



## İLAÇ KEŞFİ VE GELİŞTİRİLMESİNDE YAPAY ZEKÂ

### ARTIFICIAL INTELLIGENCE ON DRUG DISCOVERY AND DEVELOPMENT

İrem Nur ÇELİK\* , Firdevs Kübra ARSLAN , Ramazan TUNÇ , İlkay YILDIZ\* 

Ankara Üniversitesi Eczacılık Fakültesi Farmasötik Kimya Anabilim Dalı, 06560, Ankara, Türkiye

#### ÖZ

**Amaç:** Makine zekâsı olarak da bilinen Yapay Zekâ'nın ilaç keşfi ve geliştirilme sürecindeki yeri ve öneminin ortaya konması amaçlanmıştır.

**Sonuç ve Tartışma:** İlaç keşfi ve geliştirme aşamaları, insan sağlığına ve refahına katkıda bulunan en önemli çeviri bilim etkinlikleri arasındadır. Bununla birlikte, yeni bir ilacın geliştirilmesi oldukça karmaşık, pahalı ve oldukça uzun bir süreçtir. Maliyetlerin nasıl azaltılacağı ve yeni ilaç keşfinin nasıl hızlandırılacağı endüstride zorlu ve ivedi ile çözülmesi gereken bir soru haline gelmiştir. Yapay zekânın (AI) yeni deneysel teknolojilerle bir araya gelmesi, yeni ilaç arayışını daha hızlı, daha ucuz ve daha etkili hale getirmesi beklenmektedir. Bu derlemede, ilaç keşif sürecini hızlandırmak için ortaya çıkan yapay zekâ uygulamaları ele alınmıştır.

**Anahtar Kelimeler:** Bilgisayar-destekli ilaç keşfi, makine öğrenmesi, yapay zekâ

#### ABSTRACT

**Objective:** It is aimed to reveal the place and importance of Artificial Intelligence, also known as machine intelligence, in drug discovery and development.

**Result and Discussion:** The drug discovery and development stages are among the most important science activities contributing to human health and well-being. However, the development of a new drug is a complex, expensive, and lengthy process. How to reduce costs and accelerate the discovery of new drugs has become a challenging and urgent question in the industry. The combination of artificial intelligence (AI) with new experimental technologies is expected to make the search for new drugs faster, cheaper and more effective. In this review, emerging artificial intelligence applications to speed up the drug discovery process are discussed.

**Keywords:** Computer-aided drug discovery, machine learning, artificial intelligence

\* Sorumlu Yazar / Corresponding Author: İrem Nur Çelik  
e-posta / e-mail: irmnur@hotmail.com, Tel. / Phone: +90538 889 54 48

\* Sorumlu Yazar / Corresponding Author: İlkay Yıldız  
e-posta / e-mail: iyildiz@pharmacy.ankara.edu.tr, Tel. / Phone: +905335705033

## GİRİŞ

Hastalık biyolojisinin anlaşılmasındaki ilerlemelere ve teknolojiye etkileyici sıçramalara rağmen, pazara yeni ilaçlar getirmek, büyük ölçüde klinik çalışmalarda başarısızlıklarla ilişkili önemli maliyetler nedeniyle, zaman alıcı ve pahalı bir süreç olmaya devam etmektedir [1].

Küçük molekülü ligandlardan klinik olarak kullanılabilen ilaçlara kadar ilaç keşif yolu uzun ve meşakkatli bir süreç olmuştur. Mekanistik olarak, ilaçların sadece hastalık hedeflerine yeterince bağlanma affinitesi göstermesi değil, aynı zamanda bu hedeflere erişimi mümkün kılmak için çoklu fizyolojik engellerden uygun taşınmayı da sağlaması gerekir. Genetik varyasyonun ilaç yanıtlarını etkilediği farmakogenetik ve istenmeyen proteinlerle hedef dışı etkileşimlerden kaynaklanan kimyasal toksisite gibi diğer sorunların hepsinin ilaç tasarımında dikkate alınması gerekir. Bu nedenle, ilaç keşfindeki bu çok yönlü sorunlar, genellikle ilaç tasarımcıları için önemli zorluklar ortaya çıkarmıştır [2].

Pazara daha düşük maliyetle daha fazla hastaya ilaç sağlamak için yeni fikirlere, yeni ve gözden geçirilmiş ilaç keşif sürecine ilişkin kavramlara ve yenilikçi yaklaşımlara ihtiyaç vardır. Bu bağlamda, bilgisayar destekli küçük molekülü ilaç tasarımı uzun zamandır potansiyel bir fırsat olarak görülmüştür [1]. Son zamanlardaki yapay zekâ yaklaşımının yükselişi, bu zorluklara potansiyel çözümler olarak görülmektedir [2].

### Yapay Zekâ Tanımı ve Tarihçesi

Yapay zekâ (Artificial Intelligent (AI)), insan zekâsı sürecinin bilgisayarlar tarafından simülasyonudur. Süreç, bilgi edinmeyi, bilgiyi kullanmak için kurallar geliştirmeyi, yaklaşık veya kesin sonuçlara varmayı ve kendi kendini düzeltmeyi içerir. Yapay zekâ, bilgisayarların insan davranışını taklit etmesini sağlayan tekniklerin kullanımı olarak tanımlanmaktadır [3].

Yapay zekâ tabanlı algoritmalar, tahmin edilecek moleküler özellikleri belirleyen fiziksel yasaların tam olarak bilinmediği veya deneysel ilişkilerin kurulamayacak kadar karmaşık olduğu sorunlar için özellikle çok uygundur [4]. Birçok alanda yapay zekâ tabanlı yaklaşımlardan yararlanılmaktadır. Bu alanlardan biri de ilaç geliştirmedir.

"Yapay zekâ" kavramı, John McCarthy tarafından 1956'daki Dartmouth Konferansında "akıllı makineler yapma bilimi ve mühendisliğini" tanımlamak için üretilmiştir [5]. McCarthy'nin bu tanımı, günümüzde hala geçerlidir. Multidisipliner bir alan olan AI, bilgisayar bilimi, matematik, psikoloji, dilbilim, felsefe, nöro-bilim, yapay psikoloji ve diğerleri gibi çeşitli disiplinlerden gelen görüşleri entegre eder. Son entelektüel ve mühendislik gelişmeler, yaşantımızın çeşitli yönlerinde problemleri çözen akıllı sistemlerin uygulanmasına doğru tamamen teorik çalışmalardan bu alanın ilerlemesine yardımcı olmaktadır. Bu tür uygulamaların kapsamında doğal dil anlama ve işleme, konuşma anlama ve işleme, mekanik / bilgisayar vizyon, otonomik / akıllı robotlar ve alan uzmanlığı edinme gibi farklı

alanlar ve çalışmalar bulunmaktadır [6]. AI tarafından ortaya çıkan geniş çaplı sorunlara rağmen, her durumda bilgi edinme ve sürdürmeyi içeren, bilgi temsili, çözüm arama, mantık muhakemesi ve makine öğrenmesi gibi büyük rol oynayan bazı temel yöntemler vardır [7, 8].

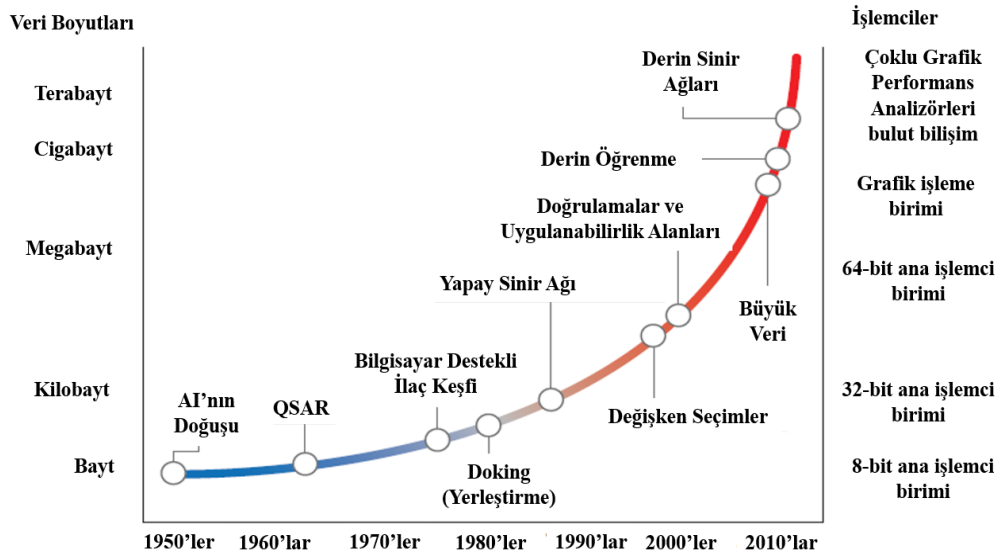
1930'ların ortalarında Alan Turing, günümüzde "evrensel Turing makinesi" olarak adlandırılan ve olası herhangi bir bilgisayarı simüle edebilen bir fikri ortaya atmıştır. Yapay zekânın tarihi gelişiminde canlanma ve çöküş gibi düzensiz bir ilerleme olmuştur [6]. 1950'ler ve 1960'lar bu konuda altın çağın erken dönemlerini oluşturmuştur. Bu dönemlerde hüristik araştırmalar, mantıksal akıl yürütme kavramları ve uygulanabilir insan-makine etkileşim için sembolik yöntemler ortaya konmuştur. 1964'te bazı matematiksel teoremlerin ve ifadelerin mantıksal çıkarımlarının makine kanıtlarını ortaya koyabilen STUDENT adlı bir makine ön zekâyâ sahip ilk makine olarak tasarlanmıştır [9]. 1966 yılında ise sınırlı bir biçimde de olsa insan diyalogunu taklit edebilen ELIZA adlı bir makine keşfedilmiştir [10]. Yapay zekâ, 1980'lerin başında ikinci zirvesine ulaşmıştır. Yapay zekâ ile ilgili matematiksel modellerde, çok katmanlı ileri beslemeli sinir ağı ve geri yayılım algoritması dahil olmak üzere önemli ilerleme kaydetmiştir [11, 12]. Bu araçlar, evrenin soyut bir modelinin oluşturulmasına ve veri girişi yapılmış modelin (geri bildirimden öğrenme) güncellenmesine olanak sağlar. Bu tür yöntemlerin kimya ve moleküler biyoloji alanına ilk girişi, protein sekans bilgisinden ikincil yapının tahmin edilmesiyle sağlanmıştır [13].

1950'lerde doğan Yapay Zekâ kavramı sonraki süreçlerde ilaç keşfinde de kullanılmıştır. Yapay zekânın özellikle yapı-aktivite ilişkilerinin modellenmesinde kullanımının ilaç keşfine uygulanması tahmin edilenden daha eskiye dayanmaktadır. Örneğin, deneysel değerleri alma ve regresyon için bir "descriptor (tanımlayıcı) set" kullanma fikri, Hammett'in [14] benzen türevlerinin reaksiyonları için reaksiyon hızlarını ve denge sabitlerini bağlayan öncü formülüne (ve muhtemelen daha da öncesine) ve bilgisayar destekli tanımlamaya ve ilaç endüstrisinde "Kantitatif Yapı-Etki İlişkileri (Quantitative Structure-Activity Relationship (QSAR))'nin babası" olarak kabul edilen Hansch'ın biyoaktif moleküllerin fizikokimyasal özelliklerinin kantitatif olarak tanımlanması ve ölçülmesine kadar gider [15,16]. Kimyasalların biyolojik etkilerinin tahmin edilebilmesi ve değerlendirilmesi amacıyla o zamanlardan beri artan sayıda medisinal kimyacılar çeşitli AI yöntemlerini uygulamışlardır [17,18,6]. Dolayısıyla, AI, 1960'larda sunulan ilk Kantitatif Yapı-Etki İlişkileri çalışmasından sonra ilaç keşfinde kullanılmıştır [15]. Kantitatif Yapı-Etki İlişkileri yaklaşımlarına dayalı olanlar gibi mevcut hesaplama modelleri, çeşitli istenen farmakolojik aktiviteler için hızlı bir şekilde çok sayıda yeni etkili bileşiğin tahmin edilmesinde kullanılmaktadır. Mevcut modeller (örn., İlaç keşif için kullanılan ticari yazılımlar) basit fizikokimyasal özelliklere (örn., LogP, çözünürlük vb.) ait tahminler de yapabilmekte ve böylece basit mekanizmalarla yeni bileşiklerin farmakokinetik özellikleri hakkında kesin olmamakla birlikte nispeten yakın tahminler ortaya konmaktadır. Ancak karmaşık biyolojik özellikler için modeller (ilaç etkinliği ve yan etkiler) optimum olmaktan hala uzaktır. İlaç keşfinin erken aşamasında (1990'lardan

önce) model geliştirmek için kullanılan yaygın hesaplama yaklaşımları doğrusal regresyonlardı [19]. Bu ilk çalışmalarda, modelleme için kullanılan kimyasal tanımlayıcılar, atom tipi ve fragman tanımlayıcıları gibi kimyasal yapı özellikleri ile de sınırlıydı. Bu yaklaşımlar, yüksek veri boyutluluğuna duyarlı ve sınırlı derecede yorumlanabilecek korelasyon özelliği içerirlerdi. Modern regresyon-temelli yaklaşımlar ise bu sorunları giderecek teknikleri içermektedirler [20]. Örnek olarak, k-En Yakın Komşu (k-Nearest Neighbor (k-NN)), Destek Vektör Makineleri (Support Vector Machine (SVM)) ve Rastgele Orman (Random Forest (RF)) gibi doğrusal olmayan modelleme algoritmalarına dayalı olarak geliştirilen yeni makine öğrenmesi yaklaşımları sıklıkla kullanılmaktadır [21].

