shines, dermatographism, conjunctivitis, cataract, and skin prick test positivity can be seen in these people.

Clinical findings consist of a severe itchy rash that presents as red dots, scaling, and papulovesiculation. This is a generalization, but eczema progresses step by step, depending on many of the previously mentioned characteristics such as the patient's disease history, age, immunity, and skin type.

1.1.2.Skin Cancer

It is necessary to know the location and appearance of moles in the body well to perceive the changes in the following processes. It is possible to confuse malignant moles with other atopic moles. Sometimes skin cancer may be mistaken for a normal mole, but there is a possibility that that mole has been skin cancer since its inception [22].

Other warning signs; peeling, leakage, bleeding or swelling, or blistering on the mole. In addition, there may be changes in feeling related to me such as itching, tenderness, and pain. However, the relevant physician (usually a dermatologist) should advise on the necessity of removing an abnormal mole or following it closely.

It is recommended to consult a doctor if the following features are observed in a mole. These features are:

- If one half of the mole is not similar to the other half (in color and/or shape),
- If the borders of the mole are irregular (indented),
- If the color of the mole is not homogeneous (two or more of the colors such as brown, black, red, gray, white is combined or have a mottled appearance),
- If the mole is larger than 6 mm in diameter (roughly larger than the diameter of an eraser pencil).

2.Methods

2.1.Methodology

The method to be used for this study is the iterative model, which is a part of the software development lifecycle. The iterative model is a method of breaking down the software development lifecycle of large applications into smaller segments [23].

In addition, the iterative method starts with a small application of the software requirements branch and includes stages that develop until the entire system is complete. In addition, the iterative approach gives the system the ability to make some changes and add new practical capabilities. The main idea of this approach is to develop the system over iterative loops in smaller chunks over a period of time.

"The key to successful use of an iterative software development lifecycle is rigorous validation of requirements and validation and testing of each version of the software with those requirements in each cycle of the model." [24].

As the main features of the iterative method; In the model, it can be shown that results are obtained periodically and early, can measure progress, develop in parallel, and do not have to change requirements or scope. On the negative side, the methodology has a few disadvantages such as having too many resources and their management. Figure 2 below shows an example of an iterative pattern diagram.

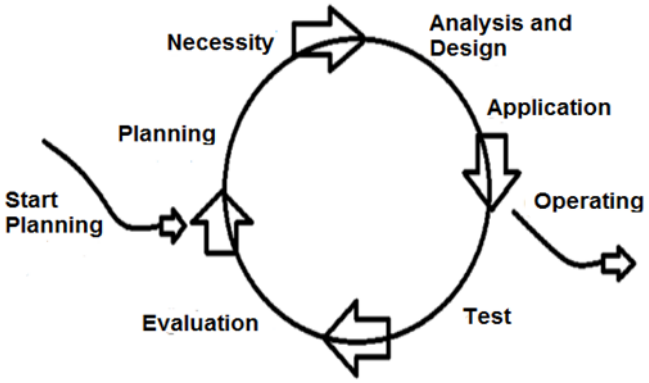


Figure 2. Iterative model diagram

The main stages of the iterative model in this study were determined as follows. These are:

- The stage of collecting and analyzing the requirements.
- Design phase: This phase analyzes the system requirements collected in the first phase and processes the system design by designing the block diagram, adjusting the MATLAB software to be used in this study to identify diseases.
- Implementation phase: In this phase, the work is implemented according to the current requirements and information, and while the system is being implemented, the work is divided into small programs called units to facilitate the implementation in MATLAB.
- Integration and testing phase.
- System deployment: After all functional and non-functional tests of the system are completed, the product is run with real data in the working environment.
- Care.

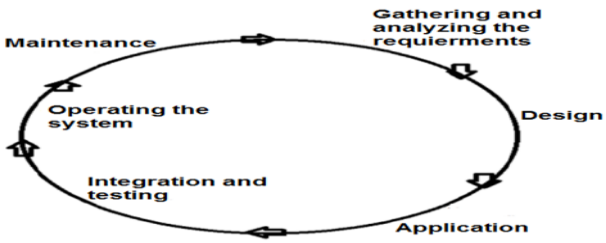


Figure 3. Iterative model stages of this study

Progress with iterative means that bug fixes are done continuously, so the result is more likely to be delivered on time and in higher-order as a result.

A function as an example: f(x)=x3-x-1
If we first write the equation x=ϕ(x)
a<b and f(a)·f(b)<0 the points a and b found.
If f(a) is closer than 0'af(b)
then x0=a, otherwise x0=b.
x1=ϕ(x0)
x2=ϕ(x1)
x3=ϕ(x2)
...
is repeated until |f(xi)-f(xi-1) |≈0

In the current algorithm, the newly added image is included in the function created by MultiSVM and the aim is to obtain results with minimum error with the iterative model. The system repeats itself every time a new image is uploaded to the system [25].

A comparative research method has been applied to this study, which helps the original image acquisition, feature extraction, and classification process based on MultiSVM. Previous studies discuss the use of colors and their effect in recognizing the surface of the human skin using the RGB YCrCb-HSV technique, distinguishing changes in the ratio of colors present on all four sides of the skin.

As the work progresses, there are clear differences in the feature extraction method part. Previous studies aimed to identify the skin area directly, which means that the primary goal was to know the location and surface of the skin. In recent studies, it is possible to follow the method of finding colors and extracting energy values from the image to detect diseases. The system in this study, which we have done, has the capacity to do this. What will be achieved in this work is to consider the extraction of multiple features, to improve the performance of the study, and to increase the capacity of the detection system to extract color features. Unlike other studies, the skin with eczema and skin cancer is also determined as normal skin. Unlike other studies, MultiSVM was used. In addition, unlike other works, MultiSVM was created with only lines of code without using MATLAB GUI. This system aims to increase the success of image classification every time the user uploads a new image.

2.2.Design and Analysis

Using the flowchart in Figure 4, the full functioning of the system can be seen.

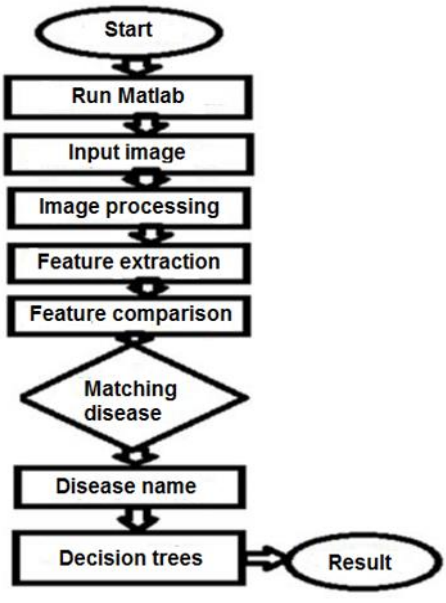


Figure 4. System flow chart

2.3.Operation and algorithm

The algorithm runs the system regularly with piecemeal progression logic. First, it is necessary to define RGB and grayscale images.

A digital color image is a type of image in which each pixel contains color information. Visually acceptable. For results, it is important to declare three color channels for each pixel, which in some color spaces are also interpreted as coordinates. In the computer environment, RGB color space is widely used. In a color image, each pixel has three values (or channels) that measure the chromaticity and intensity of the light. Luminance information in the spectral band is considered as real information stored in digital image data [26].

In grayscale, the value of each pixel is a single sample. A grayscale image is different from a single-bit two-tone black and white image. The grayscale digital image only gives density information. Also known as black and white, this image type consists of shades of gray ranging from black to white, with white at the strongest and black at the weakest. Grayscale images have many shades of gray, usually, grayscale images are produced by measuring the intensity of light in each pixel of a single band of the electromagnetic spectrum and exhibit monochromatic behavior only when a certain frequency is captured. Grayscale images can also be synthesized from a full-color image [26].

Converting RGB to grayscale; The main reason for this process is to process the brightness information of the image or to make an operation with the color information, for example, it can be examined whether an object is the same color in two different images or edge detection can be performed. To do this, it is necessary to separate the luminance information from the color information; this is mostly done using RGB to grayscale. Comparing colors in the grayscale color space is much easier than in RGB. For this reason, programmers prefer grayscale over RGB color space [27].

Feature extraction; means extracting features from images of processed lesion regions. The complexity of the classification problem in the processed images can be reduced by this method. The problem area can be distinguished from the background through certain features such as geometry and color. Feature extraction is very important to improve the accuracy of results in biomedical image processing, it is a necessary step to improve the quality of the processed image [28].

Image segmentation: It is defined as the process of dividing the image into several parts by converting the representative image into an easy form to make it more meaningful and easier to analyze [29].

The segmentation steps using K-means are as follows:

Step (1): Image reading.
Step (2): Converting the image from RGB color space to L*a*b* color space: The image used contains different colors. To distinguish these colors, it is necessary to convert the image from color space to L*a*b* space. All the color information of the space consisting of the L luminosity layer and a* b* chromaticity layer are in a* and b* layers.
Step (3): Classification of colors in a* b* space using the K-means clustering system. Clustering is a method of dividing the data into a set of clusters, this is done when the data is used without defined groups, the purpose of this algorithm is to find the groups in the data used, then the image will be divided into 3 clusters using the euclidean distance metric.
Step (4): An array is created to store the results of the clustering, the cluster containing the ROI to be converted to grayscale is selected, and then the GLCM is generated.
ROI: This name means "region of interest" containing a subset of samples selected from a dataset defined for a specific purpose [30]. The concept of ROI is widely used in many application areas. A region of interest (ROI) is a piece of the image on which filtering or other processing can be performed. More than one ROI can be identified in a single image [31].
The Gray Level Co-occurrence Matrix (GLCM) method is a method by which second-order statistical texture properties can be extracted. A GLCM is a matrix in which the number of gray levels in the image is equal to the number of rows and columns. A matrix element of two pixels with relative frequency (a, b | Δc, Δd) is separated from these pixels by a pixel distance (Δc, Δd), this matrix belonging to a certain neighborhood is formed at the density "a" and the other has density "b" [32].

