

Derin Öğrenme ile Sınıflandırma Üzerine Karşılaştırmalı Bir Çalışma A Comparative Study on Classification by Deep Learning

Abdullah CALISKAN¹, Hasan BADEM², Alper BASTURK², Mehmet Emin YUKSEL¹

¹Biyomedikal Mühendisliği Bölümü, Erciyes Üniversitesi, Kayseri, Türkiye
acaliskan@erciyes.edu.tr, yuksel@erciyes.edu.tr

²Bilgisayar Mühendisliği Bölümü, Erciyes Üniversitesi, Kayseri, Türkiye
hbadem@erciyes.edu.tr, ab@erciyes.edu.tr

Özet

Sınıflandırma problemlerinde öznetelik çıkarma yöntemlerinin kullanılması makina öğrenme algoritmalarının başarımını artırmaktadır. Derin sinir ağları eldeki veriden hiyerarşik bir şekilde otomatik öznetelik üretebilmesinden dolayı, bir çok uygulamada derin sinir ağlarının öznetelik çıkarma yöntemlerinin kullanılmasına ihtiyaç yoktur. Bu sebeple bu çalışmada ham veriden yeni öznetelikler çıkarma kabiliyeti olan özdevinimli kodlayıcılar ele alınmaktadır. Yığılanmış özdevinimli kodlayıcılar geleneksel ve alışılmış yöntemlerle karşılaştırıldığında, üzerinde çalışılan veri setlerinin sınıflandırmasında daha iyi sonuçlar ürettiği görülmüştür. Elde edilen sınıflandırma sonuçları yapılan istatistiksel analizlerle desteklenerek önerilen yöntemin karşılaştırma yapılan diğer yöntemlerden daha iyi sonuçlar ürettiği sonucuna varılmıştır.

Abstract

The use of feature extraction methods in classification problems improves the performance of machine learning algorithms. In most applications, however, there is no need for deep neural networks to employ such methods. This is because deep neural networks can automatically generate features from raw data in a hierarchical manner. For this reason, this study investigates autoencoder networks which are capable of extracting new features from raw data. Compared with the-state-art-methods, stack autoencoder has demonstrated more efficient performance in classification of the data sets used in this work. Classification results are also justified with statistical analysis which shows that the proposed method yields the best performance among the compared methods.

1. Giriş

Makine öğrenme yöntemlerinin performansı ele alınan verinin nasıl temsil edildiği ile doğrudan ilişkilidir. Bundan dolayı makina öğrenme yöntemlerinin etkinliğini ve sınıflandırma performansını arttırmak için eldeki veri içerisinde temsil oranı yüksek özneteliklerin çıkarılmasına ayrıca özen gösterilmektedir [1]. Bir başka deyişle veri içerisinde veriyi en iyi şekilde temsil edebilen öznetelikler elde edilebilirse daha başarılı sınıflandırma yapılabilir. Veriyi en iyi şekilde temsil eden özneteliklerin elde edilmesi son derece karmaşık ve zor bir süreci beraberinde getirmektedir.

Derin öğrenme yöntemleri veriyi ön işleme tabi tutmaksızın veri içerisinde öznetelikleri çıkarabilen bir yapıya sahip

olmasından dolayı son derece etkili bir yöntem olarak karşımıza çıkmaktadır [2]. Klasik yöntemlerle veri içerisinde öznetelik çıkarma işlemi son derece yorucu bir işlemdir. Bu işlemi otomatik olarak yapan derin öğrenme yöntemleri daha etkili sonuçlar ortaya koyabilmektedir. Derin öğrenme teknikleri insan beyninin çalışma mekanizmasını taklit etmeye çalışmaktadır [2].