1990'lardan bu yana kombinatoryal kimyadaki büyük ilerlemelerle, büyük kimyasal kütüphaneler yeni kimyasal geliştirme prosedürlerinin ana kaynağı haline gelmiştir. Son on yılda, bu çaba aynı zamanda yüksek verimli tarama (High-throughput screening (HTS)) tekniklerinin geliştirilmesini de teşvik etmiştir [22].

Büyük verileri kullanmanın zorlukları ve yeni veri türlerinin (örn., Görüntüler) dahil edilmesi, modern ilaç keşfinde öngörücü modellemeyi ilerletmek için yeni AI yaklaşımlarının yakın zamanda geliştirilmesini gerektirmiştir. Mevcut büyük veri çağındaki popüler AI yaklaşımları derin öğrenmeye dayanmaktadır. Şekil 1'de gösterildiği gibi geçmişten günümüze artan veri boyutu ve bilgisayarın gücüyle yapay zekânın gelişimi şematize edilmektedir [21].



Şekil 1. Artan veri boyutu ve bilgisayar gücüyle ilaç keşfinde yapay zekânın tarihsel gelişimi [21].

### Yapay Zekânın Alt Grupları

Yapay zekân teknolojileri, verilerden kavramları ve ilişkileri çıkarır ve veri modellerinden bağımsız olarak öğrenerek insanların neler yapabileceğini araştırır. Bu teknolojiler arasında bilgisayarla

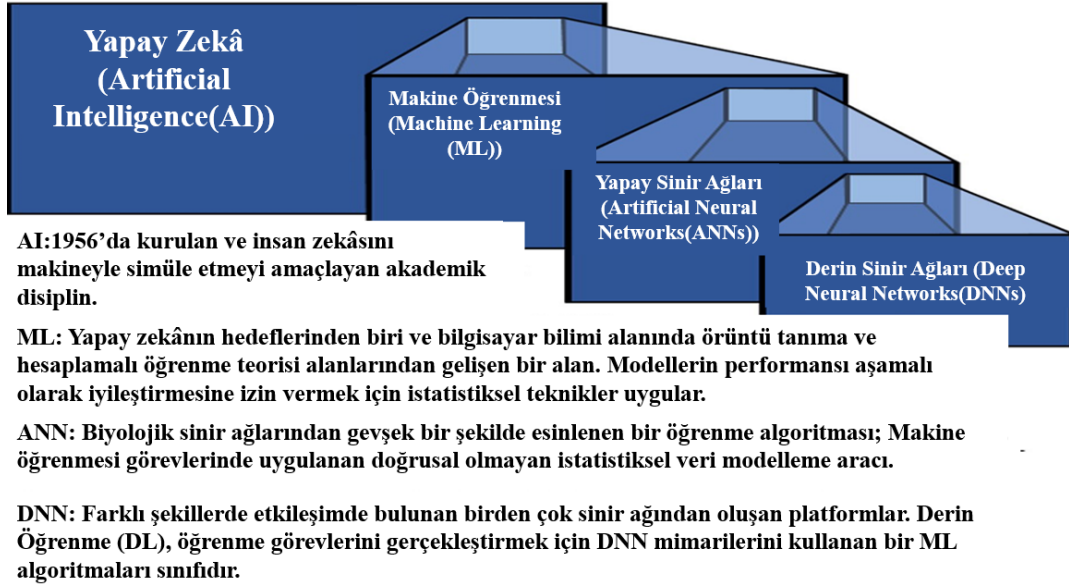
görme, derin öğrenme, makine öğrenmesi, doğal dil işleme, robotik, konuşma, denetimli öğrenme ve denetimsiz öğrenme gibi uygulamalar bulunur [23]:

*Makine Öğrenmesi (Machine Learning (ML))*: Yapılandırılmış ve yapılandırılmamış verilerden öğrenen, gizli kalıpları belirleyen, sınıflandırmalar yapan ve gelecekteki sonuçları tahmin eden bilgisayar algoritmalarıdır.

*Derin Öğrenme (Deep Learning (DL))*: Konuşma, görüntü ve video gibi kalıpları tanımak ve ayırt etmek için sinir ağları adı verilen beyne benzer bir mantık yapısını kullanan makine öğrenmesi tabanlı bir yaklaşımdır.

*Doğal Dil İşleme (Natural Language Processing (NLP))*: Hesaplama tekniklerinin doğal dil ve konuşmanın analiz ve sentezine uygulanmasıdır [23].

AI'nın hedeflerinden biri model performansının aşamalı olarak iyileştirilmesine izin veren ML platformlarının oluşturulmasıdır. Örneğin, sinir ağlarının karmaşık mimarilerine dayanan yeni bir makine öğrenmesi algoritmaları sınıfı olan derin DL, mevcut modelleri iyileştirmek veya yeni hesaplama platformları geliştirmek için olası bir araç olarak Şekil 2'de de gösterildiği gibi öne çıkmaktadır. Bu bakış açısı, derin öğrenme dahil olmak üzere yeni makine öğrenmesi yöntemlerinin ilaç keşif sürecinin farklı alanlarında nasıl uygulandığını gösteren örneklere genel bir bakış sunmaktadır [24].



**Şekil 2.** Yapay Zekâ ve Derin Öğrenme Arasındaki Temel Tanımlar ve İlişkiler [23].

Genel olarak anlaşılması gereken dört farklı makine öğrenmesi türü bulunmaktadır.

*Denetimli öğrenme (Supervised learning)*, bir modelin belirli bir sonucu tahmin etmek için etiketli veri kümeleri kullanılarak eğitilmesidir ve görev odaklıdır [25]. Bunun bir örneği, kantitatif yapı-aktivite

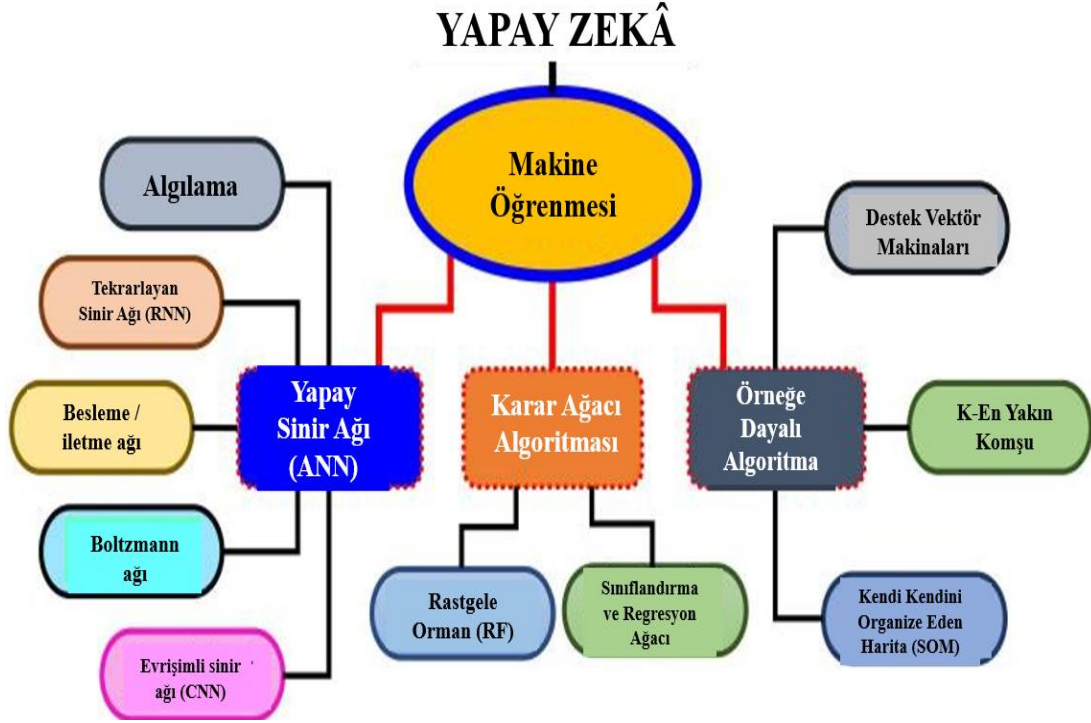
ilişkisi yaklaşımıdır. Bir kimyasalın çözünürlük ve biyoaktivite gibi özelliklerini tahmin etmek için kullanılır [26]. Denetimli öğrenmede yapay sinir ağına hem girdi hem de çıktı değerleri sunulur. Ağın ürettiği çıktı ile istenen çıktı arasındaki fark sıfır ya da sıfıra yakın bir değere ulaşmaya kadar ağırlıklar değiştirilir. Algoritma kabul edilebilir doğruluğa geldiğinde (veya daha fazla ilerleme kaydedilmediğinde) ya da eğitim veri kümesi tükendiğinde bunlardan hangisi önce gelirse yineleme sonlandırılır [27]. Bu öğrenme modelini kullanan ağlara Takviyeli Öğrenme (Reinforcement Learning (RL)), Perceptron ve İlişkili Hafızalar, Delta ve Genelleştirilmiş Delta Kuralı, Vektör Nicelik Öğrenmesi, Geri Yayılma Algoritması (Back-propagation algorithm), Stokastik Öğrenme (Stochastic Learning), Karar Ağaçları (Decision Trees), RF, SVM, Doğrusal ve Lojistik Regresyon (Linear and Logistic Regression), Naive Bayes Sınıflandırıcısı (Naive Bayes Classifier), Doğrusal Ayrımcılık Analizi (Linear Discriminant Analysis), k-NN algoritmaları ve Sinir Ağları (Neural Networks (NNs)) örnekleri verilebilir [28]. Denetimli makine öğrenmesinde iki tür vardır: Sınıflandırma ve Regresyon. Tahmin edilen çıktının sınırlı bir değerler kümesi olduğu gibi, sorun kategorik olduğunda sınıflandırma modelleri kullanılır. Regresyon modelleri, sorun bir aralık içindeki sayısal bir değeri tahmin etmeyi içerdiğinde kullanılır. Rastgele Orman, Otomatik Kodlayıcılar (AutoEncoder (AE)) ve Evrişimli Sinir Ağları (CNN) gibi çeşitli farklı türde makine öğrenmesi modelleri bu kategorideki uygulamalara örnektir [26].

*Denetimsiz öğrenme (unsupervised learning)*, verilerle ilişki bulmak için etiketli verilerle eğitime bağlı değildir. Veri odaklı olup, gizli yapıları bulmak için verilerdeki kalıpları öğrenmeyi amaçlar [25]. Denetimsiz öğrenmede, ağa sadece girdi vektörü uygulanır. Girdi değerlerine uygun bir çıktı üretilinceye kadar bağlantı ağırlıkları değiştirilir. Bu metot görüntü işleme, işaret işleme, kontrol problemlerinde etkin olarak kullanılır [28]. Bunlardan en yaygın olanları şunlardır: Hiyerarşik Kümeleme (Hierarchical Clustering), K-Ortalama Kümeleme (k-Means Clustering), Gaussian Karışım Modelleri (Gaussian Mixture Modeling (GMM)), Kendi Kendini Düzenleyen Haritalar (The Self-Organizing Map (SOM)) ve Gizli Markov Modelleri (Hidden Markov Model (HMM)) [27].