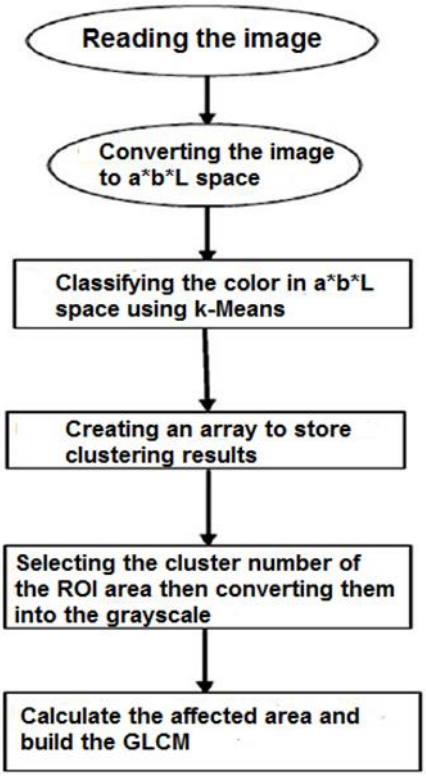


Figure 5. Block diagram of image segmentation

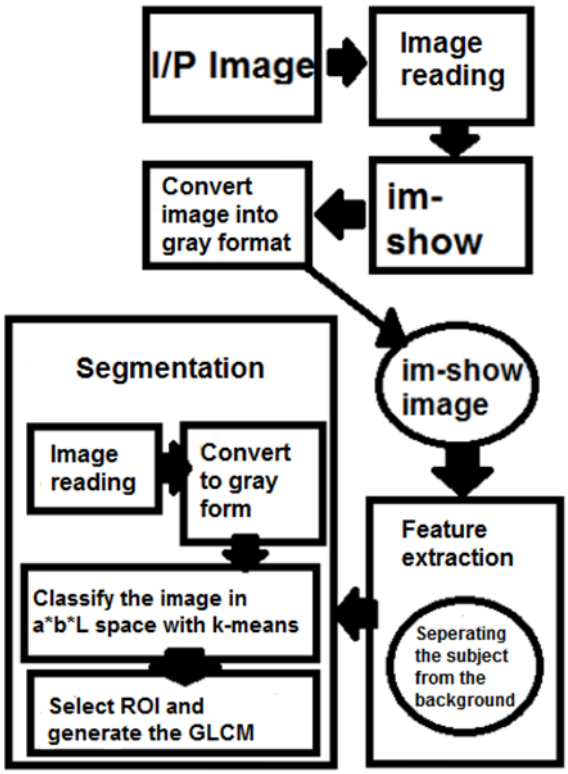


Figure 6. General block diagram of the proposed technique

2.4.MultiSVM and feature extraction

The mathematical background of the multiclass SVM (MCSVM) problem, typically with noise: (x1, y1), ..., (xN, yN), where xi: i = n length vectors from 1 to N and yi∈ {1, … represents the class of M} instances. The classical approach strategy for solving MCSVM classification problems is to think of the problem as a kind of combination of binary

classification problems. In the "One Support Vector Machine All" method, M classifiers are created, one support machine one for each class. The Mth classifier creates a hyperplane between the m class and the remaining M – 1 class. A new test sample is assigned to the class at which the distance from the positive boundary is maximum. By minimizing the optimization problem, it can be generalized to the following [33]:

$\Phi(w, \varepsilon)=\frac{1}{2}\sum_{m=1}^{M}({w_m}^T . w_m) + C \sum_{i=1}^{N}\sum_{m\neq y_i} {\varepsilon_i}^m$,
s.t. $({w_{yi}}^T . X_i)+b_{yi}\geq({w_m}^T . X_i)+b_m+2-{\varepsilon_i}^m$,
${\varepsilon_i}^m \geq 0$, for $i = 1, \ldots, N$: $m\in \{1, \ldots, M\} \setminus \{yi\}$.
The decision boundary can be written as:
$f(x)= \arg max_m[({w_m}^T . x)+b_m]$ for m= 1 ,..., M;

The optimization problem will be solved by finding Lagrange's saddle point. The final classifier is as follows:

$$F(x)= \sum a_i \, {x_i}^T + b_m$$

There are only N coefficients in the above equation (regardless of the number of classes, M), and the editing directly tries to reduce the nonzero number of α. In this study, the MultiSVM Struct structure is obtained by processing the 2 disease states and normal states with the obtained features without extracting the features. The extracted features show different properties from each other for the disease states and the normal state found in the study. Therefore, this is very important for classification. It is necessary to examine these attributes and how they are extracted.

When examining the properties of the gray level formation matrix, we obtain the values of contrast, correlation, energy, and homogeneity. These values show different properties for each image [29].

Contrast: Returns a measure of the intensity contrast between a pixel and its neighbor over the entire image.

$$\sum_{i,j} |i-j|^2 . p(i,j)$$

Correlation: Returns a measure of how related a pixel is to its neighbor over the entire image.

$$\sum_{i,j} \frac{(i-ui)(j-uj)p(i,j)}{\sigma_i \sigma_j}$$

Energy: Returns the sum of the squared elements in the gray level formation matrix;

$$\sum_{i,j} p(i,j)^2$$

Homogeneity: Returns a value that measures the closeness of the distribution of elements in GLCM to the GLCM diagonal.

$$\sum_{i,j} \frac{p(i,j)}{1+|i-j|}$$

Apart from these four features, there are different features obtained by using the RGB matrix form of the picture. For instance:

Kurtosis: A measure of how outlier-prone a distribution is.

$$k= \frac{E(x-u)^4}{a^4}$$

Here μ is the mean of x, σ is the standard deviation of x, and E(t) represents the expected value of quantity t.
Skewness: A measure of the asymmetry of the data around the sample mean.

$$s= \frac{E(x-u)^3}{a^3}$$

Here μ is the mean of x, σ is the standard deviation of x, and E(t) represents the expected value of quantity t.

Variance: Returns the variance of the A elements along with the first array size whose size is not equal to 1. For a random variable vector, A consisting of N scalar observations, the variance is defined as:

$$v=\frac{1}{N-1}\sum_{i=1}^{N}|A_i-u|^2$$

where μ is the mean of A,

$$u=\frac{1}{N}\sum_{i=1}^{N}A_i$$

Some definitions of variance use a normalization factor N instead of N, which can be specified by setting w to 1. Smoothness: The sum of all elements of the RGB matrix is set to a value and the formula 1-(1/(1+a)) is used. For all kinds of diseases and normal conditions, the operation where the distribution of color values in the RGB image is different will produce different results.

RMS (Root-mean level): Root-square-mean level of a vector x;

$$X_{rms}=\sqrt{\frac{1}{N}\sum_{n=1}^{N}|x_n|^2}$$

with aggregation carried out along the specified size.

Entropy: A statistical measure of randomness that can be used to characterize the texture of the input image.

Entropy is defined as -sum(p.*log2(p)); p contains the normalized histogram numbers returned from imhist.

Standard deviation: Returns the standard deviation of A elements along with the first array size whose size is not equal to 1.

$$s=\frac{1}{N-1}\sum_{i=1}^{N}|A_i-u|^2$$

where μ is the mean of A,

$$u=\frac{1}{N}\sum_{i=1}^{N}A_i$$

The standard deviation is the square root of the variance. Some definitions of standard deviation use the normalization factor N instead of N-1, which you can specify by setting w to 1.

Mean: For a random variable vector A consisting of N scalar observations, the mean is defined as:

$$u=\frac{1}{N}\sum_{i=1}^{N}A_i$$

IDM: Gets the measure of the distribution of GLCM elements to the GLCM diagonal.

2.5.Evaluation of the methods used

Image Processing and Decision Trees; The mean value of each feature in all images was calculated to examine the extracted features and to determine which disease they correspond to. In Table 1 below, the average value was found by moving through only 3 pictures of eczema. This shows the reader how MultiSVM is running in the background.

Table 1. Feature Inference Results and Averages of Three Skin Infected With Eczama

	A	B	C	D	E
1	Parameters				Average
2	Contrast	0,67704	0,43435	0,77974	0,63037
3	Correlation	0,87741	0,88214	0,9015	0,88701
4	Energy	0,57749	0,63881	0,2442	0,56794
5	Average	32	23	65	40
6	Standard Deviation	66	53	73	64
7	Entropy	3	2	5	3,33333

8	RMS	7	6	11	8
9	Homogenity	0,95079	0,96476	0,90031	0,93
10	Variance	3,81E+18	2,20E+18	4,71E+18	3,58E+18
11	Smoothness	1	1	1	1
12	Kurtosis	2	2	0,60633	1,33333
13	Distortion	6	7	2	5
14	IDM	255	255	255	255

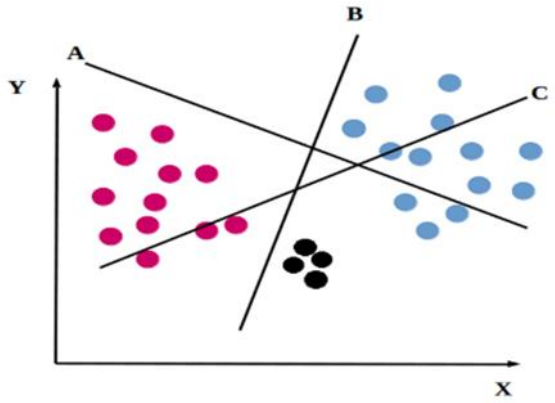


Figure 7. MultiSVM's determination of planes and boundaries [32]

Figure 7 represents the operating system of MultiSVM. In the real system, there are data taken from 25 images. Three planes have been created: Eczema, Skin Cancer, and Normal condition. Since the skin condition with the highest number of 25 images that make up the Multi SVM is skin cancer, the blue dots symbolize the skin cancer, the pink dots symbolize the normal state, and finally, the black dots with the least number symbolize the eczema disease. For each point, the health status of the skin is written in the relevant matrix. MultiSVM classification (fitcecoc) is performed.

The parameters appearing in TABLE 1 are extracted again for each added image, whether or not it is out of 25 photographs in the database, and its location in the MultiSVM plane is determined by the 'predict ()' estimation method. As a result, when a new image is given to the system, the system will determine how close it is to these three regions and accordingly classify whether it is one of the diseases or normal.

There are 51 stored images for skin cancer, 11 for eczema, and 10 for normal skin. It has the advantage of having many images. The values obtained from the images of these two diseases and normal skin in the database were compared with the data in the 'Diseasesetyeni. mat' file.

