

PROTEİNLERİN PEPTİT BAĞ AÇILARININ TAHMİNİNDE YAPAY SİNİR AĞLARI YÖNTEMİNİN KULLANILMASI

Murat DEMİR¹

Ali KARCI²

¹Fırat Üniversitesi, Muş Meslek Yüksek Okulu, Muş

² Fırat Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü , Elazığ

¹ e-posta : murat2107@yahoo.com

² e-posta : akarci@firat.edu.tr

Anahtar sözcükler : Proteinler, Peptitler, Yapay sinir ağları

ÖZET

Bu çalışmada protein makro moleküllerinin ikincil yapılarının tahmini yapılmaya çalışılmıştır. Bunun için birincil yapıdaki amino asit zincir ardışıkları kullanılmıştır. Birincil yapıda bir zincirde bulunan sıralı amino asitlerden üçer-üçer ele alınarak mevcut polipeptitlerdeki amino asitlerin merkezil karbonları arasındaki iki boyutlu açılar kullanılmıştır. Bunun için protein data banklardan alınan gerçek açı değerleri ile bir yapay sinir ağına eğitim yapılarak, sonrasında farklı protein örnekleri üzerinde denenmiş ve sonuçta farklı test protein örneği üzerinde % 70, %68.42, %49.12 oranında başarı elde edildiği görülmüştür.

1. GİRİŞ

Protein “ilk element” (Yunanca’da “proteios”) anlamına gelir. Proteinler hücrelerdeki bütün biyolojik olayların yapıtaşdırlar. Hücreler içerisinde gerçekleşen olaylar, çok sayıda farklı proteinin vazifelerini yerine getirmeleri ile devam eder.

Proteinler birbirinden birincil yapı bakımından bakıldığında amino asitlerin dizilişi bakımından ayrılmaktadırlar. Bunun yanı sıra birincil yapıları aynı olsa dahi proteinler ikincil ve üçüncül yapıları bakımından farklılık gösterebilmektedirler . Bu farklılığı sağlayan ikincil ve üçüncül yapıdaki katlanmalardır. Bu katlanmalar sayesinde molekülün kararlılığı değişmektedir. Proteinin yapısıyla ilişkili olan kararlılık terimi, doğal konformasyonu (uzaysal düzenim) sürdürme eğilimi olarak tanımlanabilir.[1]

Bu çalışmada proteinlerdeki bu katlanmalar, birincil yapıdaki amino asitlerin aralarındaki açı değerlerini kullanarak yapılmaya çalışıldı. Zincir üzerindeki amino asitlerin aralarındaki açı değerlerinden faydalanarak bir proteinin ikincil yapısı ortaya koyulabileceğinden bu elde edilen değerler katlanma şeklini ortaya koyacaktır.

Şu ana kadar bu konuda yapay sinir ağları da dahil olmak üzere yapılmış olan bir çok çalışma vardır. Bunların bir çoğu katlanmanın şekli (sarım/halka veya helezon/sarmal) esas alınarak yapılmıştır.

Baldil [2] helezon (helezon) ve sarım (sarım) yapılar üzerinde çalışmıştır. BRNN (Bi-Directional Recurrent Neural Networks) ‘den Rost ve Sender örnekleri

üzerinde çalışılmıştır. Yaklaşık %75 başarı elde edilmiştir. Francesco [3] helezon , sarım ve geniş sarmal (Extended Strand) yapılar üzerinde çalışmıştır. Brook haven databank’ tan alınan 20 protein örneği kullanılmıştır. Yaklaşık %69 başarı elde edilmiştir. Nguyen [4] PSI-BLAST yöntemi ile RS126 ve CB396 örnekleri üzerinde helezon , sarım ve geniş sarmal yapıları üzerinde çalışmıştır. Yaklaşık %77 başarı elde edilmiştir. Wu [5] PSI-BLAST yöntemi ile helezon, geniş sarmal ve diğer yapılar üzerinde çalışmıştır. Yaklaşık %80 başarı elde edilmiştir. Liv [6] BLAST algoritması yöntemiyle RS126 ve CB513 protein örnekleri üzerinde çalışmıştır. Yaklaşık % 70 başarı elde edilmiştir. Aggorval [7] helezon, sarım, geniş sarmal yapılar üzerinde çalışmıştır. Yaklaşık % 76 başarı elde edilmiştir. Robles [8] EDA algoritması yardımıyla CB513 data seti üzerinde çalışmış ve ikincil yapının olasılık tahminleri üzerinde durmuştur. Yaklaşık % 70 oranında başarı elde edilmiştir. Kurt [9] kaba karar liste öğrenicisi algoritması CB513 data seti üzerine uygulamıştır. Kural tabanları oluşturularak işlem yapılmıştır. Yaklaşık %69 civarı bir başarı elde edilmiştir. Holley [10] protein yapı tahmini için bir yapay sinir ağı modeli kullanmıştır. 20 değişik amino asit ve birde sonlardaki null değer (üst üste binme durumlarında) ile beraber 21 giriş mevcut olup pencereleme metodu ile 17 pencerelik bir boyut seçilmiştir. Yapay sinir ağına girişten rasgele amino asitler gelmektedir. Pencere boyutu 17 olarak seçildiğinden ilk 17 amino asitten 9. sıradaki merkezde olacak şekilde 0; ilk 8 tanesi -8 ‘e kadar sonraki 8 de +8’e kadar numaralandırılır. Oluşturulan tabloda tüm amino asitler pozisyonlarına göre sıralanmakta ve alabilecekleri değerler tabloya yerleştirilmektedir.