*Yarı denetimli öğrenme (Semi-supervised learning)*, daha küçük bir etiketlenmiş veri kümesi, birkaç veriden bir modelin tahmin edilmesi gereken ve daha az maliyetli olan daha büyük bir etiketlenmemiş veri kümesinin içindedir. Eğitim için etiketli veri noktalarını kullanır ve etiketli verilerden gelen bilgileri kullanarak bilinmeyen örneklerin etiketlerini tahmin etmeyi amaçlar [25, 29]. *Güçlendirilmiş öğrenme (Reinforced learning)*, sürekli öğrenen ve çevre ile etkileşimi olan ve belirli bir veri setini kullanmayı amaçlayan bir sistemdir.

Yarı denetimli öğrenmede, denetimli öğrenmeye benzer algoritmalar kullanılmıştır. Güçlendirilmiş öğrenme için kullanılan algoritmalara Q-Öğrenme (Q-Learning) ve Derin Öğrenme ağları örnek olarak verilebilir [29].

İlaç keşfi ve geliştirilmesi çalışmalarında uygulanabilecek yapay zekânın belli alt alanları Şekil 3'de topluca gösterilmiştir [30].



Şekil 3. Yapay Zekânın Yöntem Alanları [30].

Makine öğrenmesinin kapsadığı alt dallardan Derin Öğrenme, veri temsillerini öğrenmek için birçok doğrusal olmayan işlem birimi katmanıyla yapay sinir ağları (Artificial Neural Networks (ANN)) kullanan bir makine öğrenmesi algoritmaları sınıfıdır. İnsan merkezi sinir sistemini taklit etmek için bağlı yapay nöronlardan oluşan, yapay sinir ağı adı verilen geleneksel bir makine öğrenmesi algoritmasının yeniden markalandırılmasıdır [31, 32].

Derin Sinir Ağları (Deep Neural Networks (DNN)), birden çok gizli katmana sahip tamamen birbirine bağlı sinir ağlarıdır. Her gizli katman, birden fazla doğrusal olmayan işlem birimini içerir. DNN'ler, birçok farklı katmanda birden çok nöron kullanarak hiyerarşik düzeylerdeki özellikleri otomatik olarak çıkarır [25].

2014 yılından bu yana DNN'ler görüntü, ses ve metin tanıma, otonom sürüş ve diğer birçok işte insan doğruluğunu aşmıştır [33]. 2014 ve 2015 yıllarında derin öğrenmedeki ilerlemelere ilişkin ilaç endüstrisine yapılan erken sunumlar şüphecilikle sonuçlanmış ve reddedilmiştir. Fakat 2017'de birçok ilaç şirketi, yapay zekâ girişimleri için akademisyenlerle ortaklık kurmaya veya yerli Ar-Ge programları başlatmıştır. DNN'lerin küçük moleküllerin farmakolojik özelliklerini ve biyobelirteç gelişimini tahmin

etmek için transkripsiyonel yanıt verileri üzerine eğitiminden, yeni kimyanın üretilmesine kadar, derin öğrenme teknikleri biyomedikal araştırmanın birçok alanına hızla yayılmıştır.

Geleneksel makine öğrenmesi yöntemleri, manuel olarak tasarlanmış özellikleri kullanırken, derin öğrenme yöntemleri, özellikleri giriş verilerinden otomatik olarak öğrenebilir ve bu da çok katmanlı özellik çıkarma yoluyla basit özellikleri karmaşık özelliklere dönüştürebilir. Ayrıca, derin öğrenme yöntemlerinde geleneksel makine öğrenmesi yöntemlerine göre daha az genelleme hatası vardır. Bu da daha tatmin edici sonuçlar alınmasına yardımcı olur. Derin öğrenme yöntemleri esas olarak derin sinir ağlarından oluşur. Evrişimli sinir ağı, tekrarlayan sinir ağı (Recurrent Neural Networks (RNNs)), Otomatik Kodlayıcı, Kısıtlı Boltzmann Makinesi (Recurrent Boltzmann Machines (RBN)) gibi kısımlar bu sinir ağlarının parçalarıdır.

CNN, derin öğrenmedeki en temsili mimarilerden biridir. Görüntü ve ses tanımanın yanı sıra doğal dil işleme gibi birçok alanda yaygın olarak benimsenmiştir. Genellikle görüntü tanıma ve nesne sınıflandırması için kullanılan ileri beslemeli bir sinir ağıdır. Çok boyutlu diziler biçiminde olan verileri işlemek için oluşturulmuş veya geliştirilmiştir. Bu nedenle CNN, çoğunlukla gen ekspresyon veri analizinde kullanılır [34]. Geleneksel ilaç keşfinde, CNN'ler, HTS sonuçları gibi deneysel ilaç testlerinden elde edilen görüntü verilerini analiz etmek için de uygulanmıştır. Görüntü tanımadaki benzersiz avantajları nedeniyle, CNN'ler ligand-protein etkileşimlerini tahmin etmek için ve 3-boyutlu görüntüleri için de kullanılmıştır [21]. Adları evrişimsel katmanlardan gelir. Modern CNN, 1980'lerde Fukushima tarafından, Hubel ve Wiesel'in bir kedinin görsel korteksindeki alıcı alan (receptive field) araştırmalarından esinlenen Neocognitron'un geliştirilmesinden gelmiştir [35,36]. Görsel sinyalleri işlerken, yerel nöron kalıpları duyuşal boşluktaki belirli bölgeleri algılama sorumluluğunu üstlenir ve CNN, evrişimli katmanlarda iki ana karakter geliştirerek özelliklerini taklit eder: Seyrek Bağlanabilirlik ve Paylaşılan Ağırlıklar [32].

RNNs, derin öğrenmedeki diğer bir temsili türüdür. ANN'lerin "zaman serisi versiyonu" dur. Özellikle sekans verilerini işlemeyi amaçlayan RNN, NLP'de yaygın olarak kullanılmış ve büyük başarılar elde edilmiştir. Tahmin modellerinin ve dil modellerinin temelini oluştururlar. RNN, bir katmanın çıktısını kaydetme ve katmanın çıktısını tahmin etmek için girişi geri besleme prensibiyle çalışır [32]. Tekrarlayan sinir ağlarında ileri beslemeli ağların aksine, bir geri bildirim döngüsü vardır [37].

Son zamanlarda, AI teknolojileri, özellikle derin öğrenme modelleri, güçlü genelleme ve özellik çıkarma kabiliyetleri nedeniyle ilaç tasarımında büyük umutlar vaat etmektedir.

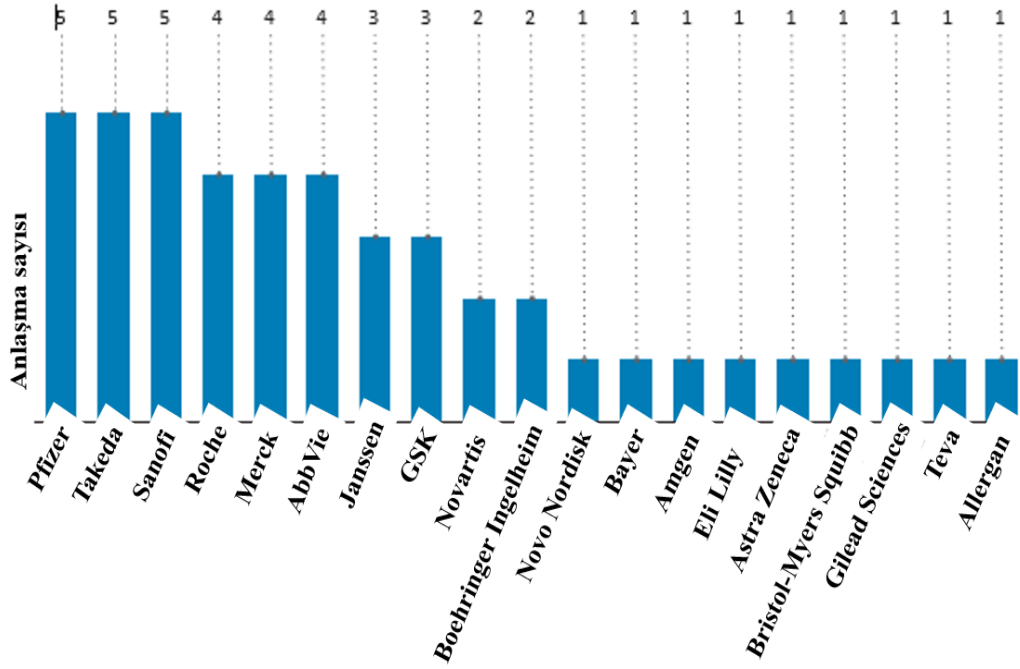
### **İlaç Keşfi ve Geliştirme Sürecinde Yapay Zekâ**

İlaç keşif süreci, halihazırda mevcut literatürü okumaktan ve analiz etmekten potansiyel ilaçların hedeflerle etkileşim yollarını test etmeye kadar uzanır. İlaç keşfinin klinik öncesi geliştirme aşaması, hayvan modellerinde potansiyel ilaç hedeflerinin test edilmesini içerir. Bu aşamada yapay zekânın





(Quantitative Structure-Properties Relationship) modellerini oluşturmada giderek daha başarılı uygulamalara doğru ilerlemektedir. Bunlarla ilgili bazı uygulamalar Tablo 1’de verilmiştir. 2012’nin başlarında Hilton grubu, kimyasal bileşik aktivitesi ve özellikleri tahmini için geliştirdikleri DL yöntemi ile Merck Kaggle yarışmasını [41] kazanmışlardır.



Şekil 5. Pazarda Açıklanan Yapay Zekâ Odaklı Büyük Biyo-İlaç Anlaşmaları [38].

İlaç keşfinde AI için en önemli fırsat, ilaçların yeniden konumlandırılarak kullanılmasıdır. Donner ve arkadaşları gen ekspresyon veri temelli bileşik fonksiyonel benzerliğini ölçmek için yeni bir metod bildirmişlerdir. Bu yaklaşım, DNN’nin avantajlarını kullanarak bir moleküldeki fonksiyonlu grubun yapısını taşıyan en yakın benzerlerini taşıyan bileşikleri bütünlüğü bozan verinin en aza indirgenmesiyle bulabilmektedir. Bu yöntemin bileşikler yapısal olarak farklı olduğunda bile paylaşılan terapötik ve biyolojik hedeflere sahip ilaçları belirleyebildiği ve böylece bileşikler arasında daha önce bildirilmemiş fonksiyonel ilişkileri ortaya çıkarabildiği bildirilmiştir. Bahsi geçen bu yaklaşım DeepCodex çevrimiçi olarak kullanıma sunulmuştur [42]. Bu ölçüm, yapısal olarak farklı olmasına rağmen biyolojik hedefleri paylaşan bileşiklerin tanımlanmasına izin vererek, bileşikler arasında önceden bilinmeyen fonksiyonel ilişkileri ortaya çıkarmıştır [43].

Xie ve arkadaşları, Entegre Ağ Tabanlı Hücrel İmzalar Kitaplığı (Library of Integrated Network-based Cellular Signatures (LINCS))’nın L1000 veritabanındaki transkriptom verilerini kullanarak ilaç-hedef etkileşimlerini öngörmek için bir DL metodu çalışması gerçekleştirmişler ve

modelin doğruluğunun %98 oranında olduğunu bildirmişlerdir [44]. Önceki verilerin derin öğrenme yaklaşımlarıyla yeniden kullanılması, ilaçların yeniden kullanılması için önemlidir [45].

