



Derin Sinir Ağları Kullanarak Diyabet Hastalığının Teşhisi

Diagnosis of Diabete mellitus Using Deep Neural Network

Ömer DEPERLİOĞLU
Bilgisayar Teknolojileri Bölümü,
Afyon Kocatepe Üniversitesi,
Afyonkarahisar, Türkiye
deperlioglu@aku.edu.tr

Utku KÖSE
Bilgisayar Mühendisliği Bölümü,
Süleyman Demirel Üniversitesi,
Isparta, Türkiye
utkukose@sdu.edu.tr

Özetçe—Diyabet hastalığının tespiti çalışmalarının temelini klinik karar destek sistemlerinin altyapısını oluşturan sınıflandırma çalışmaları oluşturmaktadır. Sınıflandırma çalışmalarındaki temel amaç sınıflandırma başarımını artırmak, teşhis oranını yükseltmektir. Bunun için çok farklı sınıflandırma yöntemleri ve farklı optimizasyon algoritmaları kullanılmaktadır. Bu bağlamda, bu çalışmada, diyabet hastalığının teşhisi için, Oto Kodlayıcı Sinir Ağları (OKSA) ile sınıflandırma çalışması yapılmıştır. Sınıflandırma çalışmasında yaygın olarak kullanılan UCI makine öğrenme laboratuvarındaki Pima Indian diyabet veri setinden yararlanılmıştır. Çalışmanın sonuçları diyabet hastalığının teşhisine odaklanan ve UCI'nin aynı makine öğrenme veri setini kullanan önceki çalışmaların sonuçları ile karşılaştırılmıştır. Elde edilen sınıflandırma doğruluğu %97,3'dür ve daha önce belirtilen sınıflandırma yöntemlerine kıyasla daha yüksektir. Elde edilen değerlendirmeler, önerilen yöntemin çok verimli olduğunu ve sınıflandırma başarımını artırdığını göstermiştir.

Anahtar Kelimeler — *Pima Yerlileri diyabet veriseti; Sınıflandırma; Oto kodlayıcı sinir ağları; Derin Öğrenme; Derin sinir ağı.*

Abstract— The basis for the determination of diabetes mellitus is the classification studies that constitute the infrastructure of clinical decision support systems. The main purpose of classification studies is to increase the classification performance and increase the diagnostic rate. Different classification methods and different optimization algorithms are used for this. In this context, in this study, a classification study with Autoencoder deep neural networks was performed for the diagnosis of diabetes mellitus. The Pima Indian diabetes dataset in the UCI machine learning laboratory, which is widely used in the classification study, was used. The results of the study were compared with the results of previous which focuses on the diagnosis of diabetes

studies using the same UCI machine learning dataset. The obtained classification accuracy is 97.3% and higher than the previously mentioned classification methods. The obtained evaluations show that the proposed method is very efficient and increases the classification success.

Keywords — *Pima Indian diabetes dataset; Classification; Autoencoder Neural Networks; Deep learning; Deep Neural Network.*

I. GİRİŞ

Diyabet hastalığı dünyadaki yaygın ve hızla artan hastalıklardan biridir ve doğal olarak dünyada önemli bir sağlık problemidir. Bu nedenle Dünya Sağlık Örgütü de (DSÖ-WHO) diyabet konusunda çalışmalar yapmaktadır. DSÖ 1965 yılından bu yana diyabetin teşhisi ve tedavisi için rehberler yayınlamaktadır. Diyabetin ve onun komplikasyonlarının tanımlanması, tanısı ve sınıflandırılması için hazırlanan rehberin son sürümü 1998 yılında yayınlanmıştır. Diyabet, insan vücudunun vücuttaki şeker miktarını düzenlemek için gereken miktarda insülini üretmediği veya üretilen insülinin yeterince kullanılmadığı bir durumdur. Aynı zamanda diyabet, kalp hastalığı, böbrek hastalığı, körlük, sinir hasarı ve kan damarı hasarı gibi çeşitli hastalıklara neden olabilmektedir [1-5].