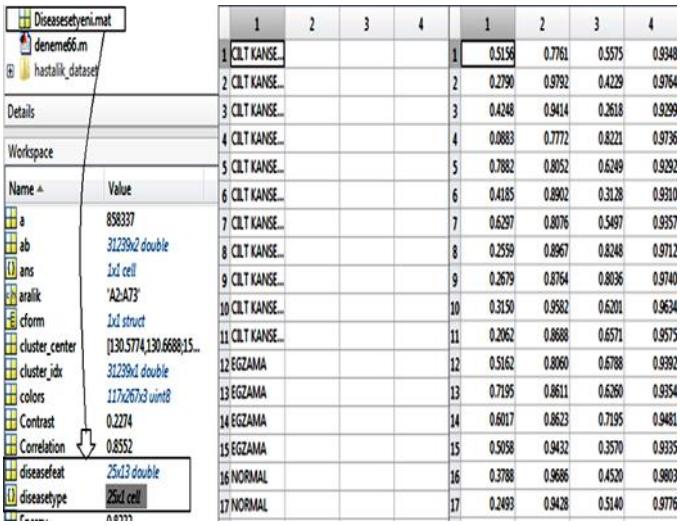


Figure 8. The matrices and their contents in Diseasesetyeni.m

Footnote: Skin cancer= Cilt kanseri, Eczama= Egzama

'Diseasesetyeni. mat' and its matrix and their values were created with 25 of the 72 images already in the system. However, the system gives consistent results when any new image that does not create the database is introduced. As mentioned before, the predict () function is used during the simulation.

Studies are carried out to find the feature with similarities, to know the obtained similarities, and to reveal the percentages. There is a limited amount of data on 'diseasesetyeni. mat'. However, this data set was sufficient to reveal the purpose and results of the current study. In the future, the study can be made more detailed by increasing the data set. The data of all images are not saved in 'Diseasesetyeni. mat'. In the study, it has been tried to prove that it gives correct results for the newly added images. For each image entered, the values are always extracted again and compared with 'Diseasesetyeni. mat'.

The values of skin cancer and eczema disease and normal skin are found in the matrices in the 'Diseasesetyeni. mat' file. The database contains all the necessary information that helps the user in the search process. The process consists of checking the values from 1 to 25 for each image and using the MultiSVM planes created from them to classify the image with the appropriate group and achieve the required similarity ratio.

The situations processed in the background of the code are:

Image is taken from the outside and displayed inside the system. The features of the image are extracted and saved in the workspace. The 'disease feat' and 'disease type' matrices in the 'Diseasesetyeni. mat' file is subjected to the MultiSVM training function and the 'svmStructDisease' structure is created. It is shown in Figure 8. Finally, the predict the disease function is performed, a match is made between the new image and the database, and the results of the predictions of the diseases or the normal state are given.

As a result, the system will notify the other party of the disease type or normal state. If the result of the image processing differs from the result of the decision trees, this will also be reported as an upstream message.

Table 2. Decision Tree Method

KAY	T3	TST	TSTT	TSH	MADTSH	Görüntü İşleme	Karar agac
1	113	9,9	3,1	2	5,90	cilt kanseri	egzama
2	105	7,3	1,5	1,5	0,10	cilt kanseri	cilt kanseri
3	114	9,9	2,4	1,5	5,70	cilt kanseri	cilt kanseri
4	106	4,2	1,2	1,6	1,40	cilt kanseri	normal
5	112	8,1	1,9	3,7	2,00	cilt kanseri	cilt kanseri
6	111	8,4	1,5	0,8	1,20	cilt kanseri	cilt kanseri
7	115	7,1	1,3	1,3	2,00	cilt kanseri	cilt kanseri
8	109	10,4	1,9	0,4	0,10	cilt kanseri	cilt kanseri
9	100	6,1	2,4	1,8	3,80	cilt kanseri	cilt kanseri
10	105	11,1	2	1	1,00	cilt kanseri	cilt kanseri
11	116	11,1	2	1,2	2,30	cilt kanseri	cilt kanseri
12	117	12,2	1,9	1,2	3,90	cilt kanseri	cilt kanseri
13	106	9,4	1,5	0,8	0,50	cilt kanseri	cilt kanseri
14	121	10,1	2,4	0,8	3,00	cilt kanseri	cilt kanseri
15	121	13,5	1,5	1,6	0,50	cilt kanseri	cilt kanseri
16	109	10	1,3	1,8	4,30	cilt kanseri	cilt kanseri
17	118	8,1	1,9	1,5	13,70	cilt kanseri	normal
18	115	8,1	1,7	0,6	2,20	cilt kanseri	cilt kanseri

Footnote: Skin cancer= Cilt kanseri

Eczama= Egzama

Image processing= Görüntü işleme

Decision tree= Karar ağacı

Record= Kayıt

Decision tree	Image processing: Normal	Cilt kanseri + Egzama	Sum
Normal	DT1	YT1	DT1+YT1
Cilt kanseri + Egzama	YT2	DT2	YT2+DT2
Sum	DT1+YT2	YT1+DT2	DT1+YT1+YT2+DT2
Accuracy=	(DT1+DT2)/	(DT1+YT1+YT2+DT2)	
Sensivity=	(DT1)/	(DT1+YT2)	
Selectivity=	(DT2)/	(YT1+DT2)	

	ROC1(Cilt K.)	ROC2(Egzama)	ROC3(Normal)	Sum
ACCURACY:	0,916666667	0,958333333	0,958333333	0,94
SENSIVITY:	0,94	0,833333333	0,9	0,89
SELECTIVITY:	0,863636364	0,983333333	0,967741935	0,94
DT1:	47	10	9	
DT2:	19	59	60	
YT1:	3	1	2	
YT2:	3	2	1	
	72	72	72	

Figure 9. Calculation of Sensitivity, Accuracy and Selectivity

Footnote: Skin cancer= Cilt kanseri, Eczama= Egzama

Table 2 and Figure 9 show the use of the decision tree method. DT1 is the number of images in which image processing for normal skin images yields parallel and accurate results against decision trees. DT2 is the number of images where image processing for skin cancer and eczema skin images yields parallel and accurate results across decision trees. YT1 is the custom of images where image processing for skin cancer and eczema skin images yields inconsistent and incorrect results against decision trees. YT2 is the custom of images where image processing for normal skin images gives opposite and wrong results against decision trees. Relevant clinical data used for decision trees Sakarya Uni. Taken from the lecture notes of the Data Mining course. In addition, the code lines in the system are the version of the Plant Disease Classification study in MathWorks File Exchange, developed with the decision tree method. In the excel data in Table 2, only some of the clinical data of 72 images in the picture folder are shown as representative. The health status of the skins was started to be classified by placing the limit of 5.2000 on the TST data found in clinical data. Decision trees progress step by step. Obtaining entropy information at each stage shows the user how well the separation progresses, and the closeness of entropy to 0 shows the opposite side that the process steps are approaching the end. Of the values included in the classification, T3 is the result of the blood test used when measuring the level of thyroid hormone, TST is the thyroid gland serum hormone test, TSH is the thyroid-stimulating hormone test, TSTT is the tubing scopic threshold test, MADTSH is the absolute difference from the thyroid-stimulating hormone test based on.

Entropy: It is mostly defined as randomness and disorder in a system and is used in many fields from statistics to technology. R is seen as a resource. It is assumed that this resource can generate n messages as {m1, m2, m3,..., mn}. All messages are generated independently of each other and the probability of generating mj messages is pj.

The entropy H(R) of the R source that produces messages with probability distribution P= {p1, p2, p3,..., pn} is as follows.

$$H(R)= -\sum_{i=1}^{n} pi * log2(pi) \text{ or } H(R)= -\sum_{i=1}^{n} pi * \log(\frac{1}{pi})$$

Finally, when the classification with decision trees is completed, the sensitivity, accuracy, and selectivity analysis of the system are performed according to the image processing by using the above formulas. It describes the results of the comparison of simulative data and numerical data.

3.Results

The operation of any system depends on the probability of occurrence of a particular event, this study includes a wide range of possibilities, it is necessary to compare images of a particular disease and define the category to which it belongs. Reference is made to studies that can compare a database of pathological images with images from outside the database and work with several models. While working on the identification and classification of diseases, it was observed that insufficient information was collected for the database, and the success of the process was low when only the image processing technique was used in the process. The database is very important and dermatology medicine needs to focus on improving the content of the database so that it is easier to identify any disease with images stored in its system. Moreover, adding the decision tree system to the study helps the user in the accuracy analysis part.

This problem was solved by extensive research on various websites associated with the idea of the study, and a sufficient number of images were found in terms of the number and quality of images. A series of images have been collected to be used directly in the database and by comparing them according to the inputs to be presented, the desired results can be obtained. In addition, all data and images were collected from public databases.

Skin diseases seriously affect people's lives and health. This study proposes an effective approach to identifying skin diseases. To increase the diagnostic accuracy of many types of dermatological diseases, it is necessary to replace traditional methods and develop automated methods. In this study, two types of skin diseases were defined, namely Eczema and Skin Cancer diseases, and it can also determine normal skin. The decision tree method adds strength to this study and distinguishes it from other studies. The program will detect and identify dermatological diseases. It was designed through the MATLAB program. This study will serve a large segment of society and will solve the problems in identifying and distinguishing the types of skin diseases in humans due to the similar effects and symptoms that humans are exposed to and will certainly contribute to the early diagnosis and treatment of skin diseases.

The detection of skin diseases has gone through several testing stages to make more accurate image processing. Simulations were made on two different diseases and normal conditions using the MATLAB program. This study is very useful in increasing the understanding of programming by using MATLAB software. After relying on the similarity rate for eczema and skin cancer diseases, the system was tested with new images. As expected, the result of the system will be an assistant for all users, including doctors and nurses. It is used with MATLAB software technology. The system can be used by any person, be it a professional, a programmer, or a regular person.

In the first part, the aim is to match the images to one of the selected diseases and to reach the basic principles for estimating the similarity ratio when an image is uploaded to the system, and a training model that can always be improved has been successfully obtained. The number of items in the database has been increased and since this study will help many people in the hospital, the system has been designed with an easy-to-use interface to make it much easier to use, and this section has been completed very successfully. Apart from the professionally working GUI interface of MATLAB, uniquely written codes were used, and this helped a lot in recognizing the backside of the process. Energy feature extraction and extraction of other features were used on images of different diseases taken from the user and not included in the database system, but which will be compared with this particular database.

The use of skin diseases detection methods will positively affect the field of medicine in terms of early detection of the disease. Since the patients will not have to stay in the hospital for a long time to perform the examination procedures, the number of patients will be improved. It also saves time for nurses and doctors, making the process go faster, and more time is allocated for other diseases that require more than one procedure to complete the examination. In all its aspects, the importance of developing a skin disease detection system is how to extract features from the images, including the acquisition phase of the images, and increase the accuracy of the study, which will make a significant difference to community users. It certainly has benefits and reduces diseases and it is recommended that government or private organizations working in the medical field support and invest in this method for the following reasons. These reasons:

1. Increasing number of skin diseases
2. Ability to use multiple features for higher resolution such as contrast,
3. Ability to extract entropy, correlation, and homogeneity features,
4. Purpose of using a camera that takes pictures of the patient directly in the future
5. Application of different images of a different disease,
6. It is listed as the purpose of having the function of determining the amount and type of drug-assigned to the patient in the future.