**Tablo 1.** Küçük moleküllü ilaç keşfinde DL'nin temsili uygulamalarının özeti [32].

Metod	Öngördüğü	Veri seti	Özellikler
RNN	Bileşimin sudaki çözünürlüğü	1144, 1026, 74, 125 genişliğinde çoklu veri seti	Moleküler grafik
DNN	İlaç-hedef etkileşimi	2710 ligand ve 836 hedef	Moleküler tanımlayıcılar ve protein özellikleri
DNN	İlaç-hedef etkileşimi	MATADOR veri kümesi ve STITCH veri kümesi	Moleküler tanımlayıcılar
DNN	Geçirgenlik	663 + 209 kimyasal bileşik	Moleküler tanımlayıcılar
DNN	Toksosite	Tox21 veri kümesi	Moleküler parmak izleri ve tanımlayıcılar
RNN	Bileşimin proteinle etkileşimi	ChEMBL veritabanı; BindingDB	Moleküler parmak izleri ve protein dizisi
CNN	Biyolojik aktivitesi	ChEMBL veritabanı, DUDE veri kümesi	Moleküler grafik (AtomNet)
CNN	Sanal gösterimi	DUDE veri kümesi	Moleküler grafik ve doking (yerleştirme) sonucu
AE	Sanal gösterimi	6252	Parmak izleri
CNN	Biyolojik aktivite/toksitesite	41,193/8014	2D kimyasal yapı görüntüsü
RNN	Biyolojik aktivitesi	756	SMILES
DNN	Biyolojik aktivitesi	ChEMBL veritabanı	Moleküler tanımlayıcılar ve parmak izleri
RNN	Odaklanmış moleküler kitaplıklar oluşturma	ChEMBL veritabanı	SMILES
RNN	Yeni moleküller üretmek	ChEMBL veritabanı	SMILES
Gan	Yeni moleküller üretmek	Birden çok veri kümesi	SMILES
CNN	Kohn-Sham kinetik enerjisi	3D elektron yoğunluğu	Parmak izleri

Gan: Çekişmeli Üretken Ağ

LINCS Programı Kütüphanesi, gen ekspresyonundaki değişiklikleri ve hücreler çeşitli rahatsız edici maddelere maruz kaldığında meydana gelen diğer hücresel süreçleri kataloglayarak, ağ tabanlı bir biyolojik anlayışı oluşturmayı amaçlamaktadır [46]. Metformin ve rapamisin, FDA onaylı rapamisinin mekanistik hedef (mechanistic target of rapamycin (mTOR)) inhibitörleridir ve mevcut klinik

uygulamalarının ötesinde önemli anti-kanser ve yaşlanma önleyici özellikler sergilerler. Bununla birlikte, her biri, advers etkiler nedeniyle etiket dışı, profilaktik kullanım için onay gerektiren sorunlarla karşı karşıyadır. Bu nedenle, metformin ve rapamisinin yaşlanmayı geciktirici etkilerini advers etkiler olmaksızın taklit eden nutrasötikleri (daha güvenli, doğal olarak oluşan bileşikler) belirlemek için bir çalışma başlatılmıştır. Metformin ve rapamisinin gen ve yol düzeyinde imzalarını haritalamak ve 800'den fazla doğal bileşik arasındaki eşleşmeleri taramak için LINCS veri kümesine çeşitli biyoinformatik yaklaşımlar ve derin öğrenme yöntemleri uygulanmıştır. Daha sonra, derin sinir ağı sınıflandırıcılarından oluşan bir toplulukla her bileşiğin güvenliğini tahmin edilmiştir. Analiz, allantoin ve ginsenosid (metformin), epigallokateşin gallat ve izoliquiritigenin (rapamisin) ve withaferin A (her ikisi) dahil olmak üzere birçok yeni aday metformin ve rapamisin mimetiklerini ortaya çıkarmıştır. Keşfedilmemiş bu dört bileşik de rapamisin ile benzer skorda bulunmuştur [46].

Bir ilaç adayı molekülün etki mekanizmasını ve *in vivo* güvenliğini öngörebilen bir yapay zekâ platformu, boşa harcanan zamanı ve maliyeti önemli ölçüde azaltabilmektedir [26]. Bu amaçla birkaç şirket örneği verilebilir. Bunlar DeepTox ve ProCTOR dur [47, 48]. Bunların her ikisi de yeni bileşiklerin toksisitesini öngörmeyi amaçlamaktadır. Bu AI platformlarının performansının, bileşiklerin toksisitesine ilişkin daha büyük ve sağlam veri kümeleri kullanılabilir hale getirildikçe artması beklenmektedir

Zhavoronkov ve arkadaşları, *de novo* küçük molekül tasarımı için üretken tensorial pekiştirmeli öğrenme (generative tensorial reinforcement learning (GENTRL)) adında bir DL modeli geliştirmişlerdir [49]. GENTRL, sentetik fizibiliteyi, yeniliği ve biyolojik aktiviteyi optimize etmektedir. Araştırmacılar, GENTRL'i kullanarak 21 gün içinde fibroz ve diğer hastalıklarda rol oynayan bir kinaz hedefi olan Diskoid Alan Reseptörü 1'i (Discoidin Domain Receptor 1 (DDR1)) için güçlü inhibitörler keşfetmişlerdir. Elde edilen sonuçları valide etmek için öncelikle bileşikler sentezlemişler ve daha sonra hücre bazlı testlerle biyokimyasal ölçümlerini yapmışlardır. Bu çalışmada elde ettikleri bir öncü aday bileşiğin farelerde farmakokinetiği kanıtlanmıştır. Bu çalışma 46 günde tamamlanmıştır. Bu çalışma bize hızlı ve etkili moleküler tasarım sağlamak için derin öğrenme yöntemlerinin potansiyelini göstermektedir.

Rastgele Orman, sınıflandırma ve regresyon problemlerini bir grup karar ağacı kullanarak çözmek için popüler bir yöntemdir. Dikkatli tanımlayıcı seçimi ve kapsamlı parametre ayarı olmadan bile QSAR modellemesinde çok iyi performans sağlayabilir [50]. NCI-60 gibi hücre hattı panelleri, ilaç adaylarının proliferasyonu inhibe etme yeteneklerini test etmek için uzun süredir kullanılmaktadır. Daha önceleri *in vitro* ilaç duyarlılığının öngörücü modelleri, gen ekspresyon mikrodizilerinden oluşturulan gen ekspresyon belirteci kullanılarak oluşturulmuştur. Bu istatistiksel modeller, orijinal NCI-60'ta olmayan hücre hatları için ilaç yanıtının öngörüsüne izin verebilmektedir [51]. Sınıflandırma ve regresyon ağaçlarına (classification and regression trees, CART) dayanan bir yaklaşım olan Rastgele Orman'ı

kullanarak ilaç cevabının regresyon modellerini oluşturan yeni birçok adımlı algoritma geliştirilerek mevcut teknikler iyileştirilmiştir. Bu yöntem, hem 19 meme kanseri hem de 7 Glioma hücre hattından oluşan bir panel için ilaç yanıtını öngörmeye başarılı olmuş, diferansiyel gen ekspresyonuna dayalı diğer yöntemlerden daha iyi performans göstermiştir. Bahsi geçen yazılım programı R dilinde yazılmış ve ilgili gen ekspresyonu ve ilaç yanıt verileriyle birlikte ivDrug paketi olarak <http://r-forge.r-project.org> adresinde mevcuttur.

Iorio ve arkadaşları [52], gen mutasyon profilleri ve ekspresyon seviyeleri göz önüne alınarak, kanser hücre dizilerinde ilaç moleküllerinin IC<sub>50</sub>'sini tahmin etmek için elastik ağ modelleri (elastic net models) oluşturmuşlar ve bileşiğe bağlı olarak bir dizi öngörülenlerde doğruluk gözlemiştir. Aynı veri setini kullanan Corte's-Ciriano ve arkadaşları ise [53] öngörü performansının bazı durumlarda her bir tahminde istatistiksel güven ölçüsüne bağlı olan RF modeli kullanılarak iyileştirilebileceğini göstermişlerdir.

ALOGPS, ilaç tasarımı (ADME / T ve HTS) ve çevre kimyası çalışmaları için bileşiklerin logP, suda çözünürlük ve pKa'larına ait öngürüler sağlamaktadır. ALOGPS 2.1, ilaç tasarımı çalışmalarında 1-oktanol/su içindeki partiyon katsayılarını, logP'yi ve nötr bileşiklerin sulu çözünürlüğünü öngörmek için geliştirilmiştir. Bu program Associative Neural Networks'ün kendi kendine öğrenme özellikleri aracılığıyla kullanıcı tarafından sağlanan yeni verileri birleştirme yeteneğine sahip olması nedeniyle ayrıcalıklı bir özelliğe sahiptir [54].

Lusci ve arkadaşları, ilaç benzeri moleküllerde suda çözünürlük tahmininde bulunmak için derin öğrenme modelleri bildirmişlerdir [55]. Bu çalışmada kimyasal yapılar grafiğe dönüştürülmüş ve modelleme için yinelemeli sinir ağı yaklaşımları uygulanmıştır.

Ferrero ve arkadaşları, Open Targets [56] platformundan çıkarılan gen-hastalık ilişkisi verilerine dayalı olarak terapötik hedeflerin belirlenmesi üzerinde çalışmışlardır [57]. Araştırmacılar, özellikle belirli bir genin bir ilaç hedefi olma olasılığını öngörebilmek amacıyla RF, SVM, NNs ve Gradyan Artırma Makinesi (Gradient Boosting Machine) gibi makine öğrenmesi modellerini kullanmışlardır [58].

Moleküler dinamik (MD) simülasyonları, reseptör işlevini mekanik olarak açıklayabilmektedir. Plante ve arkadaşları, yakın zamanda MD simülasyonlarının analizine yönelik bir makine öğrenmesi yaklaşımını tanımlamışlardır [59]. Makine öğrenmesi, ligandın neden olduğu reseptör konformasyonel değişiklikleri arasında ayırım yapmak için güçlü bir analiz aracı olarak kullanılabilir, bu da ligandın farmakolojik etkisinde yer alan kalıntıların tanımlanmasına izin verebilmektedir. Geniş bir ligand koleksiyonu incelenerek, makine öğrenmesi yaklaşımları ile potansiyel olarak yeni ilaçların farmakolojik profili sınıflandırabilir. Bu nedenle mekanizmaya dayalı ilaç keşfi için umut verici görünmektedir [60].

Dopamin D3 Reseptörünün (D3DR) etikloprite bağlı inaktif durumunun moleküler dinamik simülasyonları, dört konjenerik modülatördeki etkinlik değişikliğini rasyonelleştirme girişiminde makine öğrenmesine dayalı bir yaklaşım kullanılarak kümelenmiştir [61]. Reseptör-ligand komplekslerinin genişletilmiş MD trajektörlerini biriktirerek, ligand esnekliğindeki artışın, inaktive edilmiş reseptörün kristal yapısını aşamalı olarak nasıl destabilize ettiği gözlemlenmiştir. Bu modeli ileriye dönük olarak doğrulamak için, parsiyel bir agonist, yapısal bilgilere ve hesaplamalı modellemeye temelli rasyonel olarak tasarlanmış daha sonra sentezlenip test edilmiştir. Elde edilen sonuçların öngörülerle uyumlu çıktığı görülmüştür. Bu çalışma, genişletilmiş MD simülasyonları çerçevesinde ligand esnekliğinin araştırılmasının ilaç tasarım stratejilerine yardımcı olabileceğini, bilgi verebileceğini ve bunun *in silico'da* (bilgisayar ortamında) işlevsel testlerin güçlü bir karşılığı olarak potansiyel rolünün altını çizebileceğini ileri sürmektedir

Şiddetli akut solunum yolu sendromu koronavirüsü 2'nin (Severe Acute Respiratory Syndrome causing Coronavirus (SARS-CoV-2)) viral proteinleri üzerinde etkili olabilecek ticari olarak kullanılan mevcut ilaçları belirlemek için Molekül Transformatörü-İlaç Hedef Etkileşimi (Molecule Transformer-Drug Target Interaction (MT-DTI)) adı verilen önceden eğitilmiş (pre-trained) derin öğrenme tabanlı ilaç-hedef etkileşim modeli kullanılmış ve sonuç olarak, insan immün yetmezlik virüsünü (HIV) tedavi etmek ve önlemek için kullanılan bir antiretroviral ilaç olan atazanavirin en iyi kimyasal bileşik olduğu ve SARS-CoV-2 3C benzeri proteinaza karşı 94,94 nM'lik bir inhibitör potens ile remdesivir (113.13 nM), efavirenz (199.17 nM), ritonavir (204.05 nM) ve dolutegravir (336.91 nM) den daha etkili olduğu bildirilmiştir [62]. Bu derin öğrenme modeli sayesinde farklı bir viral enfeksiyonun tedavisinde kullanılan antiretroviralin başka bir viral enfeksiyondaki etkinliği ortaya konmuştur.

Green ve arkadaşları, 2020 yılında, kimyasal yapı oluşturma, deneysel tasarım, aktif öğrenme ve kemoinformatik araçlar için metodları entegre eden otomatik moleküler tasarım için Heterojen, Otomatikleştirilmiş İş Akışı kullanan Biyolojik Tepki Analizi ve Tasarım Sistemi (Biological Response Analysis and Design System using an Heterogenous, Automated Workflow (BRADSHAW)) adlı sistemi ortaya koymuşlardır [63]. Bu program, tasarla-yap-test et- analiz et (The Design-Make-Test-Analyse cycle (DMTA)) döngüsünde otomatik moleküler tasarıma izin vermektedir. Basit kullanıcı arayüzü, büyük ölçekli otomatik tasarıma erişimi kolaylaştırırken, çok hızlı hareket eden bir alanda kritik bir gereksinim olan yeni algoritmaları tanıtmak için gereken yazılım geliştirmeyi en aza indirmek için tasarlanmıştır. Sistem, bir otomasyon felsefesi, en iyi uygulama, deneysel tasarım ve hem geleneksel kemoinformatiklerin hem de modern makine öğrenimi algoritmalarının kullanımını içermektedir. Geriye dönük bir durum çalışmasında, BRADSHAW'ın MMP12 inhibitörlerinin tasarımı için öncü bileşik optimizasyonunda nasıl başarılı bir şekilde kullanılabildiği gösterilmiştir.