Diyabetin erken teşhisi, oluşturabileceği diğer hastalıkların önüne geçmek adına çok önemlidir [6]. Veri madenciliği teknikleri, biyomedikal verileri analiz etmek için biyo-enformatikte yaygın bir şekilde uygulanmaktadır. Han ve arkadaşları, Pima Indian Diyabet Verileri Kümesini analiz etmek için RapidMiner'ı seçerek veri madenciliği işlemini gerçekleştirmişlerdir. Temel olarak, öznitelik tanımlama ve seçimi, aykırı verileri kaldırma, veri



normalleştirme ve sayısal ayırıklaştırma, görsel veri analizi, gizli ilişkiler keşfi ve diyabet tahmin modeli oluşturma da dahil olmak üzere veri ön işleme konularına odaklanmışlardır [7]. Tıbbi veri tabanlarından bilgi keşfi, etkili tıbbi teşhis yapmak için önemlidir. Veri madenciliğinin amacı, verileri veri tabanından çıkarmak ve açık ve anlaşılır açıklamalar yapmaktır. Veri madenciliği teknikleriyle yapılan başka bir çalışmada, sayısal veriler üzerinde ilişki kuralları üretmek için yeni bir yaklaşım sunulmuştur. Devamlı değerli öznitelikleri ayırık hale getirmek için değiştirilmiş bir eşit genişlik ayırma aralığı yaklaşımı önerilmiştir. İstenilen aralıkların yaklaşık genişliği, tıp uzmanının görüşüne dayalı olarak seçilmiş ve modelin bir girdi parametresi olması sağlanmıştır [8]. AlJarullah çalışmasında, Pima Kızılderilleri Diyabet Veri Setini kullanarak, diyabet hastalarını erken teşhis için karar ağacı yöntemi kullanmıştır. Çalışma iki aşamadan oluşmaktadır. Birinci aşamada, öznitelik tanımlama ve seçimi, eksik değerleri işleme ve sayısal ayırıklaştırma da dahil olmak üzere veri ön işleme tabi tutulmaktadır. İkinci aşamada, karar ağacı yöntemini kullanarak bir diyabet öngörme modeli oluşturulmuştur [9].

Diyabet tespiti üzerine yapılan sınıflandırma çalışmalarında daha çok bulanık mantık, yapay sinir ağları ve destek vektör makineleri kullanılmıştır. Jaganathan Kuppuchamy çalışmalarında, tıbbi veri tabanlarının sınıflandırması için bulanık entropi temelli özellik uygunluk ölçümü sunmaktadırlar. Önerdikleri yöntemi test etmek için Radyal Temel İşlevli Ağ sınıflandırıcısı ve Eşik Seçimi için Sinir Ağı (ESSA) kullanmışlardır [10]. Lekkas, ve Mikhailov makalelerinde, bulanık kural tabanını, veri akışlarından örnek bazında yinelemeli olarak düzenleyerek çevrimiçi moda veri işlenmesine olanak tanıyan bulanık sınıflandırmanın gelişen bir yöntemini gözden geçirmektedir. Buna ek olarak, bu yöntemin iyileştirilebileceğini ve iki popüler tıbbi problem için teşhis alanına nasıl uygulanabileceğini göstermektedirler [11]. Yapay sinir ağları ile yapılan örneklere bakıldığında; Temurtaş vd. çalışmalarında, Levenberg-Marquardt öğrenme algoritması ile eğitilmiş çok katmanlı sinir ağı ve olasılıksal sinir ağı kullanarak sınıflandırma çalışmaları yapmışlardır. Çalışmalarının sonuçlarını, diyabet hastalığı tanısına odaklanan ve aynı UCI makine öğrenme diyabet veri setini kullanan geçmiş çalışmaların sonuçları ile karşılaştırmışlardır [12].

Bulanık mantık ve yapay sinir ağlarının birlikte kullanıldığı çalışmalarda mevcuttur. Kahramanlı ve Allahverdi çalışmalarında Yapay sinir ağı ve Bulanık sinir ağı içeren karma bir sinir ağı geliştirmişlerdir. Yöntemlerinin uygulanabilirliğini göstermek için California Üniversitesi Irvine (UCI) makine öğrenme veri tabanından elde edilen diyabet ve kalp veri setlerini kullanmışlardır. Önerilen yöntemin doğruluğunu değerlendirmek için tıbbi sınıflandırma çalışmalarında sıklıkla kullanılan doğruluk, duyarlılık ve özgünlük performans ölçütlerini kullanmışlardır [13]. Bir başka çalışmada Polat ve Güneş, Ana Bileşen Analizi (ABA) ve

Uyarlamalı sinirsel bulanık çıkarım sistemi (ANFIS) kullanarak diyabet hastalığını teşhis etmeye çalışmışlardır. Çalışmalarında, ABA ve ANFIS'i birleştirerek diyabet hastalığını tanımlama oranını iyileştirmeye çalışmışlardır. Önerilen sistem de iki aşama bulunmaktadır. İlk aşamada, 8 özellikli diyabet hastalığı veri kümesinin boyutu, ana bileşen analizi kullanılarak 4 özelliğe indirgenmiştir. İkinci aşamada, diyabet hastalığının teşhisi uyarlamalı sinirsel bulanık çıkarım sistemi (ANFIS) ile sınıflandırma yapılmıştır [14].