Temsili öğrenmenin en iyi şekilde uygulandığı yapıların başında derin sinir ağlarından yığılanmış özdevinimli kodlayıcılar (YÖK) gelmektedir. YÖK'lerin temel birimi veriden yeni öznetelikleri otomatik olarak çıkarabilen özdevinimli kodlayıcılardır (ÖK). ÖK'ler girişine uygulanan veriyi çıkışında aynen üretmek üzere eğitilen ve tek gizli katmanı olan bir yapay sinir ağlarıdır. ÖK' nin amacı giriş verisini gizli katman üzerinden geçirerek aynen çıkışa aktarırken gizli katman çıkışında verinin alternatif bir temsiliyi üretmektir. Problemin niteliğine göre ihtiyaç duyulması durumunda birden fazla ÖK ardı ardına bağlanabilir [4].

Bu çalışmada bir YÖK yapısı çeşitli veri setleriyle test edilerek yapının başarımı istatistiksel olarak analiz edilmiştir. Çalışmada ÖK'lerin yeni temsiller üretme ve temsilleri softmax katmanı aracılığıyla sınıflandırmada kullanma etkinlikleri incelenmiştir. Sonuç olarak ÖK'lerin sınıflandırma başarımını kayda değer şekilde artırdığı görülmüştür.

Yapılan çalışmanın bundan sonraki planı şu şekildedir. Çalışmanın 2. bölümünde derin öğrenme yapılarından ÖK'nin yapısı hakkında bilgi verilecektir. Bunun yanında bu bölümde bir derin sinir ağının nasıl eğitildiği anlatılmıştır. Çalışmanın 3. bölümünde elde edilen simülasyon sonuçlarının değerlendirilmesi yapılarak, istatistiksel sonuçlar sunulmuştur. Son kısımda ise sonuç bölümü bulunmaktadır.

2. Derin Sinir Ağı

Derin sinir ağı, bir yığın oluşturacak şekilde organize edilen birçok basit yapıdan oluşmaktadır. Bu basit yapıların neredeyse tamamı doğrusal olmayan işlemler gerçekleştirerek, veri boyutunu değiştirerek veriyi farklı bir uzayda temsil ederek, veri içerisinde gizli kalmış özneteliklerin ortaya konulmasına yardımcı olur [2].

2.1. Özdevinimli Kodlayıcı

YÖK tabanlı derin sinir ağı, bir veya birden fazla ÖK'ten ve softmax adı verilen sınıflandırma biriminden oluşmaktadır. ÖK girişi ve çıkışı aynı olan iki ayrı kısımdan oluşan bir yapay sinir ağıdır. Bu yapay sinir ağındaki kısımlar, ÖK'nin girişi ile çıkışı arasında yer alan ağırlıklar tarafından temsil edilen kodlayıcı

ve ÖK'nin gizli katman çıkışı ile çıkış arasındaki ağırlıklarla temsil edilen çözücü kısımlarıdır. ÖK, çıkış etiketi olarak girişi baz alır ancak yapının temel amacı ÖK'nin gizli katman çıkışını elde etmektir. Gizli katman çıkışından elde edilen yeni örnekler girişi daha iyi temsil eden özniteliklere sahiptir [4, 5]. Eğer bir ÖK'nin gizli katman çıkışı başka bir ÖK'ye eğitim verisi olarak gönderilirse ikinci ÖK'de tıpkı birinci ÖK gibi yeni örneklemde elde edecektir. Bu örüntü arzu edildiği kadar devam ettirilebilir. Her birinin boyutu M olan girişi ve çıkış etiketi $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \dots \mathbf{x}^{(T)}$ olarak verilen bir ÖK'nin amaç fonksiyonu aşağıdaki gibi olur.

$$E_{sparse} = E_T + KL(\rho \parallel \hat{\rho}_j) \quad (1)$$

Yukarıda tanımlanan amaç fonksiyonu iki kısımdan oluşmaktadır. Bu kısımlardan ilki olan ve aşağıdaki gibi ifade edilen E_T tıpkı bir yapay sinir ağı modeli gibidir.

$$E_T = \frac{1}{T} \sum_{k=1}^T e_k^2 + \frac{\lambda}{2} \left(\sum_{i=1}^N \|\mathbf{w}_i\| + \sum_{i=1}^M \|\hat{\mathbf{w}}_i\| \right); \quad (2)$$

burada λ regularize terimidir aynı zamanda ağırlık ceza terimi olarak da bilinir. Bu terim denklemin ıraksamasını engeller.