Acknowledgement

In this study, gratefully thanking to other studies with similar ideas that supported and cited as references for this article. And would also like to thank to the JIET community for supporting the publication of this work.

Author Contribution

In the study carried out, Yusuf Süer Erdem contributed to the creation of the model, literature review, data collection and processing, while Özhan Özkan contributed to the formation of the idea, the evaluation of the results and the spelling of the article.

Conflict of Interest

The authors confirm that there is no known conflict of interest or common interest with any institution/organization or person. Ethics committee approval is not required for the prepared article. There is no conflict of interest with any person/institution in the prepared article.

References

[1] Kaplan, M. Implementation of an Auxiliary System for the Diagnosis of Skin Diseases Using Artificial Neural Networks, TC Fırat University, Institute of Science, Master Thesis, 2016.

[2] Season, V. (2010). Epidemiology of dermatological diseases, disease burden, and place in primary care. Turkey Clinics J. Fam Md-Special Topics, 2010; 1(2): 15-20.

[3] Alonso, DH., Wernick, MN., Yang, Y., Germano, G., Berman, DS., Slmoka, P. Prediction of cardiac death after adenosine myocardial perfusion SPECT based on machine learning. *J Nucl Cardiol.* https://doi.org/10.1007/s12350-017-0924-x, 2018.

[4] Narula, S., Shameer, K., Salem Omar, AM., Dudley, JT., Sengupta, PP. Reply Deep learning with unsupervised features in echocardiographic imaging. *J Am Coll Cardiol*; 2017, 69: 2101–2.

[5] Esteva, A., Kupre, B., Novoa, RA., et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017, 542:115–8.

[6] Cichosz, SL., Johansen, MD., Hejlesen, O. Toward big data analytics: a review of predictive models in the management of es and its complications. *J es Sci Technol*, 2015, 10(1):27-34.

[7] Tran, BX,. Latkin, CA., Giang, VT., et al., The Current Research Landscape of the Application of Artificial Intelligence in Managing Cerebrovascular and Heart Diseases: A Bibliometric and Content Analysis. Int. J. Environ. Res. Public Health, 2019, 16:2699.

[8] Char, DS., Shah, NH., Magnus, D. Implementing Machine Learning in Health Care, Addressing Ethical Challenges. *N. Engl. J. Med.*, 2018, 378: 981–983.

[9] Celebi, V., Inal, A. Problem of Ethics in the Context of Artificial Intelligence. *The Journal of International Social Research*, 2019, 12, 66.

[10] Mujumdar, A., Vaidehi, V. Dibetes Prediction Using Machine Learning Algorithms. *Procedia Computer Science*, 2019, 165: 292–299.

[11] Farid,D., Sadeghi,H., Hajigol,E. ve Parirooy,N. Classification of Bank Customers by Data Mining: a Case Study of Mellat Bank branches in Shiraz, *International Journal of Management Accounting and Economics*, 2016, 3: 534-543.

[12] Walsh, S. Applying Data Mining Techniques Using SAS® Enterprise Miner- Course Notes, SAS Institute Inc., North Carolina, 2005.

[13] Pratt, W. K., Digital Image Processing. USA: John Wiley & Sons, 2007.

[14] Nixon, M. S., Aguado, A. S., Feature Extraction, and Image Processing. Newnes, UK, 2002.

[15] Kaur, H., Kumari, V. Predictive modeling, and analytics for diabetes using a machine learning approach. *Applied Computing and Informatics* https://doi.org/10.1016/j.aci.2018.12.004, 2018.

[16] Kavakiotis, I. et al. Machine Learning and Data Mining Methods in Diabetes Research. *Computational and Structural Biotechnology*, 2017, 15: 104–116.

[17] Araújo F.H.D. et al. Using machine learning to support healthcare professionals in making preauthorization decisions. International Journal of Medical Informatics, 2016, 94:1–7.

[18] Parikh, R.B., Kakad, M., Bates, DW. Integrating predictive analytics into high-value care: the dawn of precision delivery. *JAMA*, 2016, 315: 651-652.

[19] Bates, DW., Saria, S., Ohno-Machado, L., Shah, A., Escobar, G. Big data in healthcare: using analytics to identify and manage high-risk and high-cost patients. *Health Aff*, 2014, 33: 1123-1131.

[20] Mercaldo, F., Nardone, V., Santone, A. Diabetes Mellitus Affected Patients Classification and Diagnosis through Machine Learning Techniques. *Procedia Computer Science*, 2017, 112: 2519-228.

[21] Al-Khafaji, Muallah, S. K., Ibraheem, M. R. Detection of Eczema DISEASE by using Image Processing. *The Eurasia Proceedings of Science, Engineering & Mathematics (EPSTEM)*, 2018, Volume 2: 2602–3199.

[22] Acıbadem Web and Medical Editorial Board. Skin (Skin) Cancer. Acıbadem Healthcare Group. 2019, from https://www.acibadem.com.tr/ilgi-alani/cilt-deri-kanseri/#signs

[23] Al Shahibi, I. S. S., Koottala, S., Detection of Skin Diseases Using Matlab. Journal of Student Research Fourth Middle East College Student Research Conference, Muscat, Sultanate of Oman, 2020.

[24] Mathworks Help Center, Retrieved December 6, 2021, from https://www.mathworks.com/help/stats/fitcecoc.html

[25] https://atozmath.com/example/CONM/Bisection.aspx?he=e&q=it

[26] Houcque D., Introduction To MATLAB For Engineering Students, https://www.mccormick.northwestern.edu/documents/students/undergraduate/introduction-to-matlab.pdf, 2005.

[27] Kumar, B., Rai, S.P., Saravana Kumar, U., Verma, S.K., Garg, Pankaj K., Vijaya Kumar, S.V., Jaiswal, R., Purendera, B.K., Kumar, S.R. and Pande, N.G. Isotopic characteristics of Indian Precipitation. Published online in *Water Resources Research*, 2010, Vol. 46, DOI: 10.1029/2009WRSR008532, 2010.

[28] Yurtay, N., Adak, M. F., Dural, D., Serttaş. S. A study on use of decision tree method in the diagnosis of thyroid disease. International Science and Technology Conference, Retrieved 28 November 2021. Published, 2012.

[29] "Dermatitis" defined, Suzanne Smith, Susan Nedorost, 2012.

[30] MathWorks Help Center, Retrieved December 6, 2021, from https://www.mathworks.com/help/stats/fitcecoc.html

[31] Image Texture Feature Extraction Using GLCM Approach, P. Mohanaiah, P. Sathyanarayana, L. GuruKumar, 2013.

[32] Data Scientist Website, Retrieved December 12, 2021, from https://veribilimcisi.com/2017/07/19/destek-vektor-makineleri-support-vector-machine/

[33] https://www.researchgate.net/publication/221608588

[34] Yurtay, N. Data Mining Applications Lecture Notes, Week 4, Sakarya University, Retrieved 06 December 2021.

Journal of Investigations on Engineering & Technology
Volume 5, Issue 1, 63-75, 2022

ISSN: 2687-3052
Derleme Makalesi

Radyasyon Fiziği Çalışmalarında Bazı Kristal ve Amorf Malzemelerin Kullanılması

Oğuzhan Dervişağaoğlu[1]

(Alınış / Received: 11.04.2022, Kabul / Accepted: 22.06.2022, Online Yayınlanma / Published Online: 30.06.2022)

Anahtar Kelimeler

Katı hal fiziği
Amorf
Kristalografi
Dozimetri
Örgü
Termolüminesans Dozimetri (TLD)

Özet: Bu derleme makalesinde, bazı amorf ve kristaller hakkında yapılan çalışmalar gösterilmiştir. Kristalografinin alanına giren, kristal örgü ve yapı konusu da, katıhal fiziği perspektifi altında açıklanmıştır. Kristal büyüme yöntemleri de kısaca açıklanıp, obsidyenin termolüminesans dozimetri (TLD) için kullanılması tavsiye edilmiştir. Obsidyen, radyasyon korunması üzerine yapılan çalışmalarda kullanılmıştır. Bu makalede, obsidyenin radyasyon dedeksiyonu için uygun bir malzeme olabileceği düşünülmüştür.

Usage of Some Crystals and Amorphous Materials in Radiation Physics Studies

Keywords

Solid State Physics
Amorphous
Crystallography
Dosimetry
Lattice
Thermoluminescence Dosimetry (TLD)

Abstract: In this review article, studies of some amorphous and crystalline have been viewed. Lattice of crystalline related with cristallography has been elucidated on the perspective of solid state physics. Furthermore, usage of obsidian has been recommended for thermoluminescence dosimetry (TLD). Crystalline growth methods have been explained briefly. Obsidian has been utilized in studies which are related with radiation protection. In the article, obsidian has been contemplated as a suitable material for radiation detection.

1. Introduction

Main purpose in this review article is to introduce studies about some crystalline and amorphous materials related with radiation physics applications and also another aim is to propound a new idea with respect to amoprhous material such as obsidian whether can be used in radiation applications or not. Dosimetric application of both nuclear and radiation physics have made a huge stride on development of new dosimeters which are utterly crucial to understand the nature of some crystalline and amorphous materials. Material at higher temperatures can emit red(infrared-visible light) radiation and materials which will be used in dosimetric studies would be either semiconductors or insulators. Metals will not be appropriate tool to observe a luminescent properties. Luminescence effect has been appeared by heating materials which are suitable as mentioned above.

[1] Ege University, Nuclear Sciences Institute, 35040, Izmir, TURKEY

Oğuzhan Dervişağaoğlu / promedyum@hotmail.com

Heating is a kind of radiation and when the heat has stimulated, the material will store some energy with ionizing radiation, which means material absorbs some energy. Then, stored energy has been released as visible light but if the material has been cooled, TL properties(Thermoluminescence) of material have been terminated. Re-exposed process of ionising radiation is necessary to see visible light upon crystal or semiconductor again [1].