ADAM'lar (bir disintegrin ve metaloproteinazın kısaltması), hücre yapışması, göç, hücrel sinyalleme ve proteoliz vb. fizyolojik süreçlerde rol alan transmembran proteinler ailesine aittir [64, 65].

ADAM'ların önemli bir üyesi olan ADAM10'un Alzheimer, prion hastalığı, alerjik reaksiyonlar, glioblastoma ve pankreas kanserinin [66-69] tedavisi için umut verici bir hedef olduğu kanıtlanmıştır. ADAM10 inhibitörlerinin moleküler generatif modelini oluşturmak için, geçişli tekrarlayan birim (Gated-Recurrent-Unit (GRU)) tabanlı derin sinir ağı, transfer öğrenmesi ile birlikte başarıyla kullanılmıştır. Çalışmadan elde edilen sonuçlara göre, GRU tabanlı generatif modelin, moleküllerin SMILES gramerlerini doğru bir şekilde öğrenebildiğini ve yeni potansiyel ADAM10 inhibitörleri üretebildiğini göstermiştir. Geleneksel ligand tabanlı yöntemlerle karşılaştırıldığında, GRU tabanlı generatif model, kimyasal ligandların yalnızca SMILES bilgilerini gerektirir ve etkili geniş bir seri potansiyel yeni yapılar oluşturabilir. Bu benzersiz avantajlar, de novo ilaç tasarımı ve büyük ölçekli sanal tarama araştırmalarında son derece yararlı kılmaktadır [70].

Bu çalışmalara benzer şekilde Segler ve arkadaşları [71], SMILES notasyonunun sözdizimini öğrenmek için tekrarlayan sinir ağlarını kullanmışlar ve daha küçük bir bileşik setiyle önceden eğitilmiş bir ağa ince ayar yaparak hedef odaklı kütüphaneler oluşturmuşlardır. Bu transfer öğrenme yöntemi, biyolojik olarak aktif moleküllerin öngörülmesinde başarıyla kullanılmıştır.

CRISPR / Cas9, gen inaktivasyonu, aktivasyonu, mutasyonu ve kromozom yeniden düzenlemeleri için çok yönlü bir genomik mühendislik aracıdır [72]. Biyoteknoloji uzmanı Christof Fellmann, CRISPR-Cas'ın hedef molekülleri tanımlamaya yardımcı olmasının ilaç keşfi üzerinde çok önemli bir etkiye sahip olacağını açıklamıştır. Araştırmacılar, sistemi kasıtlı olarak genleri aktive etmek veya engellemek için kullanarak hastalığa neden olan veya engelleyen genleri ve proteinleri belirleyebilir, böylece potansiyel ilaçlar için hedefler belirleyebilmektedir. CRISPR – Cas, aynı zamanda, hastalıkları tam olarak taklit eden hücrel ve tüm hayvan modeli sistemleri oluşturmayı da kolaylaştırmaktadır. Bilim insanlarının ilaçların güvenliğini ve etkinliğini daha doğru bir şekilde doğrulamasına olanak tanıyıp, bu tür modellerin klinik çalışmalarda ne olacağına dair daha iyi öngörücüler olmasını sağlamaktadır [73].

AtomNet, ilaç keşif uygulamaları için küçük moleküllerin biyoaktivitesini tahmin etmek için tasarlanmış ilk yapı tabanlı, derin evrişimli sinir ağıdır [74]. AtomNet, atom tipleri ve yapısal protein-ligand etkileşimi parmak izleri (SPLIF'ler) gibi yapısal özellikleri çıkarmak için ilaç-hedef kompleksine yerleştirilmiş üç boyutlu (3D) bir ızgara (grid) gösterimini kullanır ve bunlar daha sonra tek boyutlu bir kayan nokta vektörüne açılır. AtomNet, bir çalışma setinden gerçek aktif bileşiklere öncelik vermede Surflex-Dock [75] ve DOCK [76] gibi yaygın olarak kullanılan yerleştirme algoritmalarından daha iyi performans göstermiştir.

Klinik araştırmalara giren çoğu ilaç başarısız olur, bu da genellikle ilaç yanıtını yöneten mekanizmaların tam olarak anlaşılmasından kaynaklanır. Makine öğrenimi teknikleri, daha iyi ilaç yanıtı tahminleri için muazzam bir umut vaat ediyor, ancak çoğu yorumlanabilirlik eksikliği ve monoterapilere odaklanmaları nedeniyle klinik uygulamaya ulaşamadı [77]. Bu zorlukları, 1.235 tümör

hücresi hattının 684 ilaca verdiği yanıtlar üzerine eğitilmiş insan kanser hücrelerinin yorumlanabilir bir derin öğrenme modeli olan DrugCell'i geliştirerek ele almışlardır. DrugCell öngörülerini hücre hatlarında doğrudur ve ayrıca klinik sonuçları sınıflandırır. DrugCell mekanizmalarının analizi, kombinatoriyal CRISPR, *in vitro* ilaç-ilaç taraması ve hastadan türetilmiş ksenograftlar ile sistematik olarak doğruladıkları sinerjik ilaç kombinasyonlarının tasarımına doğrudan götürür. DrugCell, öngörücü tıp için yorumlanabilir modeller oluşturmak için bir plan sağlamaktadır.

DeepACTION derin öğrenme tabanlı bir yöntemdir. Bu hesaplamalı yaklaşım yeni etkileşimleri (ilaç-hedef çiftleri) belirleyebilir ve ilacın yeniden kullanım sürecini hızlandırabilir. Bu sayede bilim adamlarını ilaç geliştirmeye motive etmek için eksiksiz bilgi sağlayabilmektedir. Yöntem bir doğrulama testi (5 kat CV) altında mevcut algoritmalarla karşılaştırılmış ve deneysel sonuçlar, yöntemin tüm ölçüm metrikleri açısından tahmin performansı elde ettiğini ve DrugBank veri setinden yeni çiftleri (ilaç-hedef) tahmin edebildiğini göstermektedir [78].

DeepCPI, büyük miktarda etiketlenmemiş veriden, bileşiklerin ve proteinlerin örtük ancak etkileyici düşük boyutlu özelliklerini otomatik olarak öğrenen derin öğrenme yöntemini kullanan bir algoritmadır. DeepCPI tarafından başlatılan sanal tarama ile küçük moleküllü bileşiklerin, glukagon benzeri peptid-1 reseptörü (GLP-1R), glukagon reseptörü (GCGR) ve vazoaaktif bağırsak peptid reseptörü (VIPR) dahil, G proteinine bağlı reseptör (GPCR) ailesindeki çeşitli hedeflerle etkileşimi belirlenmiştir. Bu durum DeepCPI'nin yeni bileşik-protein etkileşimlerinin öngörüsü için yararlı ve güçlü bir araç olduğunu, bu nedenle ilaç keşfine ve yeniden konumlandırma çabalarına yardımcı olabileceğini göstermektedir [79].

PASS Online, farmakolojik etkiler, etki mekanizmaları, toksik ve yan etkiler dahil olmak üzere 4000'den fazla biyolojik aktivite türünü tahmin etmeye yönelik Bayes yaklaşımının bir örneğidir [80]. Bileşiğin tahmini biyolojik aktivite profilini elde etmek için sadece yapısal formül gereklidir; böylelikle bilgisayarda tasarlanan ancak henüz sentezlenmemiş sanal yapı için bile öngöründe bulunmak mümkündür. Bayes ağının yapı öğrenmesini içeren etiket sınıflandırma yöntemi olan DRABAL, 400.000'den fazla bileşiğin 1,4 milyondan fazla etkileşimini incelemek ve PubChem BioAssay veri tabanından beş büyük HTS testi arasındaki mevcut ilişkileri analiz etmek için geliştirilmiştir [81].

Yakın zamanda, Birleşik Krallık'ta yeni kurulan Exscientia ile Japon ilaç şirketi Sumitomo Dainippon Pharma arasındaki bir iş birliği ile obsesif kompulsif bozukluk için AI kullanılmasıyla tasarlanmış bir ilaç olan DSP-1181 bildirilmiştir [82]. Gelişmiş yapay zekâ metodolojilerinin kullanılması ile, ilgili kimyasal alan verimli bir şekilde daraltılarak, yapay zekâ tarafından tasarlanmış 5-HT<sub>1A</sub> (5-Hidroksitriptamin 1A) reseptör agonistinin sadece 12 ayda kliniğe ulaşması sağlanmıştır. Normalde bu süreç birkaç yıldan on yıla kadar sürebilirdi.

Genel olarak, AI'nın çeşitli stratejiler yoluyla ilaç keşif verimliliğinde artışlar sağlayabileceği açıktır [26]. Tablo 2'de ilaç geliştirme sürecinde kullanılan bazı yapay zekâ araçları verilmektedir.



**Tablo 2.** İlaç keşfinde kullanılan yapay zekâ araçlarına örnekler [30].

Araçlar	Ayrıntılar	Web Sitesi URL
DeepChem	İlaç keşfinde uygun bir aday bulmak için python tabanlı bir AI sistemi kullanan MLP modeli	<a href="https://github.com/deepchem/deepchem">https://github.com/deepchem/deepchem</a>
DeepTox	Toplam 12.000 ilacın toksisitesini tahmin eden yazılım	<a href="http://www.bioinf.jku.at/research/DeepTox">www.bioinf.jku.at/research/DeepTox</a>
DeepNeuralNetQSAR	Bileşiklerin moleküler aktivitesinin saptanmasına yardımcı olan hesaplama araçlarıyla yönlendirilen Python tabanlı sistem	<a href="https://github.com/Merck/DeepNeuralNet-QSAR">https://github.com/Merck/DeepNeuralNet-QSAR</a>
ORGANIC	İstenilen özelliklere sahip moleküller oluşturmaya yardımcı olan bir moleküler oluşturma aracı	<a href="https://github.com/aspuru-guzik-group/ORGANIC">https://github.com/aspuru-guzik-group/ORGANIC</a>
PotentialNet	Ligandların bağlanma afinitesini tahmin etmek için NN'leri kullanır	<a href="https://pubs.acs.org/doi/full/10.1021/acscentsci.8b00507">https://pubs.acs.org/doi/full/10.1021/acscentsci.8b00507</a>
Hit Dexter	Biyokimyasal testlere yanıt verebilecek molekülleri tahmin etmek için ML tekniği	<a href="http://hitdexter2.zbh.uni-hamburg.de">http://hitdexter2.zbh.uni-hamburg.de</a>
DeltaVina	İlaç-ligand bağlanma afinitesini yeniden puanlamak için bir puanlama işlevi	<a href="https://github.com/chengwang88/deltavina">https://github.com/chengwang88/deltavina</a>
Neural graph fingerprint	Yeni moleküllerin özelliklerini tahmin etmeye yardımcı olur	<a href="https://github.com/HIPS/neural-fingerprint">https://github.com/HIPS/neural-fingerprint</a>
AlphaFold	Proteinlerin 3 boyutlu yapılarını tahmin eder	<a href="https://deepmind.com/blog/alphafold">https://deepmind.com/blog/alphafold</a>
Chemputer	Kimyasal sentez prosedürünün standart formatta raporlanmasına yardımcı olur	<a href="https://zenodo.org/record/1481731">https://zenodo.org/record/1481731</a>

### Yapay Zekâ Uygulamalarına İlişkin Dezavantajlar

Yapay zekâ teknolojisi, özellikle derin öğrenme yöntemi, farmasötik bilgileri (örneğin, QSAR ve kimyasal yapı) büyük miktarda farmasötik veriden öğrenmek için kullanılabilir. Öğrenilen bilgi daha sonra molekülü istenilen özelliklere sahip olarak tasarlamak, moleküler özellikleri optimize etmek ve molekülün klinik onay başarı oranını ilerletmek için uygulanabilir. Yapay zekâ teknolojisi, güçlü veri madenciliği yetenekleri sayesinde bilgisayar destekli ilaç tasarımına yeni bir soluk getirmiştir. Ancak bazı sorunlar mevcuttur [83].