Yukarıdaki denklemden yer alan ağırlık çıkışı ile arzu edilen çıkış arasındaki farkı ifade eden hata terimi e_k aşağıdaki şekilde ifade edilir. Denklemden yer alan $\mathbf{w}_i$ ve $\hat{\mathbf{w}}_i$ sırası ile kodlayıcı ve çözücü ağırlıklarıdır. Gizli katman boyutu ise N ile ifade edilmektedir.

$$e_k = \|\mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)}\| \quad (3)$$

Denklemden yer alan k ($k = 1 \dots T$) örnek sıra numarasını ifade eder.

Denklemin 1'de yer alan ikinci terim aşağıda yer alan denklemden hesaplanan $KL(\rho \parallel \hat{\rho}_j)$ Kullback-Leibler (KL) diverjansı olarak bilinir. Bu kısım denkleme seyreklik özelliği katarak ağda yer alan bir çok ağırlığın ya da bağlantının devre dışı bırakılmasında büyük rol oynar. Bu işin kontrolü dışarıdan bir kullanıcı tarafından girilen seyreklik ceza parametresi β ve seyreklik parametresi ρ aracılığı ile yapılmaktadır [3, 4, 5].

$$KL(\rho \parallel \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \quad (4)$$

Yukarıdaki denklemden yer alan $\hat{\rho}$ gizli katman çıkışının ortalama aktivasyon değerini ifade eder ve aşağıdaki gibi hesaplanır [4, 5].

$$\hat{\rho}_j = \frac{1}{T} \sum_{i=1}^T f_j(\mathbf{x}(i); \mathbf{W}, \mathbf{b}) \quad (5)$$

Burada f_j gizli katman çıkışında kullanılan doğrusal olmayan bir fonksiyondur. Bu fonksiyon genellikle sigmoid tipinde bir fonksiyondur bunun yanında $\mathbf{W}$ ve $\mathbf{b}$ çözücü ağırlıklarıdır.

2.2. Softmax

Derin sinir ağının son parçası sınıflandırma işlemi için kullanılacak olan softmax'dır [6]. Yapının bu katmanı iki sınıflı ayırabilen mantıksal regresyonun genişletilmiş halidir. Softmax direkt olarak giriş verisini sınıflandırma işlemi için kullanabilmektedir. Ancak bir ya da birden çok ÖK ile işlenmiş veriyi sınıflandırırken daha iyi sonuçlar elde ederek sınıflandırmanın doğruluğunu artırmaktadır.

2.3. Derin Sinir Ağı'nın Eğitimi

Eğitilecek olan derin sinir ağı ÖK katmanları ve softmax katmanından oluşmaktadır. Derin sinir ağının eğitimi ön öğrenme ve hassas öğrenme olmak üzere iki kısımdan oluşmaktadır. Ön öğrenme ile tüm yapıların ayrı ayrı eğitilmesi işlemi yapılırken hassas öğrenmede bu yapıların tamamı birleştirilerek bir kez daha eğitim işlemi yapılır. Tüm bunlar birlikte düşünüldüğünde ortaya karmaşık bir eğitim mekanizması çıkmaktadır. YÖK yapısının temel birimi ÖK'lardır. YÖK tabanlı bir derin sinir ağının eğitim aşamaları aşağıdaki gibi sıralanabilir [3, 4].