Yapay sinir ağının iki çıkışı mevcuttur. Çıkış =(0,0) ise yapı sarım , Çıkış = (0,1) ise yapı düz, Çıkış=(1,0) ise yapı helezon durumundadır. Bu yöntem ile yaklaşık % 79 başarı elde edilmiştir. Ersöz [11] hemoglobin proteininin yapısını tahmin için birleştirilmiş yapay sinir ağları metodu kullanmıştır. Proteinlerin yapısında bulunan 20 farklı amino asit 20 bitlik dizge ile yapılarının helezon veya sarım oluşu ise 2 bit ile kodlanmıştır.

Alanine = 1000000000000000000 Helezon(H) = 1,2

Girişten yapay sinir ağlarına gelen amino asit ardışıkları pencereleme metoduna tabi tutulmaktadır. Pencerede ortadaki amino asit $W = 2*r+1$ formülü yardımıyla bulunarak r . sıraya denk gelen amino asit pencerenin ortası olup yapısına bakılıp sonra pencerede r . den sonra sıra gelen amino asit pencere ortası seçilip aynı işlem ardışıl sonuna kadar devam etmektedir. İşlem sonunda gerçekte olması gereken helezon ve sarım sayısı ile yapay sinir ağının bulunduğu helezon ve sarım sayısı oranlandığı zaman başarı oranı ortaya çıkmaktadır. Bu çalışmada farklı pencere boyutları için sonuçlar elde edilmiştir. 20 alfa-141 ve 20 beta-146 hemoglobinin zincirleri için elde edilen sonuçlar yaklaşık % 90 civarındadır.

Bu çalışmada ise amino asitlerin merkezi karbonları arasındaki açıların ne olduğunu tahmin etmeye çalışan bir yapay sinir ağı modeli kullanıldı. Yapay sinir ağı modeli olarak ileri beslemeli bir yapay sinir ağı modeli kullanılmıştır. Öğrenme algoritması olarak Levenberg-Marquardt öğrenme algoritması metodu seçilmiştir.

Kullanılan olan yapay sinir ağı 3 katmandan oluşmaktadır. İlk katman olan giriş katmanı 10 , orta katman 15, çıkış katmanı ise 1 nöron içermektedir. Giriş vektörümüzün boyutu da amino asitler üzer üzer ele alındığından 3'dür.

Yapay sinir ağına eğitim için protein data banklarından elde edilen 290 tane eğitim örneği verilerek yapay sinir ağının eğitimi tamamlanmıştır. Eğitim için kullanılan açılar, her biri üç amino asidin merkezil karbonları arasındaki açıyı gösteren bir açı değerini göstermektedir.

Eğitim sonunda yapay sinir ağı, protein data bank'lerden elde edilen 1a89, 1a9a, 1bbe proteinleri ile teste tabi tutulmuştur.

Değerlendirilme yapılırken açı değerleri sınıflara bölünmüştür. Olayın çok kompleks olmasından dolayı proteinlerde aynı dizilime sahip olan zincirlerde bile bazen açı değerlerinin diğer moleküllerin etkileşiminden kaynaklanan sapmalar olduğu görülür. Bu yüzden açı aralıkları seçilip çok fazla sapma olmadan yaklaşım yapılmaya çalışılmıştır. Sonuçta gerçek değerlere yaklaşık üç farklı örnek üzerinde % 70, %68.42, %49.12 oranında bir yaklaşım olduğu görülmüştür.