OSL (Optically Stimulated Dosimetry) is a procedure or natural material such as quartz or feldspar [2]. Because of that reason, optically stimulated dosimetry has been suggested by scientists [3]. Radiation dosimetry generally has been used in the application such as health physics, x-ray studies, radiation protection and so on. Borate phosphors is effective and close to human tissue with an atomic number which is Z=7.4. Because of that reason, coefficient of absorption of borate phosphors is as same as biological materials. Borate phosphors has some advantages which are as follows; wide energy band gap, sensitivity of neutron, thermal stability, optical characteristic and etc [4], [5]. Piezoelectric quartz has been studied in dosimetric application as dosimeter. In a study which focuses on development of quartz dosimeter and determining of different corrosivity caused by photooxidizing agents and volatile organic acids has been resulted by using reactive lead coating was applied to quartz crystal. Quartz vibrates coating lead that has an effect on lower frequency [6]. The other study on energy band gap of potassium lithium borate glass has been carried out by the applying FTIR(Fourier Transform Infra Red) and UV-VIS(Ultraviolet-Visible Spectrophotometry) spectroscopy techniques [7]. Next study has been related with the thermoluminescence of C doped lanthanum aluminate crystalline ($LAlO_3$:C) synthesized by **Solid State Reaction** with U.V(Ultraviolet) dosimetry [8]. Another research has been done about the subject which is thermoluminescence and radioluminescence of a sodium doped feldspar. Sodium has been used for hole traps(Correcher et al, 2007) [9].

One of the oldest studies in dosimetry of alkali halide crystal has been irradiated with gamma and x -rays [10]. Photoconductivity has been efficient on imperfection lattice of crystals by doping H(hydrogen) or F(fluorine) [11], [12]. Another study in luminescence about crystalline has been done by scientists (Jang et al, 2012) on Eu^{+3} doped fluorthalenite ($Y_3Si_3O_{10}F$) for laser excitation spectroscopy for understanding temporal behavior of crystal[13]. Next research has been done on Ag doped and undoped lithium tetraborate($Li_2B_4O_7$) by using **Czochralski Method**(Kuralı et al,2016) [5].Furthermore, diamond dosimeter has been studied and these types of dosimeters are adequate devices for small field dosimetry. Single crystal devices have shown efficient performance, where pollycrystalline can be into bigger size and they also have demonstated low quality due to structural imperfection [14].

Another study on radiophotoluminescent dosimetry for carbon and magnesium doped AI_2O_3 crystal [15]. Sm:$CaNb_2O_6$ single crystal has been studied by Czochralski method which includes high x-ray diffraction, optical absorption and fluorescence [16]. Thermoluminescence dosimetry properties of Ge doped photonic crystal fibres (PCFs) have been evaluated with **Modified Chemical Vapour Deposition** method (MCVD) by scientists[17]. LiF material has been doped with elements which are P, Mg and Cu in a study to grasp the effect of thermoluminescence [18]. Hydrogenated amorphous silicon has been studied due to its own intrinsic radiation hardness for dosimetry and particle detection such as x-rays. Hydrogenated amorphous silicon has been thought as efficient candidate of radiation resistance quality [19], [20].TLD property of pencil lead graphite has been evaluated in another scientific study [21]. Another investigation have been done K_2YF_5:Tb^{+3} as well as undoped K_2YF_5 has been synthesized under hydrothermal conditions. Linearity of dose response and reproductivity of dose have been emerged by scientists [22]. Lucin metavariscite ($AlPO_4.2H_2O$) has been studied with different characterization techniques. TL has an important role to see structural changing in material under 150^0-165^0C. Transformation of metavariscite into berlinite has caused by loss of water molecules and crystal lattice changing from monocilinic to trigonal [23]. Kinetic parameters of La doped phosphate glass under the TL glow curves of beta radiation by applying some methods such as peak shape (PS) methods and computerized glow curve deconvulation code (CGCD) [24]. In another paper, magnesium oxide (MgO) with co-doped Li^+, Ce^{+3} has been synthesized with the method called **Solid State Reaction Technique** (Guckan et al ,2021) [25].Other research has been done on boron doped hydrogenated amorphous silicon by using co-sputtering technique related with PECVD film samples [26].In another research, amorphous silicon sensor has been put on the top of integrated circuit and radiation detectors based on amorphous silicon industrialized in very thin films and also amorphous silicon has been used in solar cells for space application[27], [28], [29]. A new study on susceptibility of obsidian has been done in Japan and also obsidian has been characterized with different crystals such as microlites and phenocrysts [30]. Goksu and Turetken had studied on natural sources such as obsidian [31]. Other research is about to obsidian pellets in a powder will be adding teflon and these pellets have been exposed to gamma radiation of ^{60}Co in dose interval between 10 Gy-10 kGy and TL,OSL, TSEE techniques have been used in study of obsidian materials which can be a candidate of detector [32]. Alzahrani et al, investigated on obsidians under U.V. the situation of TL and PTTL[33].

The writer has written an article related with obsidians and he has made a mistake by writing a sentence "% 75 of obsidian contains quartz" which has not been proved in any study.These both materials are similar to each other chemically because of SiO_2 compound and furthermore, the writer of this article had thought that obsidian can be useful for radiation detection and it can be an adequate tool as target for particle accelerators. He has declared an opinion about this subject upon his own previous article [12].

2. Basic Information about Amorphous and Crystal

In the beginning of the 1900's, development of quantum mechanics had a crucial impact on forming the theory related with the energy band gap of solids. Correlation between crystals can be managed with the mechanic, thermal, electrical and magnetic qualities of solids. The main aim of crystal physics is to understand relationships between particles of microscopic and crystal structure [34]. Crystal physics studies consist of geometrical forms and physical properties such as doing experiment upon x-rays, electron, neutron beams and etc. Johannes Kepler had thought on the reason why snowflake has 6 corners instead of has 5 or 7 corners. He had made assumptions on the geometry of crystalline. Robert Hooke and Rene Just Haüy had studied on providing the mathematical theories[37].Picture1 shows amorphous and crystal stones which are taken from general chemistry notes[35]. Picture 2 demonstrates lattice models of both amorphous and crystal solid[36].

Figure 1. Galena(PbS), Pyrite(FeS_2) and Quartz have crystal structure and obsidian is an amorphous material, (General Chemistry Notes)

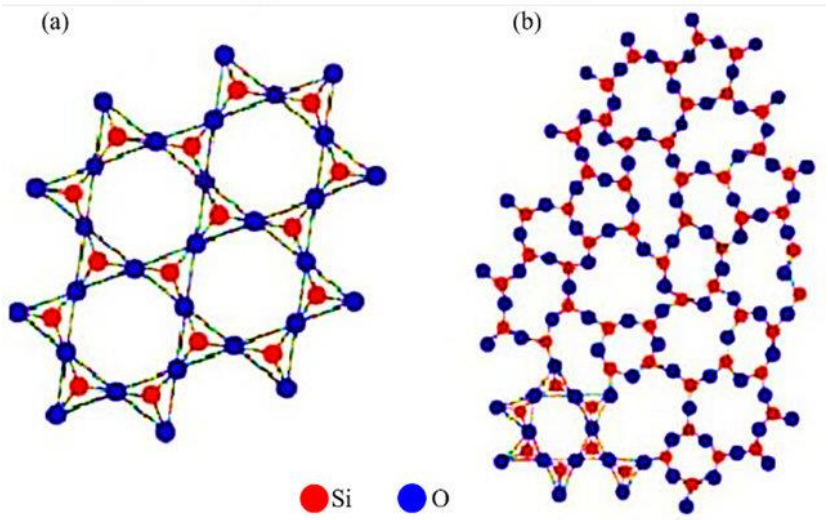


Figure 2. Crystal and amorphous lattice (Atif et al,2020)

Amorphous means " shapeless" and that word originally comes from ancient Greek word which is "ámorphos" [35]. Crystal lattices are consisted of three dimensional system. These systems are originated by atoms or group of atoms. In three dimensional space, according to position of atoms in crystal there are 14 different lattices or Bravais lattice and the word crystal comes from "krystallos" which means "ice" [37].

2.1. Classification of Crystal Lattices

Cubic system: In this type of lattice system there are three different models which are; primitive, body-centered and face-centered. Connection of any two different cubic lattices is impossible. Tetragonal system: This type has included primitive and centered and there is no Bravais connection. Orthorhombic system: in this system there are four types of lattice that can be categorized as; primitive, base centered, body centered. There is Bravais connection between primitive and base centered which is defined as "phase transition" [38]. Trigonal or Rombohedral: It has three equal planes and two angles are also equal except one angle. Hexagonal system: It has six rotational axis and every each corner has one atom. One atom has placed in the surface and in the center, there are three atoms. Monocilinic system: It has different sides and one angle is not equal to 90^0. Triclinic system has no equal sides and angles are not equal to 90^0. Table 1, can give an information to the readers with different properties of crystals with respect to their geometric shapes [39]. Coordinate system have been defined as three dimensions and have been composed of non-coplanar vectors. Basis vectors can be called as a,b and c. Angle of intervectors α, β, θ are called as metric parameters. Different evaluations of basis have been originated by scalar products of all basis vectors. Changing from one basis to another basis has been explained by transformation matrix (P).This transformation can be described as in the followings;

$$(a, b, c)P = (a', b', c') \tag{1}$$

$$G' = P^T G\, P \tag{2}$$

G and G' stands for metric tensors [40].

Table1. Types of lattice in crystals (Kashif,2018).