1. Şekil 1(a) da görüldüğü gibi birinci ÖK'ye giriş verisi gönderilir bu giriş verisi aynı zamanda çıkış etiketi olarak kullanılır. Ağı optimum ağırlıkları bölüm 2.1'de anlatıldığı gibi bulunur. Bu aşama tamamıyla danışman-sız öğrenme olarak değerlendirilir.
2. Şekil 1(b) de görüldüğü gibi birinci ÖK'nin gizli katman çıkışı ikinci ÖK'ye giriş olarak gönderilerek ikinci ÖK 1. adımdaki gibi eğitilir.
3. 2. adım bir sonraki ÖK için tekrar edilir ve bu işlem arzu edilen ÖK sayısı kadar tekrar edilir.
4. Şekil 1(c) de görüldüğü gibi son ÖK'nin gizli katman çıkışı softmax katmanına giriş olarak gönderilir. Softmax katmanı ağırlıklarını, hedeflenen çıkış ile kendi çıkışı arasındaki fark minimum olana kadar yeniler. Bu kısımda danışmanlı eğitim gerçekleştirilir. Buraya kadar gerçekleştirilmiş tüm eğitim aşamaları ön öğrenme olarak adlandırılır.
5. Şekil 1(d) de görüldüğü gibi tüm ÖK'ler gizli katman çıkışları birbirleriyle birleştirilerek YÖK yapısı oluşturulur bu yapının hemen ardına softmax katmanı eklenir. Bu yapının ağırlıklarının ilk değerleri ön öğrenme esnasında belirlendiğinden bu aşamada hassas öğrenme işlemi yapılarak ağırlıkların son değeri bulunur.

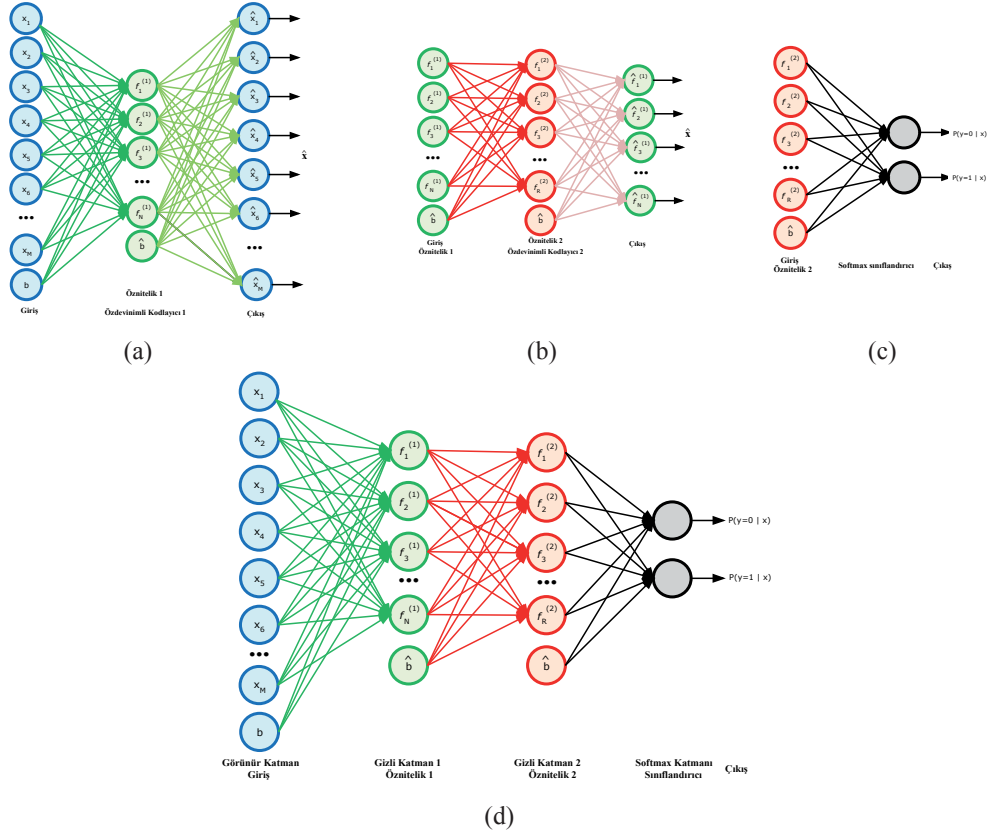
Tüm eğitim aşamalarında L-BFGS optimizasyon algoritması [7] kullanılmaktadır.