Destek vektör makineleri ile yapılan sınıflandırma çalışmalarına örnek verecek olursak, Kumari ve Chitra radyal temelli çekirdek kullanarak destek vektör makineleri ile yaptıkları sınıflandırma da %78 sınıflandırma doğruluk oranı elde etmişlerdir [5]. Köse vd. hem Destek Vektör Makineleri hem de Bilişsel Gelişim Optimizasyon Algoritması vasıtasıyla oluşturulmuş bir diyabet teşhis sistemi önermişlerdir. Destek vektör makinelerinin eğitimi boyunca, Gauss çekirdek işlevinin sigma parametresini belirlemek için Bilişsel Gelişim Optimizasyon Algoritması kullanılmış ve sonuçta sınıflandırma işlemi yapılmıştır [15]. Benzer şekilde başka bir çalışmada Destek Vektör Makineleri tabanlı bir diyabet teşhis sistemi önerilmiştir. Destek Vektör Makinelerinin eğitimi sırasında, Gauss çekirdek fonksiyonunun sigma parametresinin belirlenmesi için Girdap Optimizasyon Algoritması kullanılmış ve Pima Indian diyabet veri seti için sınıflandırma işlemi gerçekleştirmişlerdir [16].

Son zamanlarda yaygın olarak kullanılmaya başlayan *Derin Öğrenme*, birbirleriyle iletişim kuran çok sayıda soyut katmanın bulunduğu gelişmiş bir makine öğrenme tekniğidir. Her katman önceki katmana derinden bağlıdır ve kararlarını önceki katman tarafından beslenen çıktıya dayanarak yapar. Başka bir deyişle *Derin Öğrenme*, Yapay Zekada yapısal olmayan veya etiketlenmemiş verilerin denetimsizliğini öğrenebilen ağlara sahip makine öğrenmesinin bir alt kümesidir. Derin Öğrenme veya Derin Sinir Ağı olarak da bilinmektedir [17].

Bu çalışmada Pima Indian Diyabet veri seti ve OKSA kullanarak diyabet hastalığının tespiti üzerinde durulacaktır. OKSA kullanılarak, herhangi bir optimizasyon işlemine veya karma yöntemlere gerek kalmadan, kolayca sınıflandırma başarısının artırılabilirliğini göstermektedir.

II. MATERYAL METOT

Bu çalışmada, Pima Indian Diabet veriseti kullanılarak OKSA ile diyabet teşhisi için sınıflandırma çalışması yapılmıştır. Sınıflandırma çalışmasının aşamaları sırasıyla aşağıda verilmiştir.

A. Pima Indian Diyabet Veriseti

Sınıflandırma çalışmasında, web sayfası son olarak 28 Şubat 2011 tarihinde güncellenmiş olan, UCI, Kaliforniya Üniversitesi, Irvine Makine öğrenmesi veri tabanındaki Pima Indian diyabet veri seti kullanılmıştır [18, 19]. Veri

setindeki örnekler için tüm hastalar en az 21 yaşında Pima Indian kökenli kadınlardan seçilmiştir. Veri setinde toplam 768 örnek bulunmaktadır ve her bir örnek için 8 nitelik mevcuttur. Bu nitelikler Tablo I'de verilmiştir. Veri setindeki veriler, sağlıklı-diyabet negatif ve diyabetli-diyabet pozitif olmak üzere 2 gruptan oluşmaktadır.