Systems	Axis/Sides	Angles	Lattice Symbol	Examples
Cubic	a=b=c	α=β=θ=90	P,I,F	Fe,Cu,NaBr,Au,Diamond,NaCl
Hexagonal	a=b≠ c	α=β=90, θ=120	P	Mg,Zn,Graphite,,ZnO,Ice
Tetragonal	a=b≠ c	α=β=θ=90	P,I	NH_4Br, SnO_2, Sn
Rhombohedral	a=b=c	a=b=c<120(≠90)	P	Bi, $NaNO_3$, KNO_3, As, Bs
Orthorhombic	a≠b≠c	α=β=θ=90	P,I,F,C	$BaSO_4$, K_2SO_4,
Monoclinic	a≠b≠c	α=θ=90≠ β	P,C	Sugar, Sulfur
Triclinic	a≠b≠c	α≠ β≠ θ	P	$K_2Cr_2O_7$, H_3BO_3

2.2. Amorphous Materials

Macroscopic crystals are made of microscopic sized crystals. Positions of every each atom in spectacular crystal have been extrapolated by the formula as following;

$$r_{atom} = M(r_0, atom) + u.\vec{a} + v.\vec{b} + w.\vec{c}$$ [41], [42]. (3)

Positions of atom in amorphous materials have not been calculated as positions of atom in crystals. Amorphous materials have been supposed as disordered crystalline solids [43]. Most of amorphous solids do not shape up into crystalline form. Amorphousness is generally explained by what it is not rather than what it is in lattice of solid. Three methods have been used to determine the identity of amorphous solids in which has been carried out for experiment related with solid-state physics.(i) calorimetry to calculate temperature transition of glass (ii) x-ray scattering (iii) NMR to comprehend correlation of bonds in solid. Problem of disordered crystalline can be repaired but a new method should have been considered to grasp the structure of random atom in amorphous material. Set of random data has been set up with theorical point of view. The meaning of "disordered" is generally described in dictionaries as unpredictable and something opposite to the order [41].Studies in amorphous liquids have been accelerated with theorical approach by Bernal[44] . In the last two decades, study interests have been divided into two different streams which are; (i) studies related with packing of both spheres and molecules [45], (ii) simulations for atoms by **ab-initio method**[46]. Simulation modelling has been applied in the packing of spheres randomly with extremely cold conditions such as the rate of the order 10^{15} K/s. Even so these methods succeeded to simulate the structure of amorphous materials by considering atomic arrangements which possessed imperfections. Ideal amorphous structure has not been officially accepted and lattice of atomic arrangements in amorphous materials have been keeping this mystery [47]. Adams and Matheson had studied on computation of dense random packings of hard spheres[48]. ,Finney also has done scientific study on random packings and the structure of simple liquids[49]. Bennet studied on defining of packing of spheres for creation of ideal amorphous solids. Bennett started with regular tetrahedron and formed random packing. It has been proved precisely which one has imperfection by Clarke and Jonnson [50], [51.]There are different scientists that made a precious contribution on the subject of packing spheres either related with lattice or not related with lattice [52], [53], [54]. Obsidian is the natural amorphous material which can solidify when it interact with oxygen. Amorphous solids can be categorized into two main groups;

- Crystallisable (metallic and metalloid glassses)
- Naturally non-crystallisable (cross- link polymers, atactic organic polymers) [41] .

When ab-initio has been emerged with Hamiltonian equation that let us to conceive intuitional with perspective of quantum mechanics

$$I\theta(t)\rangle = exp\left(-\frac{iHt}{h}\right) I\theta(0)\rangle \quad (4)$$

This equation managed interaction between atoms and also carried out the tendency of self assembly and furthermore, fluctuations took over the dynamic of the system. Random Gaussian distribution has been followed by spontaneous density fluctuations [55] , to put it simply, all materials comprise larger or lesser defects which refer to crystalline structure, or contain flaws which correspond to amorphous materials.
Ideal amorphous solid model is to represent the spheres/atoms which are randomly packed and non-intersection spheres and there are two rules for the perfectly random arrangements of atoms;

- There are three non-touching spheres and centers of these three spheres which are adjacent should create irregular triangles [56].
- Spheres should be in fixed position for mono atomic amorphous solid which is ideal in fixed position when $4 \leq k \leq 9$ touching contacts with spheres which are neighbours, for $k>9$ sphere is in fixed position permanently [57].

Amorphous solid which is obsidian can be heated into quartz tube at the temperature in the range between 100-700^{0}C. TL measurements were determined by Harshaw 3500 TLD apparatus with light pulses. First TLD read out confirmed to use the time from the room temperature which goes linearly to 400^{0}C and second readout belongs to black body radiation[33]. Black obsidian has been crushed into small pieces and then it has been pulvarized by agate mortar and pestle porcelain and powder has been weighed, washed and dried for 24 hours. It was mixed with teflon by using liquid nitrogen to confirm homogeneity. Obsidian+ teflon have been irradiated at the center of radiation technology, IPEN, Atomic Energy of Canada LTD with Co-60. All test which are done related with dosimetric properties of material with respect to TL, OSL and TSEE dosimetry. Obsidian can be a good candidate as dosimeter[64].

3. Methods of Crystal Growth and Usage of Amorphous in Radiation Studies

Crystal growth from liquid can be categorized into six groups[58] ;

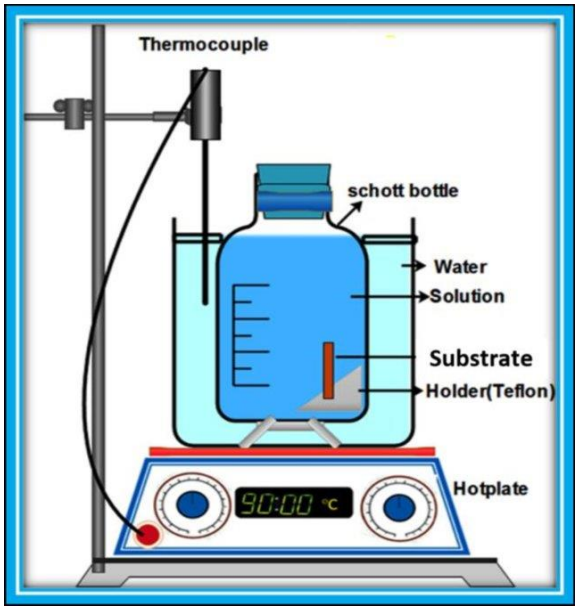
- Gel growth
- High temperature
- Hydrothermal growth
- Electrocristallization
- Low temperature solution
- Melt growth

Alves et al, have done a research on lanthanium aluminate polycrystals by using solid state reaction method which has been mixed as equimolar ratios of both lanthanium oxide and aluminium oxide. Mixed powder has been granulated in agate mortar after powder were weighted with 0.1% weighted carbon atoms. The first approach in methodology in this study, Lanthanium oxide and aluminum oxide were put into agate mortar with 0.1 weighted of graphite and this mixture has been put into sintered process in hydrogen atmosphere at 1770^{0}C for 2 hours. The second way of the method is aluminium oxide powder has been mixed with 0.1 % carbon and has been sintered in the same conditions. The third method is that lanthanium oxide has been put into same procedure. U.V radiation has been applied the materials and captured charged carriers removed from traps by using Regaku D/Max Ultima X-ray diffractometer. Thermoluminescent analysis has been done by using Harshaw TLD-4500 reader[8].

High temperature crystal growth has been originated by the category of high temperature solutions growth. Constituents of crystallized material were dissolved in convenient solvent and crystallization occupied as a solution which has been supersaturate. Supersaturation can be advanced by solvent evaporation with respect to cooling process of solvent or transport process in which solute is confirmed to flow from hot region to cooler region. High temperature crystal growth can be categorized into two groups which are;(i) growth from single component system and (ii) multi-component system. In this method, solid has been taken as the solvent and growth have stood under melting point. This method can be useful for oxide compounds that have high melting point [58].

Hydrothermal growth, some materials tend to dissolve in water under high temperature and pressure in NaOH and $CaCO_3$ which are mineralization agents. Single crystal has been grown under high temperature and pressure.

Quartz can be grown by using this method which is generally called hydrothermal growth method. According to this method crystals have been grown into apparatus named autoclave and solution has placed at the bottom of hot region. Seed crystal has been placed into cooler region. Autoclave will be full of solvent which is definite amount that also can be a character as interior pressure in the system. Solution in the system will cause a rise in the heat because of temperature differences between upper and lower solution and bring solution at high temperature into growth region where material being crystallized in seed crystal[59]. Hydrothermal method has been shown in the Picture 3[65].



Pigure 3.Hydrothermal method;(Mohammadet al, 2020)

Another method on crystal growth is Czochralski method figüre 4 [62]that allows obtaining crystal of many intermetallic compounds. Crystallization methods have been carried out by Czochralski in 1916 and this method focused on taking crystal out from pure metals such as Pb Zn in air[60], [61]. Then this method has been applied on the crystal growth in semiconductors and oxide for electronic ındustries. In this method, Crystal has been taken out from melt after seed crystal immerse into melt region and it began to grow under an adequate temperature conditions. Grown crystal has been detached from melt region when growth process has been terminated[62] .

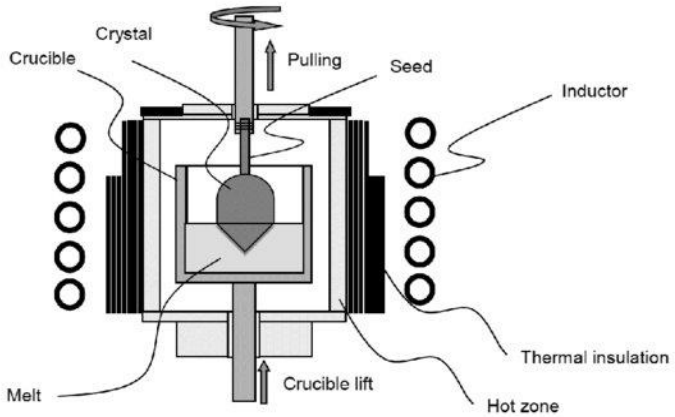


Figure 4. Czochralski method; **(**Kouwenberg,2018)

Bridgman growth method (BGO) is also used for crystal growth generally called directional solidification method. According to schematic diagram, , it consists of furnace and crucible holder [63]. Seed crystal has been put into little chamber with respect to vessel and at the bottom, quartz tube has been placed vertically. Crystal solidified from the seed by changing heat element under control with increasing temperature, intersection of solid-liquid along with it will be elevated. Most of scintillation crystals have been made with this method related to the subject

of crystal growth. BGO Crystals with this method also used in satellite for discovery of dark matter in cosmology studies.

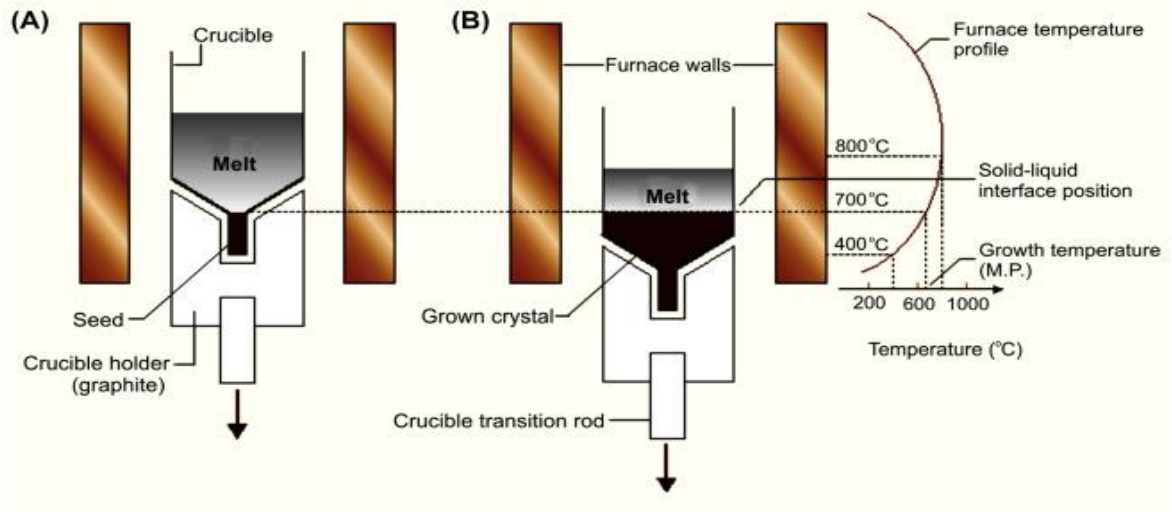


Figure 5. Bridgman method, (Venkataraman,2020)

Heat exchanger method (HEM)is directional solidification technique and this method includes gas-cooled heat exchanger which has been under the seed crystal. Growth has been begun by increasing cooled gas flow and decreasing temperature of furnace. Heat exchanger has a role of controlling temperature at the growth interface of crystal. Heat exchanger output has been controlled by helium gas. Sapphire crystal can be grown by HEM method. Evaporation method, circulation method, sublimation method, evaporation method, top seeded solution method, cold crucible method, floating zone method also can be used for the process of crystal growth [59]. Picture 6 has been cited from a study of Khattak and Schmid [66].