3. Deneysel Sonuçlar

Bu çalışmada, YÖK yapısının performansını test edebilmek için kullanılan veri setleri Data Mining Repository of University of California, Irvine (UCI) [8] dan alınmıştır. Çalışmada Arrhythmia, Breast Cancer ve Cleveland olmak üzere 3 farklı veri seti kullanılmıştır. Arrhythmia veri setinde 452 örneklem ve 279 öznitelik bulurken, Breast Cancer veri setinde 286 örneklem ve 9 öznitelik bulunmaktadır. Cleveland veri setinde ise 303 örneklem ve 13 öznitelik bulunmaktadır. Tasarlanan ağ yapısının, eğitim aşamasında kullanılan çeşitli parametreleri Çizelge 1'de görülmektedir. Bu parametreler ile elde edilen ağ yapısı Intel i7 2600 3.4 Ghz işlemcili, 12 GB DDR3 RAM bellekli sistemde, 10-kat çapraz doğrulama ile oluşturulmuş ve deneysel sonuçlar elde edilmiştir. Elde edilen deneysel sonuçlar literatürde daha önceden gerçekleştirilen yöntemler ile karşılaştırılmıştır.

3.1. Simülasyon Sonuçları ve İstatistiksel Analiz

YÖK yapısının sınıflandırma performansını gözlemleyebilmek için literatürde yaygın kabul gören *Destek Vektör Makinesi (SVM)*, *K En Yakın Komşu Algoritması (KNN)*, *Karar Ağacı (DT)* ve *Navie Bayes (NB)* yöntemleri ile karşılaştırılmıştır.



Şekil 1. Yığılanmış özdevimli kodlayıcı modeli

Çizelge 1. YÖK yapısında kullanılan parametreler ve değerleri

	Arrhythmia	Breast Cancer	Cleveland
Nöron Sayısı	4	4	4
ρ	0,2	0,1	0,01
β	5	5	5

Çizelge 2. YÖK yapısının her bir veri seti için sınıflandırma başarımlar dağılımı

	ED	EY	MED	ORT	SS
Arrhythmia	74,76	76,19	75,00	75,21	0,46
Breast Cancer	73,24	75,11	73,99	74,02	0,60
Cleveland	83,14	83,85	83,48	83,52	0,20

Tüm yöntemler her bir veri seti için 10 defa 10 katmanlı çapraz doğrulama ile koşturulmuştur. Her bir yöntemin ürettiği doğruluk değerlerinin en düşüğü (ED), en yükseği (EY), ortalaması (ORT), medyanı (MED) ve standart sapmaları (SS) sırasıyla Çizelge 2, 3, 4, 5, 6'da verilmiştir.

Çizelge 2-6 incelendiğinde, Arrhythmia veri setinde 75, 21

Çizelge 3. SVM algoritmasının her bir veri seti için sınıflandırma başarımlar dağılımı

	ED	EY	MED	ORT	SS
Arrhythmia	73,33	74,76	74,40	74,26	0,55
Breast Cancer	74,38	75,16	74,72	74,67	0,24
Cleveland	79,48	80,52	80,10	80,09	0,37

Çizelge 4. KNN algoritmasının her bir veri seti için sınıflandırma başarımlar dağılımı

	ED	EY	MED	ORT	SS
Arrhythmia	61,19	64,05	63,45	63,17	0,99
Breast Cancer	68,66	74,38	72,59	72,20	1,57
Cleveland	73,07	76,78	75,41	75,22	1,03

Çizelge 5. DT algoritmasının her bir veri seti için sınıflandırma başarımlar dağılımı