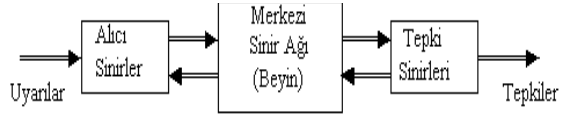
2. YAPAY SİNİR AĞLARI

Biyolojik sinir sistemi, merkezinde sürekli olarak bilgiyi alan, yorumlayan ve uygun bir karar üreten beynin bulunduğu 3 katmanlı bir yapıdan oluşmuştur.

Alıcı sinirler (receptor) organizma içerisinden ya da dış ortamlardan algıladıkları uyarıları, beyne bilgi ileten elektriksel sinyallere dönüştürür. Tepki sinirleri

(effector) ise, beynin ürettiği elektriksel darbeleri organizma çıktısı olarak uygun tepkilere dönüştürür.

Aşağıdaki şekilde bir sinir sisteminin blok diyagramı verilmiştir [12].

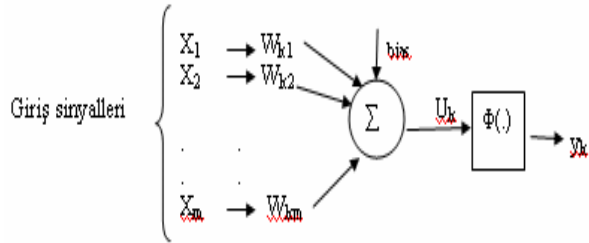


Şekil 1 sinir sistemi blok diyagramı

Merkezi sinir ağında bilgiler, alıcı ve tepki sinirleri arasında ileri ve geri besleme yönünde değerlendirilerek uygun tepkiler üretilir. Bu yönüyle biyolojik sinir sistemi, kapalı çevrim denetim sisteminin karakteristiklerini taşır.

Yapay sinir ağları da benzer mantıkla giriş olarak aldığı verileri işleyerek sonuçta bir çıktı üretir ve arada kapalı sistem bir yapı vardır. Bu yönüyle yapay sinir ağlarına kara kutu benzetmesi yapılmıştır. [13]

Yapay sinir ağları, insan beyninin özelliklerinden olan öğrenme yolu ile yeni bilgiler üretebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yetenekleri herhangi bir yardım almadan otomatik olarak gerçekleştirebilen bilgisayar sistemleridir. Örneklerden elde ettikleri bilgiler ile kendi deneyimlerini oluşturur ve daha sonra, benzer konularda benzer sonuçlar verebilecek hale gelebilirler [13]. Bir yapay sinir ağının modeli basitçe verilecek olursa, aşağıdaki gibi temsil edilebilir.



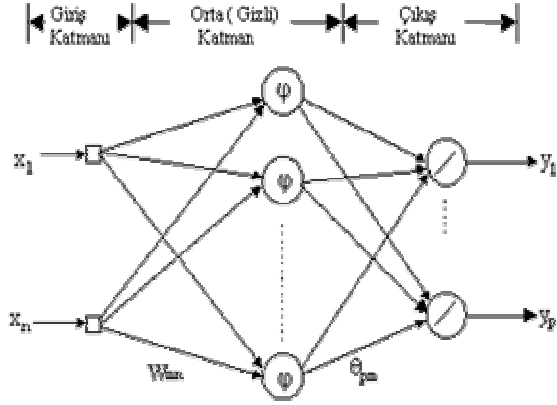
Şekil 3 yapay nöron modeli

X_1, X_2, \dots, X_m girişleri ve W_1, W_2, \dots, W_m ağırlıkları ifade etmektedir. Bias değeri olan U_k oryantasyonu (yönlendirme yada belli bir eğere takılıp kalmama) sağlamaktadır.

$$U_k = \sum_{j=1}^m w_{kj} \cdot x_j$$

$Y_k = (U_k + b_k)$ ifadesi ise bias eklenmiş ağırlıklı toplamın aktivasyon fonksiyonundan geçmiş durumudur.

Yapay sinir hücreleri bir araya gelerek yapay sinir ağlarını oluşturmaktadırlar. Bir çok farklı geliştirilmiş yapay sinir ağı modeli vardır. Bu çalışmada kullanılan ileri beslemeli yapay sinir ağı modeli verilecektir.