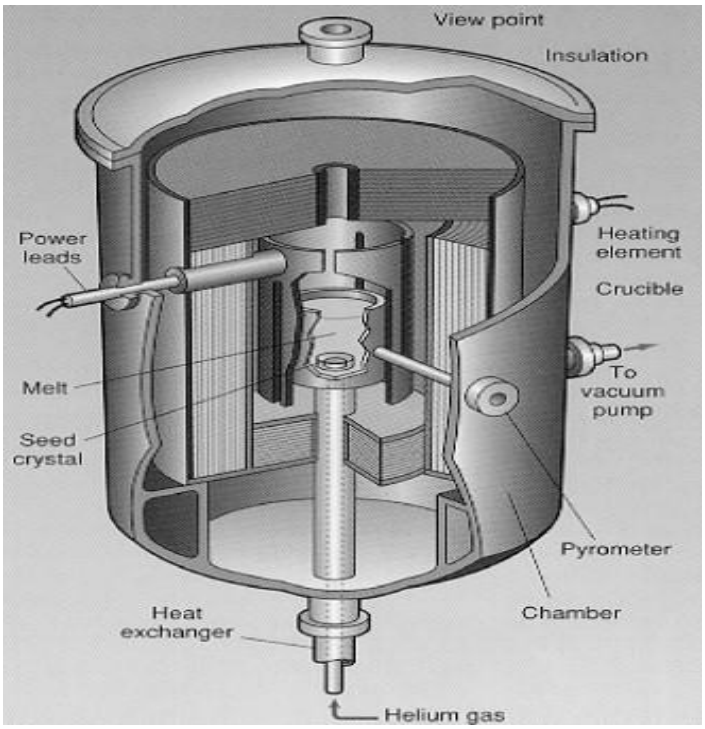


Figure 6. Heat exchanger method (Khattak and Schmid, 2001).

Understanding of phase equilibrium is utterly necessary knowledge to provide efficient crystal growth. Homogeneous part of macroscopic mass system shows the same atomic arrangement which is phase of material and each phase uniform structure along its volume. Phases have been ruled by Gibbs which gives a knowledge about equilibrium state of a material also it can be described as following[59];

$$F = C - P + 2 \quad (5)$$

F: number of intensive degree of free state.
C: number of components in the system.
P: number of phases.
Number 2 demonstrated both pressure and temperature which can be changed.

Vapor Crystal System

Equilibrium temperature is T_0 and saturation vapor pressure can be shown with P_0. If $P_1 > P_0$, vapor will be in supersaturated state and P_1 stands for supersaturated vapor pressure. If vapor is assumed as ideal gas, metastable phase defined by P_1 and change in free energy (Gibbs free energy) should be; [59]

$$\Delta G = \int_{P_0}^{P_1} RT_0 \frac{dP}{P} = RT_0 \, In \frac{P_1}{P_0} \quad (6)$$

(Ye, et al,2017)

R: constant of gas

$\Delta G =$ is force for crystal growth from vapor, released energy along phase transition

Solution Crystal System

Solution and crystal in the system saturation concentration will be C_0. Constant pressure and temperature C_1 bigger than C_0. Solution is supersaturated in metastable state. Driving force in this system related crystal growth will be Gibbs energy for growth process of material[59].

$$\Delta G = RT_0 \, In \frac{C_1}{C_0} \quad (7)$$

Melt Crystal System

Molten phase and solid crystal phase endures in an equilibrium state [59];

$$\Delta T = T_m - T \quad (8)$$

$$\Delta H(T_m) = T_m \Delta S(T_m) \quad (9)$$

$\Delta H(T_m)$ = molar enthalpy; (10)

$T_m \Delta S(T_m) =$ Entropy difference between crystal and melt states (11)

4. Results

In this review article, lattices of both crystal and amorphous were explained concisely by guiding oldest studies which are also related with luminescence dosimetry properties. Theorical approach has been made a clear shortly. Amorphous structure has been clarified by considering that it has flaw and crystal has disorder. Some hydrogenated amorphous silicon has been doped by H or F to observe photoconductivity properties. Thermoluminescence, carried out interaction between radiation sources and response of material in which goes through process of crystal growth method or mixing process (doping as PECVD method) for amorphous. Obsidian can be an efficient dosimeter for development of dosimetry and it can be useful on the subject of different studies related high energy physics such as particle accelerators for discovery of dark matter, neutrino and so on.

References

[1] Bos A.J.J. (2007), Theory of thermoluminescence, Delft University of Technology, Faculty of Applied Sciences, 2629 JB Delft, The Netherlands.

[2] McKeever S.W.S. (2001). Optically stimulated luminescence Dosimetry, *Nuclear Instruments and Methods in Pyhsics Research B 184*, Pages 29-54.

[3] V.V Antonov_Romanksi, I.F. Keirum- Marcus, M.S. Porishina, Z.A. Trapeznikova (1955). *Conference of the Academy od Science of the USSR on the peaceful Uses of Atomic Energy, Moscow*, USAEC reportAEC-tr2435(Pt.1)259,1956.

[4] K.H. Gavhane, M.S. Bhadane, P.P. Kulkarni, V.N. Bhoraskar, S.D. Dhole, S.S. Dahi. (2020). Investigation of novel Eu doped SrDy2O4 microphosphor for thermoluminescence dosimetry, *J. Lumin. 117781*, doi:10.1016/j.jlumin.2020. 117781.

[5] Kuralı, D., Ekdal Karalı, E., Kelemen, A., Holovey, V., Can, N., and Karalı, T. (2016), Thermoluminescence characterization of Ag-doped Li2B4O7 single crystal materials. Luminescence, doi: 10.1002/bio.3252

[6] Odlyha M., Jakiela S., Slater J.M., Bozec L., Bergsten C.J., Gronthoft T., Dahlin E., Colombini M.P., Bonaduce I., Thickett D., Larsen R., Scharff M., PIEZO-ELECTRIC QUARTZ CRYSTAL -BASED DOSIMETRY AND DAMAGE ASSESSMENT OF PARCHMENT AND ARTISTS'VARNISH.

[7] Aboud H., Wagiran H., Hossain I. Hussin H.(2012). Infrared Spectra and Energy band gap of Potassium Lithium Borate glass dosimetry, *International Journal of Physical Sciences Vol. 7(6)*, pp. 922 - 926, Available online at http://www.academicjournals.org/IJPS DOI: 10.5897/IJPS11.1744.

[8] Alves N., Ferraz W.B., Faria L.O. (2019).Thermoluminescent properties of LaAlO3:C crystals synthesized by solid state reaction applied to UV Dosimetry, *BRAZILIAN JOURNAL OF RADIATION SCIENCES* pages 01-14, ISSN: 2319-0612.

[9] Correcher V., Garcia- Guinea J., Sanchez-Munoz L., Rivera T. (2007).Luminescence characterization of a sodium rich feldspar, *IX International Symposium/XIX Nacional Meeting on Solid State Dosimetry,* pages 28-41.

[10] D. Kahn .(1956). Radiation dosimetry using alkali halide crystals and the absorption of betatron bremsstrahlung in water and bone, Acta Radiologica, **46**:3,1956 563-569, DOI: 10.3109/00016925609171447.

[11] Ruud E.I. Schropp, M.Zeman(1998). "Amorphous and Microcrystalline Silicon Solar Cells: Modeling, Materials and Device Technology", *Kluver Akademik Yayınevi, Boston*.

[12] Dervişağaoğlu, O.(2020). Usage of Obsidian (Igneous Energy) in Both Radiation Detection and Fixed-Target Machine in Discovery of Subatomic Particles . *Journal of Investigations on Engineering and Technology,***3**(1),6-9 https://dergipark.org.tr/tr/pub/jiet/issue/55612/703683.

[13] Jang K. H., Khaidukov N.M., Tuyen V. P., Kim S. I., Yu M. Y., Seo H. J.(2012). Luminescence properties and crystallographic sites for Eu3+ ions in fluorthalenite Y3Si3O10F, *Journal of Alloys and Compounds 536, 47-51 pages.*

[14] C. Talamonti , K. Kanxheri , S. Pallotta and L. Servoli .(2021). Diamond Detectors for Radiotherapy X-Ray Small Beam Dosimetry, https://doi.org/10.3389/fphy.2021.632299

[15] Akselrod M. S., Akselrod A. E. (2006). NEW Al2O3:C,Mg CRYSTALS FOR RADIOPHOTOLUMINESCENT DOSIMETRY AND OPTICAL IMAGING, doi:10.1093/rpd/nci663.

[16] Di J., Xu .X., Xia C., Zeng H., Cheng Y., Li D., Zhou D., Wu F., Cheng J., Xu J.(2012). Crystal growth and optical properties of Sm:CaNb2O6 single crystal, *Journal of Alloys and Compounds 536* Pages 20-25.

[17] Begum M., Mizanur Rahman A.K.M., Abdul- Rashid H.A., Yusoff Z., Siti Nurasiah Mat Nawi., Mayeen Uddin Khandaker., Bradley D.A.,(2021). Photonic crystal fibre as a potential medium for radiotherapy dosimetry. *Applied Radiation and Isotopes,* 174 109771.