Tıbbi araştırmalarda yapay zekânın karşılaştığı veri eksikliği, birlikte çalışabilirlik eksikliği ve boyutluluk gibi birçok zorluklar bulunmaktadır [26]. Veri eksikliği, yapay zekâyı uygulamak isteyen her sektörde tekrar eden bir sorundur. Geleneksel bir biyolojik çalışmanın geçerli olması için minimum örnek sayısı beş olmalıdır. Bununla birlikte, çoğu makine öğrenmesi algoritmasının iyi performans göstermesi için yüzlerce veya binlerce veri noktası / örneği üzerinde eğitilmesi gerekir. Ayrıca,

etiketlenmiş verileri elde etmek zor olabilir çünkü bu genellikle bir tür manuel girdi gerektirir. Bazı durumlarda istenen verileri içeren büyük veritabanları mevcut olmayabilir. Bununla mücadele etmek için stratejilerden biri veri artırmadır. Veri artırma, gerçek verilerden yapay veri oluşturma sürecidir. Çeşitli veri artırma yaklaşımları vardır; nihayetinde yeni veri toplamadan eğitim modelleri için mevcut verileri artırır [26]. Diğer bir yaklaşım transfer öğrenme teknolojisinin geliştirilmesi olabilir [83].

Tahmine dayalı modeller oluşturmak için deneysel verileri kullanırken bir başka zorluk ve belirsizlik kaynağı, verilerin yanlışlıkla doğru bildirilmemesidir. Yanlış raporlama, bilimsel literatürde rapor edilen ve veri tabanlarında saklanan rapor edilen değerlerde, gen tanımlayıcılarında, birimlerde veya diğer parametrelerde basit yazım hataları şeklinde olabilir. Yanlış bildirilen bir veya iki veri noktası bile tahmine dayalı bir modelin sonuçlarını çarpıtma potansiyeline sahiptir. Veri iyileştirme ve veri raporlamadaki olası hataların belirlenmesi, yapay zekânın ilaç keşfiyle ilgili olabileceği başka bir alandır [84].

Makine öğrenmesinin karşı karşıya kaldığı bir diğer zorluk da modelin hala gizemini koruyor olmasıdır. Uygulanabilirlik alanlarının ortaya çıkmasına benzer şekilde, yeni AI araçlarını doğrulamak ve değerlendirmek için hangi uygun kontrollerin olduğu konusunda topluluk arasında bir fikir birliğine varılması gerekir [85]. 'Kara kutu modeli' terimi genellikle bir modelin nasıl belirli tahminlerde bulunduğunu ve performans gösterdiğini açıklamak zor olduğunda kullanılır. Çoğu makine öğrenmesi modeli, tahmin edilen sonuçların altında yatan biyolojik mekanizmalar hakkında bilgi sahibi olmadan veya bunlara dikkat etmeden tahmin doğruluğu için optimize edilmiş "kara kutular"dır [86, 87]. Her katman modele karmaşıklık kattığından, her katmanın çıktılarını katlanarak karmaşık hale getirebileceğinden ve katman sayısı arttığından, bu durum daha olasıdır. Ancak, Yerel Yorumlanabilir Model-Agnostik Açıklamalar (Local Interpretable Model-agnostic Explanations (LIME)) [88] ve SHapley Katkı Açıklamaları (Shapely Additive Explanations (SHAP)) gibi daha fazla açıklanabilirliği aydınlatmak için çeşitli araçlar geliştirilmektedir [89]. LIME, modelin davranışının yerel doğrusal bir yaklaşımını benimser, SHAP ise model çıktısını açıklamak için oyun teorisine dayalı bir yaklaşım kullanır. Hem LIME hem de SHAP ve diğer benzer stratejilerin makine öğrenmesinde yaygın uygulama haline geleceği ve kliniğe daha fazla AI teknolojisi getirmek için gerekli olacağı öngörülmektedir [26].

Tıbbi verilerde yapay zekâ ile tekrar eden bir sorun, boyutluluk laneti olarak bilinir. Bu, kullanılan veri setlerinin az sayıda örneğe ve birçok özelliğe sahip olduğu zamandır. Bu, tipik olarak binlerce özellik ve 100'den az örnek verdikleri için tıbbi omik veri setlerinde yaygın bir durumdur; böylece mevcut veriler seyrekleşir. Bu sorun, çeşitli boyut indirgeme teknikleriyle ele alınabilir [26]. Sinir ağı modellerini eğitmek birçok parametrenin ayarlanmasını içerir, ancak yalnızca birkaç pratik kılavuz vardır ve bu modellerin optimizasyonları için eksiksiz teorik sistemler hala ulaşılamazdır [83].

## SONUÇ VE TARTIŞMA

Hesaplama gücünün hızlı gelişimi ve hesaplamalı kimya ve biyolojinin hızlı gelişimi sayesinde, bilgisayar destekli ilaç tasarım teknikleri, araştırma sürecini hızlandırmak için ilaç keşif ve geliştirme hattının hemen hemen her aşamasında başarıyla uygulanmıştır [83]. Ana hedef prelinik ve klinik araştırmalarla ilgili maliyeti ve riski azaltmaktır. Yapay zekâ teknolojisi, makine öğrenimi teorisinin gelişimi ve farmakolojik verilerin birikmesi sayesinde, güçlü bir veri madenciliği aracı olarak, sanal tarama, aktivite öngörme, QSAR analizi de nova ilaç tasarımı ve *in silico* absorpsiyon, dağılma, metabolizma, atılım ve toksisite (ADME/T) özelliklerinin değerlendirilmesi gibi ilaç tasarımının çeşitli alanlarında rol oynamaktadır. Yapay zekâ tabanlı modellerin fiziksel bir açıklamasını sağlamak hala zor olsa da aslında çok yönlü sistemler aracılığıyla ilaç keşfinin manipüle edilmesine yardımcı olmak için büyük bir güç olarak hareket etmektedir. Son zamanlarda, güçlü tümevarım yeteneği ve güçlü özellik ekstraksiyon yeteneği nedeniyle, moleküler özelliklerin öngörülmesinin yanı sıra, ilaç tasarımı alanında AI teknolojilerinin uygulanmasını daha da teşvik edecek olan istenen molekülleri oluşturmada derin öğrenme yöntemleri kullanılmaktadır.

Yapay zekâ, ilaç keşfinin ilk aşamalarından başlayarak tamamıyla *in silico* ortamda istenen aktiviteyi gösteren yeni öncü bileşikler geliştirmek için kullanılmaktadır [90]. Hesaplamalı *de novo* tasarımı ile yapay zekâyı birleştirmek, bilinen yararlı bileşiklerin öğrenilmesine ve planlı bir biyolojik aktivite ile kimyasal olarak doğru ve sentezlenebilir yapıların üretilmesine olanak sağlayabilmektedir [91]. Bu uygulamalar sayesinde emek, zaman, maliyet ve bunun gibi birçok konuda tasarruf ve avantaj sağlanmaktadır. Yapay zekâ uygulamalarının giderek yaygınlaşmasının sebebi bu avantaj ve tasarruflardır.

Şu anda, birçok ilaç şirketi ilaç geliştirme programlarında artan maliyetler ve verimliliğin azalması nedeniyle zorluklarla karşılaşmaktadır [3]. Son zamanlarda, bu süreçleri daha maliyet ve zaman açısından verimli hale getirebilecek birçok etkileyici yapay zekâ yöntemi ve aracı geliştirilmiştir. İlaç taramasında AI / ML kullanımı buna bir örnektir. Geleneksel yüksek verimli tarama kütüphanesi genellikle yaklaşık bir milyon bileşik içerir ve her bir bileşik tipik olarak 50-100 Amerikan Doları tutarındadır. Bu nedenle, bir ilk tarama süreci birkaç milyon Amerikan dolarına ve birkaç aylık çalışmaya mal olabilir. Sonraki öncü bileşik optimizasyonunun, prelinik ilaç adaylarının belirlenmesi birkaç yıl sürebilir. Aksine, AI yardımıyla, birkaç milyar molekülden oluşan sanal bir bileşik kitaplığı birkaç gün içinde taranabilir. Yapay zekâ tabanlı bir hesaplama hattı kullanarak klinik öncesi adayları belirlemek yalnızca birkaç ay ile 1 yıl sürebilir [88, 92].

Yapay zekâ tabanlı hesaplamalı yaklaşımların ilaç geliştirme üzerindeki büyük etkisi göz önüne alındığında, bu alandaki yönelimlerin ve bu alana başvuranların sayısı hızla artmaktadır. Dahası, birçok

ilaç şirketi, 2017'den beri yapay zekâ girişimleri ve akademik kurumlarla iş birliği yapmanın yanı sıra dahili yapay zekâ tabanlı Ar-Ge programlarına yatırım yapmaktadır.

Bugüne kadar, yapay zekânın klinik deneyler için hızlı bir şekilde kullanılabilir hale getirilebilecek ve onaylanırsa sağlık hizmetleri ile birleştirilebilecek yeni aday tedavileri belirleme potansiyeli benzersiz olup yapay zekâyı ileri teknolojilerin merkezi haline getirmektedir. Bu nedenle AI, insan hastalıkları, özellikle de Koronavirüs Hastalığı 2019 (Coronavirus Disease 2019 (COVID-19)) gibi yeni ortaya çıkan hastalıklar için ilaç geliştirme ve mevcut ilaçların yeniden kullanılması için umut vericidir [93].

Son olarak, DL yöntemlerinin hala "karanlık bir sır" veya "kara kutu" olduğunu vurgulamak önemlidir [88]. Eğitim aşamasında, bir sinir ağına yalnızca etiketli belirli bir girdi verilir. Özellikler açıkça belirtilmemiştir ve ağı oluşturan kişi bile ara aşamalarda neyin incelendiğini veya modelin neden belirli bir sonuca vardığını bilemeyebilir [94]. Sonuç olarak, ilaç keşif döngüsünü hızlandırıp yapay zekâ araçlarını sisteme dahil etmek için büyük miktarda çalışma yapılmıştır; ancak ilaç keşfinde yapay zekânın tam potansiyeli gerçekleştirilmeden önce bu araçların daha başarılı uygulamaları gerekli olacaktır [39].

## YAZAR KATKILARI

Kavram: *R.T., İ.Y.*; Tasarım: *İ.N.Ç., F.K.A., R.T., İ.Y.*; Denetim: *İ.N.Ç., F.K.A., İ.Y.*; Kaynaklar: -; Malzemeler: -; Veri toplama ve/veya işleme: -; Analiz ve/veya yorumlama: *İ.N.Ç., İ.Y.*; Literatür taraması: *İ.N.Ç., F.K.A., R.T.*; Makalenin yazılması: *İ.N.Ç., İ.Y.*; Kritik inceleme: *İ.Y.*; Diğer: -

## ÇIKAR ÇATIŞMASI BEYANI

Yazarlar bu yazı için gerçek, potansiyel veya algılanan çıkar çatışması olmadığını beyan etmişlerdir.