TABLO I. PİMA İNDİAN DİYABET VERİ SETİNDEKİ NİTELİKLER

Nitelik No.	Niteliğin Açıklaması
1	Hamile Kalma Sayısı
2	Plazma Glukoz Konsantrasyonu
3	Diastolik Kan Basıncı (mm Hg)
4	Kol Kası Cilt Kıvrım Kalınlığı (mm)
5	2-saat serum insülin (mu U/ml)
6	Vücut Kütle İndeksi (kg/(boy(m)) ²)
7	Diyabet Soyağacı Fonksiyonu
8	Yaş (Yıl)
9	Sınıf Değişkeni (sağlıklı 0 veya diyabetli 1)

B. Sınıflandırma

Bu çalışmada sağlıklı-diyabet negatif ve diyabetli-diyabet pozitif olmak üzere 2 gruptan oluşan veriler Oto Kodlayıcı Sinir Ağları (OKSA) ile sınıflandırılmıştır.

1) Oto Kodlayıcı Sinir Ağları

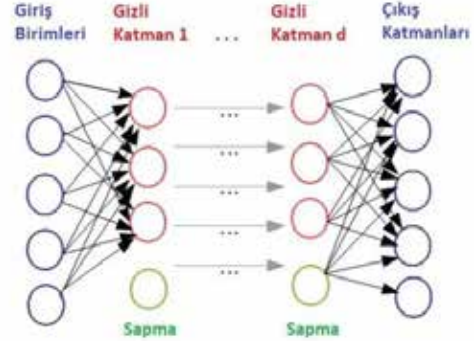
Derin sinir ağlarının veya derin öğrenmenin bir çeşidi olan Oto Kodlayıcı Sinir Ağlarından, ilk olarak 1990'lı yıllarda bahsedilmiş 2000'li yıllarda da yaygınlaşmıştır. OKSA temelde üç ana katmana sahip bir sinir ağıdır. Bunlar: bir giriş katmanı, gizli (kodlama) katmanları ve bir de çıkış (kod çözme) katmanıdır. Şekil 1'de d kadar gizli katmana sahip olan bir OKSA mimarisi verilmiştir. OKSA, girdilerini yeniden yapılandırmak için eğitilir ve gizli katmanı, girdilerin iyi temsillerini öğrenmeye zorlar. OKSA, hedef değerlerini girdilere eşit olacak şekilde ayarlayan, geri yayılımı uygulayan denetimsiz bir Makine öğrenme algoritmasıdır. OKSA, girişini çıktısına kopyalamaya çalışacak şekilde eğitilir. İçerisinde, girdiyi temsil etmek için kullanılan bir kodu açıklayan gizli bir katman vardır [17, 20]. OKSA ağındaki gizli katman, ağı veriyi düşük boyutlu bir gösterime sıkıştırması için zorlayan, bir bilgi darboğazı oluşturur. Verilerin bu modellemesi, tahminler yapmak için kullanılabilir.

Şekil 1'deki d gizli katmana sahip basit bir OKSA için, gizli birim etkinliklerinin vektörü, h (1)'deki gibi hesaplanabilir:

$$h = f(W_d \cdot a + sapma_d) \quad (1)$$

Burada f aktivasyon fonksiyonudur. W_d bir parametre matrisi ve $sapma_d$ ise önyargı parametrelerinin bir vektörüdür. Bu çalışmada kodlama katmanında Ölçekli Eşlenik Eğitim Algoritması (ÖEEA-Scaled Conjugate Gradient Algorithm) ve çapraz entropi maliyet (Cross entropy cost) fonksiyonu kullanılmıştır.

Verilerin gizli gösterimi, daha sonra (2)'deki gibi kod çözme işlevini kullanarak bir a alanına eşlenir:



Şekil 1. d adet gizli katmanlı bir OKSA mimarisi.

$$\hat{a} = f(W_d \cdot h + sapma_d) \quad (2)$$

Burada W_d , kod çözme matrisidir ve önyargı parametrelerinin bir vektörünü ön plana çıkarır. OKSA parametrelerini, yeniden yapılandırma Levenberg-Marquardt algoritmasını ve kod çözme hatasını en aza indirmek için a ve \hat{a} arasındaki ortalama karesel hata (OKH-mean square error-MSE) yöntemini kullanarak gerçekleştirir. OKH (3)'teki gibi elde edilebilir:

$$OKH(a, \hat{a}) = \|a - \hat{a}\|_2^2 = \|a - (W_d \cdot h + sapma_d)\|_2^2 \quad (3)$$

Gizli katmanın a 'ya göre daha az boyutu varsa, OKSA eğitim verisinin sıkıştırılmış bir temsili öğrenir. Aslında, k doğrusal gizli birimlere sahip bir OKSA, veriyi ilk k ana bileşenlerine yansıtmayı öğrenecektir. Doğrusal olmayan gizli birimler, ek gizli katmanlarda olduğu gibi, bir OKSA'nın daha karmaşık kodlama işlevleri öğrenmesine izin verir [20].