[18] J. I. Lee, J. S. Yang, J. L. Kim, A. S. Pradhan, J. D. Lee, K. S. Chung, and H. S. Choe,(2006).Dosimetric characteristics of LiF:Mg,Cu,Si thermoluminescent materials, :*Applied Physics Letters 89, 094110* ; doi: 10.1063/1.2345280.

[19] M. Menichelli, M. Boscardin, M. Crivellari, J. Davis, S. Dunand, L. Fanò,a A. Morozzi, F. Moscatelli, M. Movileanu-Ionica, D. Passeri, M. Petasecca, M. Piccini, A. Rossi, A. Scorzoni, L. Servoli, G. Verzellesib and N. Wyrsch, (2020). Hydrogenated amorphous silicon detectors for particle detection, beam flux monitoring and dosimetry in high-dose radiation environment.

[20] M. Menichelli, , M.Bizzarri, M.Boscardin, M. Caprai , A.P. Caricato , G.A.P. Cirrone, M. Crivellari, I. Cuppar7, G. Cuttone , S.Dunand, L. Fanò , O. Hammad , M. Ionica , K. Kanxheri , M. Large, G. Maruccio , A.G. Monteduro, A. Morozzi , F. Moscatelli, A. Papi , D. Passeri, M. Petasecca, G. Petringa , G.Quarta , S. Rizzato , A. Rossi1, G. Rossi , A. Scorzoni1, L. Servoli , C. Talamonti, G. Verzellesi, N. Wyrsch, (2021). Testing of planar hydrogenated amorphous silicon sensors with charge selective contacts for the construction of 3D- detectors.

[21] Khandaker M. U., Nawi S.N. M., Bradley D. A., Lam S.E., Sani S.F.A., Sulieman A. (2021). Studies of thermoluminescence kinetic parameters of polymer pencil lead graphite under photon exposure, *Applied radiation and isotopes* 174-109757.

[22] Faria L.O., Lo D., Kui H. W., Khaidukov N.M. and Nogueira M. S. (2004). Thermoluminescence Response of K_2YB_5:TB^{+3} Crystals to Photon Radiation Fields, Radiation Protection Dosimetry, Vol.112, No.3, pages 435-438.

[23] Prado-Herrero,P.,.Garcia-Guinea, J.,Crespo-Feo, E. and Correcher, V.(2010) 'Temperature-induced transformation of metavariscite to berlinite', Phase Transitions, 83: 6, 440 — 449. Temperature-induced transformation of metavariscite to berlinite: Phase Transitions: Vol 83, No 6 (tandfonline.com).

[24] El-Nashar HF, El-Kinawy M, El- Faramawy N.(2019). Investigations of the kinetic energy parameters of irradiated (La)-doped phosphate glass. *Luminescence;1–7*. https://doi.org/10.1002/bio.3703.

[25] V. Guckan, S.W. Bokhari, V. Altunal, A. Ozdemir, W. Gao, Z. Yegingil.(2021). Luminescence of Ce3+ and Li+ co-doped MgO synthesized using solid-state reaction method, *Nuclear Instruments and Methods in Physics Research B 503* pages 53-61.

[26] De Lima Jr M.M., Freire Jr F.L. and Marques F.C.(2002). Boron Doping of Hydrogenated Amorphous Silicon Prepared by rf-co sputtering, *Brazilian journal of Physics,* pages 379-382.

[27] N. Kishimoto et al.(1998), "Radiation resistance of amorphous silicon in optoelectric properties under proton bombardment", *Journal of nuclear materials*, 258-263, pp. 1908-1913.

[28] V. Perez Mendez et al.(1988). "Hydrogenated amorphous silicon pixel detectors for minimum ionizing particles", *Nucl. Instr and Meth*. A, 273, p. 127.

[29] M. Despeisse, G. Anelli, P. Jarron, J. Kaplon, D. Moraes, A. Nardulli, F. Powolny, N. Wyrsch. (2008). Hydrogenated Amorphous Silicon Sensor Deposited on Integrated Circuit for Radiation Detection.

[30] Sano K.(,2021). Measurement of magnetic susceptibility of obsidian from Shirataki, Hokkaido, Japan, to identify the source of obsidian tools, *Jour. RRM, Univ. Hyogo, no.1,* p. 35-43.

[31] Goksu H.Y., Turetken N.(1979).Source Identification of Obsidian Tools by Thermo_luminescence. *PACT Journal 3*, 356-359.

[32] Patrícia L. Antonioa, , Raquel A.P. Oliveirab, Helen J. Khouryc, Linda V.E. Caldas, (2019). Evaluation of the thermally and optically stimulated response of an Italian Obsidian irradiated in 60Co beams, *Radiation Physics and Chemistry*, pages 115-120.

[33] J.S.Alzahrani, C.Soliman, D.A.Aloraini A.A. Alzahrany.(2016). Phototransferred Thermoluminescence from Obsidian Using Ultraviolet Radiation, *Journal of Natural Sciences Research*, Vol.6, No.16, pages 53-59.

[34] Woodley, S., Catlow, R.(2008). Crystal structure prediction from first principles. *NatureMater* 7, 937–946 .https://doi.org/10.1038/nmat2321.

[35] General Chemistry Notes https://chem.libretexts.org/Bookshelves/General_Chemistry

[36] Atif R., Khaliq J., Combrinck M., Hassanin A. H., Shehata H., Elnabawy E., Shyha I.(2020). Solution Blow Spinning of Polyvinylidene FluorideBased Fibers for Energy Harvesting Applications:A Review, doi:10.3390/polym12061304, 1-29 pages.

[37] Aydoğan Ş. (2011).Katı Hal Fiziği, *Nobel Bilim ve Araştırma Merkezi yayın* No: 1620 ISBN 978-605-395-431-6.

[38] Pitteri M., Zanzotto G. (1996) On the definition and classification of Bravais lattices.

[39]Kashif Q.Q. (2018). Crystal System- Bravais Lattices, http://dx.doi.org/10.13140/RG.2.2.11512.14083.

[40] Burzlaff H., Grimmer H., Gruber B., De Walff P.M. and Zimmermann H. (2016). International Tables of Crystallography, Vol A, Chapter 3.1, pages 698-719.

[41] Stachurski Z.H. (2011), On Structure and Properties of Amorphous Materials, Materials 2011, 4, 1564-1598; doi:10.3390/ma4091564.

[42] Engel P. (2002). Geometric Crystallography. *Springer Kluwer Science Publishers*.

[43] Cheng, Y.T.; Johnson, W.L (1987). Disordered materials: *A survey of amorphous solids. Science* 235, 997-1002.

[44] Bernal, J.D.(1959). A geometrical approach to the structure of liquids. *Nature, 4665*, 141-147.

[45] Torquato, S.(2002).Random Heterogenous Materials; *Springer-Verlag*: New York, NY, USA.

[46] Hui, X.; Fang, H.Z.; Chen, G.L.; Shang, S.L.; Wang, Y.; Qin, J.Y.; Liu, Z.K. (2009). Atomic structure of Zr41.2Ti13.8Cu12.5Ni10Be22.*5 bulk metallic glass alloy. Acta Mater*., 57, 376-391.

[47] To, L.-Th.; Daley, D.; Stachurski, Z.H. (2006).On the definition of an ideal amorphous solid of uniform hard spheres. *Solid State Sci., 8*, 868-896).

[48] Adams, D.J.; Mathesons, A.J.(1972), Computation of dense random packings of hard spheres. J. *Chem. Phys. 56*, 1989-1994.

[49] Finney, J.L. (1970).Random Packings and the structure of simple liquids. I. The geometry of random close packing. *Proc. R. Soc. Lond. A , 319*, 479-493.

[50] Clarke, A.S.; Jonsson, H., (1993). Structural changes accompanying densification of random hard-sphere ´ packings*. Phys. Rev. E, 47/6,* 3975-3984.

[51] Donev, A.; Torquato, S.; Stillinger, F.H. (2005). Pair correlation function characteristics of nearly jammed disordered and ordered hard-sphere packings. *Phys. Rev. E, 71,* 0501-0514.

[52] Conway, J.N.; Sloan, N.J.A,(1998). Sphere Packings, Lattices and Groups; *Springer-Verlag*: New York, NY, USA.

[53] Torquato, S.(2002).Random Heterogenous Materials; *Springer-Verlag:* New York, NY, USA.

[54] Stoyan, D.; Kendall, W.S.; Mecke, J.(1995). Stochastic Geometry and Its Applications; Wiley: Chichester, UK.

[55] Zaccarelli, E.; Foffi, G.; Sciortino, F.; Tartaglia, P.; Dawson, K.A. (2001),Gaussian density fluctuations and mode coupling theory for supercooled liquids. *Europhys. Lett , 55*, 157-163.

[56] Stachurski Z.H. and Welberry T.R.(2011).Metall. and Mater. *Trans. A42,* p.21-29.

[57] Stachurski Z.H. (2003). *Phys. Rev. Lett., 90/15*, 155502-1.

[58] B.Subashini, Mrs.Geetha. (2017). Introduction to Crystal Growth Techniques, *International Journal of Engineering and Techniques - Volume 3 Issue 5*, ISSN: 2395-1303.

[59] Ye N., Wang J.Y., Boughton R.I., and Hong M.-C. (2017). Functional Crystals, *chapter 20*, pages 575-611.

[60] J. Czochralski, Z. (1918., *Phys. Chem*. *92*, 219.

[61] Talik E., Oboz M. (2013). Czochralsi Method for Crystal Growth of Reactive Intermetallics, *124 (2).*

[62] Kouwenberg, J., 2018, Fluorescent Nuclear Track Detectors for Alpha Particle Measurement, DOI: 10.4233/uuid:6b7b09bf-5eb0-4a2e-bd93-a3bfff3a2772.

[63] Venkataraman, chapter 4 semiconductor detectors, *Handbook of Radioactivity Analyisis* pages 409-491, 2020. https://doi.org/10.1016/B978-0-12-814397-1.00004-2.

[64] Antonio P. L., Oliviera R.A.P, Khoury H.J., Caldas L. V.E., 2019,Evaluation of the thermally and optically stimulated response of an Italian Obsidian irradiated in 60Co beams, *Radiaiton Physics and Chemistry 155,* p. 115-120.

[65] Sabah M. Mohammad et al. (2020). J. *Phys.: Conf. Ser. 1535 012009*.

[66] Chandra P. Khattak, Frederick Schmid (2001). Growth of the world's largest sapphire crystals,*Journal of Crystal Growth*,Volume *225, Issues 2–4*,Pages 572-579, ISSN 0022-0248,https://doi.org/10.1016/S0022-0248(01)009551. https://www.sciencedirect.com/science/article/pii/S0022024801009551.