## KAYNAKLAR

1. Schneider, P., Walters, W. P., Plowright, A. T., Sieroka, N., Listgarten, J., Goodnow, R. A., Fisher, J., Jansen, J. M., Duca, J. S., Rush, T. S., Zentgraf, M., Hill, J. E., Krutoholow, E., Kohler, M., Blaney, J., Funatsu, K., Luebke, C., Schneider, G. (2020). Rethinking Drug Design in The Artificial Intelligence Era. *Nature Reviews Drug Discovery*, 19(5), 353–364. <https://doi.org/10.1038/s41573-019-0050-3>
2. Lo, Y. C., Ren, G., Honda, H., Davis, K. L. (2019). Artificial Intelligence-Based Drug Design and Discovery. *ChemInformatics and Its Applications. Drug Discovery Today*. <http://dx.doi.org/10.5772/intechopen.89012>

3. Mak, K. K., Pichika, M. R. (2019). Artificial intelligence in drug development: present status and future prospects. *Drug Discovery Today*, 24(3), 773–780. <https://doi.org/10.1016/j.drudis.2018.11.014>
4. AI for Chemistry Web Site. Retrieved December 20, 2020, from <https://chemintelligence.com/ai-for-chemistry>
5. McCarthy, J., Hayes, P. (1969). Some Philosophical Problems From the Standpoint of Artificial Intelligence. In *Machine Intelligence*; Edinburgh University Press: Edinburgh, Retrieved from <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.85.5082>
6. Yang, X., Wang, Y., Byrne, R., Schneider, G., Yang, S. (2019). Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. *Chemical Reviews*, 119(18), 10520–10594. <https://doi.org/10.1021/acs.chemrev.8b00728>
7. Barr, A.; Feigenbaum, E. A.; Cohen, P. R. (1982). *Handbook of Artificial Intelligence*; Addison-Wesley Longman: Boston, MA, USA.
8. Popovic, D., Bhatkar, V. P. (1994). *Methods and Tools for Applied Artificial Intelligence*; Marcel Dekker: New York.
9. Bobrow, D. G. (1964). Natural Language Input for a Computer Problem Solving System. In *Semantic Information Processing*; MIT Press: Cambridge.
10. Weizenbaum, J. (1966). ELIZA---a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/365153.365168>
11. Baum, E. B. (1988). On the capabilities of multilayer perceptrons. *Journal of Complexity*, 4(3), 193–215. [https://doi.org/10.1016/0885-064X\(88\)90020-9](https://doi.org/10.1016/0885-064X(88)90020-9)
12. Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986). Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press: Cambridge.
13. Qian, N., Sejnowski, T. J. (1988). Predicting the secondary structure of globular proteins using neural network models. *Journal of Molecular Biology*, 202(4), 865–884. [https://doi.org/10.1016/0022-2836\(88\)90564-5](https://doi.org/10.1016/0022-2836(88)90564-5)
14. Hammett, L. P. (1937). The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *Journal of the American Chemical Society*, 59(1), 96–103. <https://doi.org/10.1021/ja01280a022>
15. Hansch, C., Fujita, T. (1964).  $\rho$ - $\sigma$ - $\pi$  Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *Journal of the American Chemical Society*, 86(8), 1616–1626. <https://doi.org/10.1021/ja01062a035>
16. Miller, E., Hansch, C. (1967). Structure-Activity Analysis of Tetrahydrofolate Analogs Using Substituent Constants and Regression Analysis. *Journal of Pharmaceutical Sciences*, 56(1), 92–97. <https://doi.org/10.1002/jps.2600560119>

17. Kopecký, J., Boček, K., Vlachová, D. (1965). Chemical Structure and Biological Activity on m- and p-Disubstituted Derivatives of Benzene. *Nature*, 207(5000), 981–981. <https://doi.org/10.1038/207981a0>
18. Wessel, M. D., Jurs, P. C., Tolan, J. W., Muskal, S. M. (1998). Prediction of human intestinal absorption of drug compounds from molecular structure. *Journal of Chemical Information and Computer Sciences*, 38(4), 726–735. <https://doi.org/10.1021/ci980029a>
19. Martin Y. C. (2010). *Quantitative Drug Design: A Critical Introduction*. Boca Raton, FL: CRC Press. 2nd ed.
20. Basile, A. O., Yahi, A., Tatonetti, N. P. (2019). Artificial Intelligence for Drug Toxicity and Safety. *Trends in Pharmacological Sciences*, 40(9), 624–635. <https://doi.org/10.1016/j.tips.2019.07.005>
21. Zhu, H. (2020). Big data and artificial intelligence modeling for drug discovery. *Annual Review of Pharmacology and Toxicology*, 60(1), 573–589. <https://doi.org/10.1146/annurev-pharmtox-010919-023324>
22. Bunney, P. E., Zink, A. N., Holm, A. A., Billington, C. J., Kotz, C. M. (2017). Orexin activation counteracts decreases in nonexercise activity thermogenesis (NEAT) caused by high-fat diet. *Physiology & Behavior*, 176(1), 139–148. <https://doi.org/10.1016/j.physbeh.2017.03.040>
23. Properzi, F., Taylor, K., Steedman, M. (2019). Accelerating drug discovery. *Intelligent drug discovery powered by AI*. 2-7. Retrieved from <https://blogs.deloitte.co.uk/health/>
24. Panteleev, J., Gao, H., Jia, L. (2018). Recent applications of machine learning in medicinal chemistry. *Bioorganic and Medicinal Chemistry Letters*, 28(17), 2807–2815. <https://doi.org/10.1016/j.bmcl.2018.06.046>
25. D'Souza, S., Prema, K. V., Balaji, S. (2020). Machine learning models for drug–target interactions: current knowledge and future directions. *Drug Discovery Today*, 25(4), 748–756. <https://doi.org/10.1016/j.drudis.2020.03.003>
26. Linton-Reid, K. (2020). Introduction: An Overview of AI in Oncology Drug Discovery and Development. *Artificial Intelligence in Oncology Drug Discovery and Development*, (M1), 1–13. <https://doi.org/10.5772/intechopen.92799>
27. Ippolito, M., Ferguson, J., Jenson, F. (2020). Improving facies prediction by combining supervised and unsupervised learning methods. *Journal of Petroleum Science and Engineering*, 200, 108300. <https://doi.org/10.1016/j.petrol.2020.108300>
28. Civelek, Ö. (2003). Bulanık Mantık Nedir Yapay Zekâ Nedir. *Türkiye Mühendislik Haberleri Dergisi*, 423(1), 40–50.
29. Bohr, H. (2020). Drug discovery and molecular modeling using artificial intelligence. *In Artificial Intelligence in Healthcare*, pp. <https://doi.org/10.1016/b978-0-12-818438-7.00003-4>
30. Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., Tekade, R. K. (2020). Artificial intelligence in drug discovery and development. *Drug Discovery Today*, 26(1), 80–93. <https://doi.org/10.1016/j.drudis.2020.10.010>

31. Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug Discovery Today*, 23(6), 1241–1250. <https://doi.org/10.1016/j.drudis.2018.01.039>
32. Jing, Y., Bian, Y., Hu, Z., Wang, L., Xie, X. Q. S. (2018). Deep Learning for Drug Design: an Artificial Intelligence Paradigm for Drug Discovery in the Big Data Era. *AAPS Journal*, 20(3), 58. <https://doi.org/10.1208/s12248-018-0210-0>
33. Zhavoronkov, A. (2018). Artificial Intelligence for Drug Discovery, Biomarker Development, and Generation of Novel Chemistry. *Molecular Pharmaceutics*, 15(10), 4311–4313. <https://doi.org/10.1021/acs.molpharmaceut.8b00930>
34. Gunavathi, C., Sivasubramanian, K., Keerthika, P., Paramasivam, C. (2020). A review on convolutional neural network based deep learning methods in gene expression data for disease diagnosis. *Materials Today: Proceedings*. <https://doi.org/10.1016/j.matpr.2020.10.263>
35. Hubel, D. H., Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1), 106–154. <https://doi.org/10.1113/jphysiol.1962.sp006837>
36. Hubel, D. H., Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148(3), 574–591. <https://doi.org/10.1113/jphysiol.1959.sp006308>
37. Bilski, J., Rutkowski, L., Smolağ, J., Tao, D. (2021). A novel method for speed training acceleration of recurrent neural networks. *Information Sciences*, 553, 266–279. <https://doi.org/10.1016/j.ins.2020.10.025>
38. Big pharma is using AI and machine learning in drug discovery and development to save lives Web Site. Retrieved December 20, 2020, from <https://www.businessinsider.com/ai-machine-learning-in-drug-discovery-development-2020>
39. Chan, H. C. S., Shan, H., Dahoun, T., Vogel, H., Yuan, S. (2019). Advancing Drug Discovery via Artificial Intelligence. *Trends in Pharmacological Sciences*, 40(8), 592–604. <https://doi.org/10.1016/j.tips.2019.06.004>
40. Rubio, D. M. G., Schoenbaum, E. E., Lee, L. S., Schteingart, D. E., Marantz, P. R., Anderson, K. E., Platt, L. D., Baez, A., Esposito, K. (2010). Defining translational research: Implications for training. *Academic Medicine*, 85(3), 470–475. <https://doi.org/10.1097/ACM.0b013e3181ccd618>
41. Merck Molecular Activity Challenge. Retrieved December 20, 2020, from <https://www.kaggle.com/c/MerckActivity>
42. DeepCodex: a deep code for gene expression data. Retrieved December 20, 2020, from <http://deepcodex.org>
43. Donner, Y., Kazmierczak, S., Fortney, K. (2018). Drug Repurposing Using Deep Embeddings of Gene Expression Profiles. *Molecular Pharmaceutics*, 15(10), 4314–4325. <https://doi.org/10.1021/acs.molpharmaceut.8b00284>
44. Xie, L., He, S., Song, X., Bo, X., Zhang, Z. (2018). Deep learning-based transcriptome data classification for drug-target interaction prediction. *BMC Genomics*, 19(7), 93-102.

<https://doi.org/10.1186/s12864-018-5031-0>

45. Vanhaelen, Q., Mamoshina, P., Aliper, A. M., Artemov, A., Lezhnina, K., Ozerov, I., Labat, I., Zhavoronkov, A. (2017). Design of efficient computational workflows for in silico drug repurposing. *Drug Discovery Today*, 22(2), 210–222. <https://doi.org/10.1016/j.drudis.2016.09.019>
46. Aliper, A., Jellen, L., Cortese, F., Artemov, A., Semper, D. K., Moskalev, A., Swick, A. G., Zhavoronkov, A. (2017). Towards natural mimetics of metformin and rapamycin. *Aging*, 9(11), 2245–2268. <https://doi.org/10.18632/aging.101319>
47. Gayvert, K. M., Madhukar, N. S., Elemento, O. (2016). A Data-Driven Approach to Predicting Successes and Failures of Clinical Trials. *Cell Chemical Biology*, 23(10), 1294–1301. <https://doi.org/10.1016/j.chembiol.2016.07.023>
48. Mayr, A., Klambauer, G., Unterthiner, T., Hochreiter, S. (2016). DeepTox: Toxicity prediction using deep learning. *Frontiers in Environmental Science*, 3(80). <https://doi.org/10.3389/fenvs.2015.00080>
49. Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., Terentiev, V. A., Polykovskiy, D. A., Kuznetsov, M. D., Asadulaev, A., Volkov, Y., Zholus, A., Shayakhmetov, R. R., Zhebrak, A., Minaeva, L. I., Zagribelnyy, B. A., Lee, L. H., Soll, R., Madge, D., Xing, L., Guo, T., Aspuru-Guzik, A. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, 37(9), 1038–1040. <https://doi.org/10.1038/s41587-019-0224-x>
50. Hessler, G., Baringhaus, K. H. (2018). Artificial intelligence in drug design. *Molecules*, 23(10), 2520. <https://doi.org/10.3390/molecules23102520>
51. Riddick, G., Song, H., Ahn, S., Walling, J., Borges-Rivera, D., Zhang, W., Fine, H. A. (2011). Predicting in vitro drug sensitivity using random forests. *Bioinformatics*, 27(2), 220–224. <https://doi.org/10.1093/bioinformatics/btq628>
52. Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R., Menden, M. P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., Cokelaer, T., Greninger, P., Syk, E. V., Chang, H., Silva, H. D., Heyn, H., Deng, X., Egan, R.K., Liu, Q., Mironenko, T., Mitropoulos, X., Richardson, L., Wang, J., Zhang, T., Moran, S., Sayols, S., Soleimani, M., Tamborero, D., Lopez-Bigas, N., Ross-Macdonald, P., Esteller, M., Gray, N. S., Haber, D. A., Stratton, M. R., Benes, C. H., Wessels, L. F. A., Saez-Rodriguez, J., McDermott, U., Garnett, M. J. (2016). A Landscape of Pharmacogenomic Interactions in Cancer. *Cell*, 166(3), 740–754. <https://doi.org/10.1016/j.cell.2016.06.017>
53. Cortés-Ciriano, I., Van Westen, G. J. P., Bouvier, G., Nilges, M., Overington, J. P., Bender, A., Malliavin, T. E. (2016). Improved large-scale prediction of growth inhibition patterns using the NCI60 cancer cell line panel. *Bioinformatics*, 32(1), 85–95. <https://doi.org/10.1093/bioinformatics/btv529>
54. Tetko, I. V., Bruneau, P. (2004). Application of ALOGPS to predict 1-octanol/water distribution coefficients, logP, and logD, of AstraZeneca in-house database. *Journal of Pharmaceutical Sciences*, 93(12), 3103–3110. <https://doi.org/10.1002/jps.20217>
55. Lusci, A., Pollastri, G., Baldi, P. (2013). Deep architectures and deep learning in chemoinformatics: The prediction of aqueous solubility for drug-like molecules. *Journal of*