Şekil 3 ileri beslemeli yapay sinir ağı modeli

Bu modelde tüm hücrelerin çıkışı bir sonraki katmandaki sinir hücrelerine giriş olmaktadır. Temelde 3 katman vardır giriş katmanı ara katman ki ara katman bazen birden fazla katman olabilir son katman ise çıkış katmanıdır. Katmanlar arası ileri doğru bir veri iletimi ile çıkış elde edilmektedir.

3. PROTEİNLER

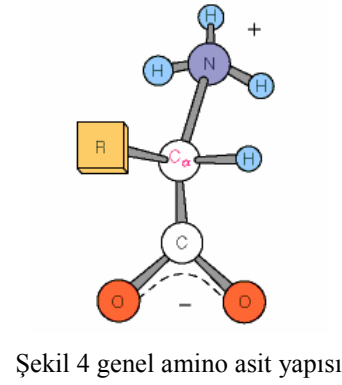
Vücut organlarının temel yapısını proteinler oluşturur. Zigottan itibaren bir canlının oluşması, organların, sistemlerin yapılması ve çalışması proteinler sayesinde olmaktadır. Proteinler de kendilerinden çok daha küçük parçalardan oluşur. Bu parçalar, "amino asit" adı verilen ve karbon, azot, hidrojen gibi atomların farklı şekillerde birleşmesiyle oluşan moleküllerdir.

İşin en önemli yanı ise, amino asitlerin bir proteini oluşturmak için mutlaka belirli bir sıra içinde dizilmeleri zorunluluğudur. Canlı bedenlerinde kullanılan 20 farklı türde amino asit vardır. Protein yapısındaki tek bir amino asidin bile eksilmesi veya yerinin değişmesi, o proteini işe yaramaz bir molekül yığını haline getirir. Bu nedenle her amino asit, tam gereken yerde, tam gereken sırada yer almalıdır [15]. Şekil 4' te bir amino asidin genel yapısı verilmiştir.

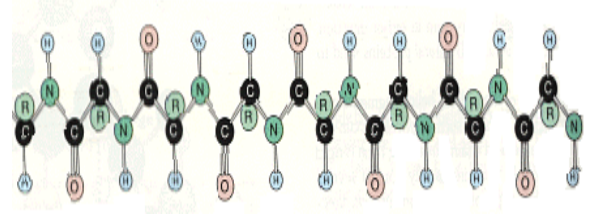
Proteinin yapısına göre amino asitler farklı yapılarda ve farklı sıralarda bir araya gelerek polipeptitleri ve proteinleri meydana getirirler. Proteinler 3 boyutlu yapıları itibarıyla dört farklı konfigürasyonda bulunurlar.

1-Primer yapı 2- Sekonder yapı 3- Tersiyer yapı 4- Kuaterner yapı

Primer yapı : Bir proteinin primer yapısı yalnızca amino asit moleküllerinin yan yana gelip zincir oluşturmalarından ibarettir.



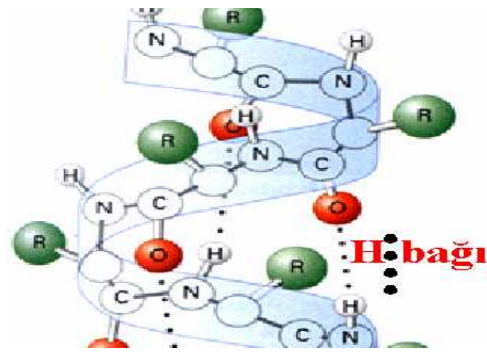
Şekil 4 genel amino asit yapısı



Şekil 5 primer yapı

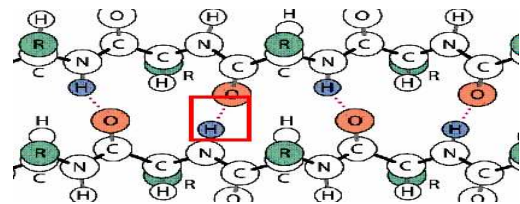
Sekonder Yapı : Bu yapıda da 4 farklı yapı vardır :alfa helezon , beta kırılmalı, beta bender ve tesadüfi kıvrılmalar

Alfa helezon : Bu yapıda zincir sarmal bir yapıda dönerek devam eder [15].



Şekil 6 sekonder alfa helezon yapısı

Beta kırılmalı : İki farklı zincir karşılıklı gelip birbirine bağlanırlar. karşılıklı gelen zincirlerdeki H ve O arasında bir bağ oluşur [15].



Şekil 7 sekonder beta kırılmalı yapısı