III. SINIFLANDIRMA

OKSA ile yapılan sınıflandırma çalışmasında ağda toplam 24 gizli katman bulunmaktadır. Kodlama aşamasında öğrenme algoritması olarak Ölçekli Eşlenik Eğitim Algoritması ve performans fonksiyonu olarak da çapraz entropi maliyet fonksiyonu kullanılmıştır. Kod çözme aşamasında ise Levenberg-Marquardt algoritması ve ortalama karesel hata yöntemi kullanılmıştır.

Pima Indian diyabet veri seti ile yapılan çok sayıda denemelerin sonucunda OKSA'nın parametreleri elde edilmiştir. Elde edilen OKSA ile veri setinin %80'i eğitim verisi ve %20'si test verisi olacak şekilde sınıflandırma gerçekleştirilmiştir. Farklı eğitim ve test verileri ile sınıflandırma işlemi 20 kez tekrarlanmıştır. Bu çalışmalarda elde edilen Doğruluk (Accuracy) oranlarının ortalaması alınmıştır. Yapılan 20 deneme sonunda eğitim ve test çalışmaları birlikte olmak üzere ortalama doğruluk oranı 97,3 olarak elde edilmiştir.

IV. SONUÇ VE ÖNERİLER

OKSA ile yapılan sınıflandırma çalışmasında, sınıflandırma başarısını karşılaştırmak için geçmişte Pima Indian diyabet veri seti kullanılarak, farklı yöntemlerle



yapılan sınıflandırma işlemlerinde elde edilen *Doğruluk* oranları karşılaştırılmıştır. Sınıflandırma *Doğruluk* oranları Tablo II 'de verilmiştir.

Tabloda 1. sırada bu çalışmada elde edilen doğruluk oranı, 2. sırada Eşik Seçimi için Sinir Ağ (ESSA) kullanılmışlardır [6], 3. sırada Gelişen bulanık kural tabanlı sınıflandırma (EBKTS) [7], 4.sırada olasılıksal yapay sinir ağı (OYSA) [8], 5. sırada Levenberg-Marquardt algoritması (LM-YSA) [8], 6. sırada Yapay sinir ağı (YSA) ve bulanık sinir ağı (BSA) [9], 7. sırada, Ana Bileşen Analizi (ABA) ve Uyarlamalı sinirsel bulanık çıkarım sistemi (ANFIS) [10], 8. sırada, Radyal temelli çekirdek fonksiyonu ile destek vektör makinesi (RT-DVM) [5], 9. sırada DVM ve Bilişsel Gelişim Optimizasyon Algoritması (BGOA) [11], 10.sırada DVM ve Girdap Optimizasyon Algoritması (GOA) kullanılmıştır [12].

TABLO II. SINIFLANDIRMA DOĞRULUK ORANLARI

Sıra No	Sınıflandırma Çalışması	Doğruluk Oranı
1	OKSA (Bu çalışma)	%97,30
2	ESSA [6]	%76,04
3	EBKTS [7]	%79,37
4	OSA [8]	%78,13
5	LM-YSA [8]	%82,37
6	YSA ve BSA [9]	%84,20
7	ABA ve ANFIS [10]	%89,47
8	RT-DVM [5]	%78,00
9	DVM ve BGOA [11]	%87,50
10	DVM ve GOA [12]	%92,70

Tablodan da görüldüğü gibi tabloda koyu renkle belirtilen ve OKSA ile yapılan sınıflandırma çalışması en yüksek sınıflandırma oranına sahiptir. Bu da OKSA' nın başarısını ortaya koymaktadır. Buradan amacımızın gerçekleştiği ve OKSA kullanılarak, herhangi bir optimizasyon işlemine veya karma yöntemlere gerek kalmadan, kolayca sınıflandırma başarısının artırılacağı ortaya çıkmaktadır.

Burada örnek olarak sadece diyabet veri seti için sınıflandırma başarımı verilmiştir. Fakat OKSA' nın bu güçlü yapısı ile diğer veri setlerinde başarılı olacağı düşünülmektedir. Aynı yapı kalp hastalıkları ve göğüs kanseri gibi çok sık kullanılan diğer veri setleri içinde denenebilir.