- Chemical Information and Modeling*, 53(7), 1563–1575. <https://doi.org/10.1021/ci400187y>
56. Koscielny, G., An, P., Carvalho-Silva, D., Cham, J. A., Fumis, L., Gasparyan, R., Hasan, S., Karamanis, N., Maguire, M., Papa, E., Pierleoni, A., Pignatelli, M., Platt, T., Rowland, W., Wankar, P., Pento, A. P., Burdett, T., Fabregat, A., Forbes, S., Gaulton, A., Gonzalez, C. Y., Hermjakob, H., Hersey, A., Jupe, S., Kafkas, Ş., Keays, M., Leroy, C., Lopez, F. J., Magarinos, M. P., Malone, J., McEntyre, J., Fuentes, A. M. P., O'Donovan, C., Papatheodorou, I., Parkinson, H., Palka, B., Paschall, J., Petryszak, R., Pratanwanich, N., Sarntivijal, S., Saunders, G., Sidiropoulos, K., Smith, T., Sondka, Z., Stegle, O., Tang, Y. A., Turner, E., Vaughan, B., Vrousou, O., Watkins, X., Martin, M. J., Sanseau, P., Vamathevan, J., Birney, E., Barrett, J., Dunham, I. (2017). Open Targets: A platform for therapeutic target identification and Validation. *Nucleic Acids Research*, 45(1), 985–994. <https://doi.org/10.1093/nar/gkw1055>
  57. Ferrero, E., Dunham, I., Sanseau, P. (2017). In silico prediction of novel therapeutic targets using gene-disease association data. *Journal of Translational Medicine*, 15(1), 1–16. <https://doi.org/10.1186/s12967-017-1285-6>
  58. Cavasotto, C. N., Di Filippo, J. I. (2021). Artificial intelligence in the early stages of drug discovery. *Archives of Biochemistry and Biophysics*, 698, 108730. <https://doi.org/10.1016/j.abb.2020.108730>
  59. Plante, A., Shore, D. M., Morra, G., Khelashvili, G., Weinstein, H. (2019). A Machine Learning Approach for the Discovery of Ligand-Specific Functional Mechanisms of GPCRs. *Molecules*, 24(11), 2097. <http://dx.doi.org/10.3390/molecules24112097>
  60. Díaz, Ó., Dalton, J. A. R., Giraldo, J. (2019). Artificial Intelligence: A Novel Approach for Drug Discovery. *Trends in Pharmacological Sciences*, 40(8), 550–551. <https://doi.org/10.1016/j.tips.2019.06.005>
  61. Ferraro, M., Decherchi, S., De Simone, A., Recanatini, M., Cavalli, A., Bottegoni, G. (2020). Multi-target dopamine D3 receptor modulators: Actionable knowledge for drug design from molecular dynamics and machine learning. *European Journal of Medicinal Chemistry*, 188, 111975. <https://doi.org/10.1016/j.ejmech.2019.111975>
  62. Beck, B. R., Shin, B., Choi, Y., Park, S., Kang, K. (2020). Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Computational and Structural Biotechnology Journal*, 18, 784–790. <https://doi.org/10.1016/j.csbj.2020.03.025>
  63. Green, D. V. S., Pickett, S., Luscombe, C., Senger, S., Marcus, D., Meslamani, J., Brett, D., Powell, A., Masson, J. (2020). BRADSHAW: a system for automated molecular design. *Journal of Computer-Aided Molecular Design*, 34(7), 747–765. <https://doi.org/10.1007/s10822-019-00234-8>
  64. Camodeca, C., Nuti, E., Tepshi, L., Boero, S., Tuccinardi, T., Stura, E. A., Poggi, A., Zocchi, M. R., Rossello, A. (2016). Discovery of a new selective inhibitor of A Disintegrin and Metalloprotease 10 (ADAM-10) able to reduce the shedding of NKG2D ligands in Hodgkin's lymphoma cell models. *European Journal of Medicinal Chemistry*, 111, 193–201. <https://doi.org/10.1016/j.ejmech.2016.01.053>
  65. Healy, E. F., Romano, P., Mejia, M., Lindfors, G. (2010). Acetylenic inhibitors of ADAM10 and ADAM17: In silico analysis of potency and selectivity. *Journal of Molecular Graphics and*

- Modelling*, 29(3), 436–442. <https://doi.org/10.1016/j.jmgm.2010.08.006>
66. Tippmann, F., Hundt, J., Schneider, A., Endres, K., Fahrenholz, F. (2009). Up-regulation of the  $\alpha$ -secretase ADAM10 by retinoic acid receptors and acitretin. *The FASEB Journal*, 23(6), 1643–1654. <https://doi.org/10.1096/fj.08-121392>
67. Altmeppen, H. C., Prox, J., Krasemann, S., Puig, B., Kruszewski, K., Dohler, F., Bernreuther, C., Hoxha, A., Linsenmeier, L., Sikorska, B., Liberski, P. P., Bartsch, U., Saftig, P., Glatze, M. (2015). The sheddase ADAM10 is a potent modulator of prion disease. *ELife*, 2015(4), 1–50. <https://doi.org/10.7554/eLife.04260>
68. Kohutek, Z. A., DiPierro, C. G., Redpath, G. T., Hussaini, I. M. (2009). ADAM-10-Mediated N-Cadherin Cleavage Is Protein Kinase C- $\alpha$  Dependent and Promotes Glioblastoma Cell Migration. *Journal of Neuroscience*, 29(14), 4605–4615. <https://doi.org/10.1523/JNEUROSCI.5126-08.2009>
69. Woods, N., Trevino, J., Coppola, D., Chellappan, S., Yang, S., Padmanabhan, J. (2015). Fendiline inhibits proliferation and invasion of pancreatic cancer cells by interfering with ADAM10 activation and  $\beta$ -catenin signaling. *Oncotarget*, 6(34), 35931–35948. <https://doi.org/10.18632/oncotarget.5933>
70. Shi, T., Huang, S., Chen, L., Heng, Y., Kuang, Z., Xu, L., Mei, H. (2020). A molecular generative model of ADAM10 inhibitors by using GRU-based deep neural network and transfer learning. *Chemometrics and Intelligent Laboratory Systems*, 205, 104122. <https://doi.org/10.1016/j.chemolab.2020.104122>
71. Segler, M. H. S., Kogej, T., Tyrchan, C., Waller, M. P. (2018). Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Central Science*, 4(1), 120–131. <https://doi.org/10.1021/acscentsci.7b00512>
72. Luo, J. (2016). CRISPR/Cas9: From Genome Engineering to Cancer Drug Discovery. *Trends in Cancer*, 2(6), 313–324. <https://doi.org/10.1016/j.trecan.2016.05.001>
73. Scott, A. (2018). A CRISPR path to drug discovery. *Nature*, 555, 10–11. <https://doi.org/10.1038/d41586-018-02477-1>
74. Wallach, I., Dzamba, M., Heifets, A. (2015). *AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery*. 1–11. Retrieved from <http://arxiv.org/abs/1510.02855>
75. Spitzer, R., Jain, A. N. (2012). Surflex-Dock: Docking benchmarks and real-world application. *Journal of Computer-Aided Molecular Design*, 26(6), 687–699. <https://doi.org/10.1007/s10822-011-9533-y>
76. Allen, W. J., Balius, T. E., Mukherjee, S., Brozell, S. R., Moustakas, D. T., Lang, P. T., Case, D. A., Kuntz, I. D., Rizzo, R. C. (2015). DOCK 6: Impact of new features and current docking performance. *Journal of Computational Chemistry*, 36(15), 1132–1156. <https://doi.org/10.1002/jcc.23905>
77. Kuenzi, B. M., Park, J., Fong, S. H., Sanchez, K. S., Lee, J., Kreisberg, J. F., Ma, J., Ideker, T. (2020). Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer

- Cells. *Cancer Cell*, 38(5), 672-684. <https://doi.org/10.1016/j.ccell.2020.09.014>
78. Hasan Mahmud, S. M., Chen, W., Jahan, H., Dai, B., Din, S. U., Dzisoo, A. M. (2020). DeepACTION: A deep learning-based method for predicting novel drug-target interactions. *Analytical Biochemistry*, 610, 113978. <https://doi.org/10.1016/j.ab.2020.113978>
79. Wan, F., Zhu, Y., Hu, H., Dai, A., Cai, X., Chen, L., Gong, H., Xia, T., Yang, D., Wang, M. W., Zeng, J. (2019). DeepCPI: A Deep Learning-based Framework for Large-scale in silico Drug Screening. *Genomics, Proteomics and Bioinformatics*, 17(5), 478-495. <https://doi.org/10.1016/j.gpb.2019.04.003>
80. Lagunin, A., Zakharov, A., Filimonov, D., Poroikov, V. (2011). QSAR modelling of rat acute toxicity on the basis of PASS prediction. *Molecular Informatics*, 30(2-3), 241-250. <https://doi.org/10.1002/minf.201000151>
81. Soufan, O., Ba-Alawi, W., Afeef, M., Essack, M., Kalnis, P., Bajic, V. B. (2016). DRABAL: novel method to mine large high-throughput screening assays using Bayesian active learning. *Journal of Cheminformatics*, 8(1), 1-14. <https://doi.org/10.1186/s13321-016-0177-8>
82. Malandraki-Miller, S., Riley, P. R. (2021). Use of artificial intelligence to enhance phenotypic drug discovery. *Drug Discovery Today*. <https://doi.org/10.1016/j.drudis.2021.01.013>
83. Zhong, F., Xing, J., Li, X., Liu, X., Fu, Z., Xiong, Z., Lu, D., Wu, X., Zhao, J., Tan, X., Li, F., Luo, X., Li, Z., Chen, K., Zheng, M., Jiang, H. (2018). Artificial intelligence in drug design. *Science China Life Sciences*, 61(10), 1191-1204. <https://doi.org/10.1007/s11427-018-9342-2>
84. Kalliokoski, T., Kramer, C., Vulpetti, A., Gedeck, P. (2013). Comparability of Mixed IC50 Data -A Statistical Analysis. *PLoS ONE*, 8(4), 1-11. <https://doi.org/10.1371/journal.pone.0061007>
85. Solomon, S. M. (2020). Genome editing in animals: why FDA regulation matters. *Nature Biotechnology*, 38(2), 142-143. <https://doi.org/10.1038/s41587-020-0413-7>
86. Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., Agapow, P.M., Zietz, M., Hoffman, M. M., Xie, W., Rosen, G. L., Lengerich, B.J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A. E., Shrikumar, A., Xu, J., Cofer, E.M. Lavender, C.A., Turaga, S. C., Alexandari, A. M., Lu, Z., Harris, D. J., DeCaprio, D., Qi, Y., Kundaje, A., Peng, Y., Wiley, L. K., Segler, M. H. S., Boca, S. M., Swamidass, S. J., Huang, A., Gitter, A., Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*, 15(141). <https://doi.org/10.1101/142760>
87. Carlini, N., Liu, C., Erlingsson, Ú., Kos, J., Song, D. (2019). The secret Sharer: Evaluating and testing unintended memorization in neural networks. In *Proceedings of the 28th USENIX Security Symposium*, (pp. 267-284). Santa Clara, CA, USA.
88. Voosen, P. (2017). The AI detectives. *Science*, 357(6346), 22-27. <https://doi.org/10.1126/science.357.6346.22>
89. Tishby, N., Zaslavsky, N. (2015). Deep learning and the information bottleneck principle. In *Proceedings of the 2015 IEEE Information Theory Workshop (ITW)*, (pp. 1-5). Jeju Island, Korea. <https://doi.org/10.1109/ITW.2015.7133169>
90. Merk, D., Friedrich, L., Grisoni, F., Schneider, G. (2018). De Novo Design of Bioactive Small Molecules by Artificial Intelligence. *Molecular Informatics*, 37(1-2).

<https://doi.org/10.1002/minf.201700153>

91. Lake, F. (2019). Artificial intelligence in drug discovery: what is new, and what is next? *Future Drug Discovery*, 1(2). <https://doi.org/10.4155/fdd-2019-0025>
92. Hassan Baig, M., Ahmad, K., Roy, S., Mohammad Ashraf, J., Adil, M., Haris Siddiqui, M., Khan, S., Amjad Kamal, M., Provazník, I., Choi, I. (2016). Computer Aided Drug Design: Success and Limitations. *Current Pharmaceutical Design*, 22(5), 572–581. <https://doi.org/10.2174/1381612822666151125000550>
93. Zhou, Y., Wang, F., Tang, J., Nussinov, R., Cheng, F. (2020). Artificial intelligence in COVID-19 drug repurposing. *The Lancet Digital Health*, 2(12), 667–676. [https://doi.org/10.1016/S2589-7500\(20\)30192-8](https://doi.org/10.1016/S2589-7500(20)30192-8)
94. Lecun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>