	ED	EY	MED	ORT	SS
Arrhythmia	69,29	73,33	72,14	71,93	1,18
Breast Cancer	65,05	67,21	66,08	66,04	0,62
Cleveland	69,75	72,43	71,53	71,42	0,83

Çizelge 6. NB algoritmasının her bir veri seti için sınıflandırma başarımlar dağılımı

	ED	EY	MED	ORT	SS
Arrhythmia	74,76	75,48	75,24	75,19	0,22
Breast Cancer	72,53	72,99	72,89	72,78	0,18
Cleveland	82,45	83,18	82,82	82,82	0,28

Çizelge 7. Her bir veri seti için YÖK ve SVM doğruluk değeri arasındaki istatistiksel Mann Whitney U testi sonuçları

	Z	U	Ort. Fark.	p	AF*
Arrhythmia	-3,495	5,000	9,52E-01	0,000	YÖK
Breast Cancer	-2,652	15,000	-6,51E-01	0,008	SVM
Cleveland	-3,787	0,000	3,44E+00	0,000	YÖK

*Anlamlı fark $p < 0,05$

Çizelge 8. Her bir veri seti için YÖK ve KNN doğruluk değeri arasındaki istatistiksel Mann Whitney U testi sonuçları

	Z	U	Ort. Fark.	p	AF*
Arrhythmia	-3,811	0,000	1,20E+01	0,000	YÖK
Breast Cancer	-3,100	9,000	1,82E+00	0,002	YÖK
Cleveland	-3,788	0,000	8,31E+00	0,000	YÖK

*Anlamlı fark $p < 0,05$

doğruluk değeri ile ortalamada başarıyı en yüksek olan yöntem YÖK iken 63,17 doğruluk değeri ile ortalamada başarıyı en düşük olan yöntemin KNN olduğu görülmektedir. Breast Cancer veri setinde 74,67 doğruluk değeri ile ortalamada başarıyı en yüksek olan yöntem SVM iken 66,04 doğruluk değeri ile ortalamada başarıyı en düşük olan yöntem ise DT olduğu anlaşılmaktadır. Cleveland veri setinde ise YÖK yapısının 83,52 doğruluk değeri ile ortalamada en başarılı yöntem olduğu ve DT yönteminin ise 71,42 doğruluk değeri ile ortalamada en başarısız yöntem olduğu görülmektedir.

Arrhythmia, Breast Cancer ve Cleveland veri setlerinden elde edilen doğruluk değerleri üzerinden YÖK ve SVM yöntemlerinin sınıflandırma başarıyı için kullanılan Mann Whitney U test sonuçları Çizelge 7’de verilmiştir. Çizelge 7 incelendiğinde; Arrhythmia ve Cleveland veri setlerinde YÖK yapısının lehine anlamlı farkın olduğu görülmektedir ($p < 0,05$). Breast Cancer veri setinde ise, SVM algoritmasının lehine anlamlı farkın olduğu anlaşılmaktadır ($p < 0,05$).

YÖK, KNN ve DT yöntemlerinin sınıflandırma başarıyı istatistiksel analizi için kullanılan Mann Whitney U test sonuçları Çizelge 8 ve Çizelge 9’da verilmiştir. Çizelge 8 ve Çizelge 9 incelendiğinde; Arrhythmia, Breast Cancer ve Cleveland veri setlerinde YÖK yapısının lehine anlamlı farkın olduğu görülmektedir ($p < 0,05$).

YÖK ve NB yöntemlerinin sınıflandırma başarıyının istatistiksel analizi için kullanılan Mann Whitney U test sonuçları Çizelge 10’da verilmiştir. Çizelge 10 incelendiğinde; Breast Cancer ve Cleveland veri setlerinde YÖK yapısının lehine anlamlı farkın olduğu görülmektedir ($p < 0,05$). Fakat Arrhythmia veri setinde ise anlamlı bir fark çıkmamıştır ($p > 0,05$).