KAYNAKÇA

- [1] World Health Organization and International Diabetes Federation "Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia", Report of World Health Organization and International Diabetes Federation, 2005.
- [2] International Diabetes Federation, Diabetes atlas (6th ed.). Brussels, Belgium, 2013 https://www.idf.org/sites/default/files/EN_6E_Atlas_Full_0.pdf, Son erişim 23 Ocak 2018.
- [3] International Diabetes Federation, Diabetes atlas (4th ed.). Brussels, Belgium, 2009 <https://www.idf.org/sites/default/files/IDF-Diabetes-Atlas-4th-edition.pdf>. Son erişim 23 Ocak 2018.

- [4] Veresiu, A.I., Bondor C. I., Florea B., Vinik E.J., Vinik A.I. ve Găvan N.A., Detection of undisclosed neuropathy and assessment of its impact on quality of life: a survey in 25,000 Romanian patients with diabetes, Journal of Diabetes and Its Complications, vol. 29, pp. 644–649, 2015.
- [5] Kumari V. A. ve Chitra R., Classification Of Diabetes Disease Using Support Vector Machine", International Journal of Engineering Research and Applications (IJERA). vol. 3, issue: 2, pp. 1797-1801, 2013.
- [6] Deperioglu O., The Effects of Different Training Algorithms on the Classification of Medical Databases Using Artificial Neural Networks, European Conference on Science, Art & Culture ECSAC 2018, Antalya, Turkey.
- [7] Han J., Rodriguze J. C. ve Beheshti M., Diabetes Data Analysis and Prediction Model Discovery Using RapidMiner, Second International Conference on Future Generation Communication and Networking, pp. 97-99, 2008.
- [8] Ganji M. F., ve Abadeh, M. S. "Using fuzzy Ant Colony Optimization for Diagnosis of Diabetes Disease", Proceedings of ICEE 2010, May 11-13, 2010.
- [9] AlJarullah, A. A. "Decision Tree Discovery for the Diagnosis of Type II Diabetes", International Conference on Innovations in Information Technologies, 303-307, 2011.
- [10] Jaganathan P., Kuppuchamy R., A threshold fuzzy entropy based feature selection for medical database classification, Computers in Biology and Medicine, vol. 43, 2013, pp. 2222–2229.
- [11] Lekkas S. ve Mikhailov L., Evolving fuzzy medical diagnosis of Pima Indians diabetes and of dermatological diseases. Artificial Intelligence in Medicine vol. 50, 2010, pp. 117–126.
- [12] Temurtas H., Yumusak N. ve Temurtas F., A comparative study on diabetes disease diagnosis using neural networks, Expert Systems with Applications, vol. 36, 2009, pp. 8610–8615.
- [13] Kahramanli H. ve Allahverdi N., Design of a hybrid system for the diabetes and heart diseases, Expert Systems with Applications, vol. 35, 2008, pp. 82–89.
- [14] Polat K., Günes S., Arslan A., A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine, Expert Systems with Applications, vol. 34, 2008, pp. 482–487.
- [15] Kose U., Guraksin G. E., Deperioglu O., Cognitive Development Optimization Algorithm Based Support Vector Machines for Determining Diabetes, BRAIN. Broad Research in Artificial Intelligence and Neuroscience, vol. 7, no:1, 2016, pp. 80-90.
- [16] Kose U., Guraksin G. E., Deperioglu O., Diabetes determination via vortex optimization algorithm based support vector machines. Medical Technologies National Conference (TIPTEKNO), pp. 1-4, 2015.
- [17] Chandrayan P., Deep Learning: Autoencoders Fundamentals and types, <https://codeburst.io/deep-learning-types-and-autoencoders-a40ee6754663>. Son erişim 25 Ocak 2018.
- [18] UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/datasets/pima+indians+diabetes>, Irvine, CA: University of California, School of Information and Computer Science, Son erişim 10 Kasım 2017.
- [19] Gurney K., An introduction to neural networks, Taylor & Francis e-Library, UCL Press Limited, London, 2004.
- [20] Chicco D., Sadowski P., Baldi P., Deep autoencoder neural networks for gene ontology annotation predictions, Proceeding of BCB '14, Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, pp. 533-540, 2014.