4. Sonuç

Bu çalışmada ham veriden çeşitli özniteliklerin elde edilmesiyle sınıflandırma yapan çoğu makine öğrenme yöntemlerinin aksine; ham veriyi doğrudan kullanarak sınıflandırma işlemini başarıyla yapan derin öğrenme yöntemlerinin başarıyı test edilmiştir. Üç farklı veri seti literatürde sıklıkla kul-

Çizelge 9. Her bir veri seti için YÖK ve DT doğruluk değeri arasındaki istatistiksel Mann Whitney U testi sonuçları

	Z	U	Ort. Fark.	p	AF*
Arrhythmia	-3,797	0,000	3,29E+00	0,000	YÖK
Breast Cancer	-3,782	0,000	7,98E+00	0,000	YÖK
Cleveland	-3,790	0,000	1,21E+01	0,000	YÖK

*Anlamlı fark $p < 0,05$

Çizelge 10. Her bir veri seti için YÖK ve NB doğruluk değeri arasındaki istatistiksel Mann Whitney U testi sonuçları

	Z	U	Ort. Fark.	p	AF*
Arrhythmia	-0,506	43,500	2,38E-02	0,613	-
Breast Cancer	-3,787	0,000	1,24E+00	0,000	YÖK
Cleveland	-3,560	3,000	6,99E-01	0,000	YÖK

*Anlamlı fark $p < 0,05$

lanılan 4 farklı sınıflandırma yöntemiyle teste tabi tutularak elde edilen sonuçlar önerilen yöntemle karşılaştırılmıştır. Gerek doğruluk oranları gerek istatistiksel sonuçlar önerilen yöntemin başarıyı ortaya koymaktadır. Derin öğrenme yöntemlerinin özellikle ham veriden veriyi daha iyi temsil eden yeni öznitelikler çıkararak sınıflandırma yapılabilmesi diğer makine öğrenme yöntemleri karşısında ciddi bir üstünlüktür. Derin sinir ağlarının karmaşık bir yapıya sahip olması ve yapının amaç fonksiyonunun optimizasyonunun zor olması dezavantajlar olarak görülse de ham veriyi doğrudan kullanabilme yeteneği bu dezavantajları anlamsız hale getirmektedir. Bu kapsamda ilerleyen dönemlerde derin öğrenme yöntemlerinin hızlı yaygınlaşacağı ve kullanım alanının artacağı düşünülmektedir.

5. Kaynaklar

- [1] Bengio Y., Courville A. ve P. Vincent, "Representation learning: A review and new perspectives", *IEEE Trans. Pattern Anal. Mach. Intell.*, cilt. 35, no. 8, sf. 1798-1828, 2013
- [2] LeCun Y., Bengio Y. ve Hinton G., "Deep Learning", *Nature*, cilt.521, sf.436-444, 2015.
- [3] Ranzato M., Poultney C., Chopra S. ve LeCun Y., "Efficient Learning of Sparse Representations with an Energy-Based Model", *Proceedings of Neural Information and Processing Systems*, 2006.
- [4] Bengio Y., Lamblin P., Popovici D. ve Larochelle H., "Greedy Layer-Wise Training of Deep Networks", *19. Advances in Neural Information Processing Systems*, 2006, sf. 153-160.
- [5] Ngiam J., Coates A., Lahiri A., Prochnow B., Le Q. V. ve Ng A. Y., "On optimization methods for deep learning", in: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, sf. 265-272.
- [6] Zhang Y., Zhang E. ve Chen W. "Deep neural network for halftone image classification based on sparse auto-encoder", *Engineering Applications of Artificial Intelligence*, cilt. 50, 2016.
- [7] Liu D.C. ve Nocedal J. "On the Limited Memory Method for Large Scale Optimization", *Mathematical Programming B*, cilt. 45, no. 3, sf. 503-528, 1989.
- [8] Lichman, M. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2013.