

**YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**YAPAY SİNİR AĞLARI İLE PROTEİNLERİN İKİNCİL
YAPILARININ KESTİRİMİ**

Elektrik ve Elektronik Müh. Ertan ATAR

F.B.E. Elektronik ve Haberleşme Mühendisliği Anabilim Dalı Elektronik Programında Hazırlanan

YÜKSEK LİSANS TEZİ

**Tez Danışmanı : Y.Doç.Dr. Lale ÖZYILMAZ
İkinci Tez Danışmanı : Prof.Dr. Okan ERSOY**

İSTANBUL, 2005

İÇİNDEKİLER

	Sayfa
SİMGE LİSTESİ	iv
KISALTMA LİSTESİ	v
ŞEKİL LİSTESİ	vi
ÇİZELGE LİSTESİ	vii
ÖNSÖZ.....	ix
ÖZET	x
ABSTRACT	xi
1 GİRİŞ.....	1
2 ÖNCEKİ ÇALIŞMALAR	4
3 PROTEİNLER.....	8
3.1 Amino Asitler	8
3.2 Yaşam İçin Özel Hızlandırıcılar: Enzimler	12
3.3 Proteinlerin Yapısı	12
3.3.1 Proteinlerin Primer (Birincil) Yapısı: Amino Asit Dizilimi.....	13
3.3.2 Proteinlerin Sekonder (İkincil) Yapısı: Sarmal, Tabaka ve Sarım	13
3.3.3 Proteinlerin Tersiyer (Üçüncül) Yapısı	17
3.4 DNA.....	18
3.4.1 Genetik Bilginin Değerlendirilmesi.....	19
3.5 Protein Sentezi	20
3.6 Gen.....	22
3.6.1 Genetik Kodun Özellikleri.....	23
4 YAPAY SİNİR AĞLARI	25
4.1 Yapay Sinir Ağlarının Özellikleri.....	25
4.1.1 Doğrusal Olmama	25
4.1.2 Öğrenme	26
4.1.3 Genelleme	26
4.1.4 Uyarlanabilirlik.....	26
4.1.5 Hata Toleransı.....	26
4.2 Yapay Sinir Ağları ve Biyolojik Beyinler	27
4.3 Yapay Sinir Ağı Modeli.....	30
4.4 Yapay Sinir Ağlarında Kullanılan Öğrenme Algoritmaları.....	32
4.4.1 Öğreticili Öğrenme-Eğitici-Denetimli (Supervised Learning).....	32
4.4.2 Öğreticisiz Öğrenme-Eğiticisiz-Denetimsiz (Unsupervised Learning).....	32
4.4.3 Takviyeli Öğrenme (Reinforcement Training)	33
4.5 YSA Ağ Yapıları	33

4.5.1	İleri Beslemeli Ağlar.....	34
4.5.2	Geri Beslemeli Ağlar	34
4.6	Aktivasyon Fonksiyonları.....	35
4.6.1	Eşik (Threshold)	36
4.6.2	Parçalı Lineer (Purelin).....	36
4.6.3	Signum (Sign).....	37
4.6.4	Sigmoid (Logsig)	37
4.6.5	Yarışmalı (Competitive)	38
4.7	Ağ Yapıları	38
4.7.1	Statik Ağlar.....	38
4.7.2	Dinamik Ağlar	38
4.8	Öğrenme Kuralları	38
4.8.1	Çok Katmanlı Algılayıcı (MLP: Multilayer Perceptron).....	38
4.8.1.1	Geri Yayılım Algoritması	39
4.8.2	Radyal Temelli Fonksiyon.....	42
4.8.2.1	Olasılıklı Yapay Sinir Ağları (PNN: Probabilistic Neural Network)	43
4.8.2.2	Genelleştirilmiş Regresyon Yapay Sinir Ağları (GRNN: Generalized Regression Neural Network)	45
4.8.3	Kendi Kendine Organize Olan Haritalar (SOMs: Self Organizing Maps)	46
4.8.4	Kendi Kendine Organize Olan Global Sıralamalı Sınıflama Algoritması (SOGR: Self Organizing Global Ranking Algorithm)	50
5	YSA ile PROTEİNLERİN İKİNCİL YAPILARININ KESTİRİMİ.....	52
5.1	Özellik Sunma	52
5.2	Proteinlerin İkincil Yapıları Kestiriminin Önemi	54
5.3	Proteinlerin İkincil Yapıları Kestirimine Genel Bir Bakış	55
5.4	Yapılan Çalışmanın Ayrıntıları.....	55
5.4.1	Data Oluşturma	55
5.4.2	Dataların Özellikleri	55
5.4.3	Giriş ve Çıkış Kodlaması.....	59
5.4.4	YSA ile Proteinlerin İkincil Yapılarının Kestirimi.....	60
6	SONUÇLAR, TARTIŞMA ve ÖNERİLER.....	63
	KAYNAKLAR	85
	ÖZGEÇMİŞ.....	88

SİMGE LİSTESİ

a	Fonksiyon çıkışı
a_i^p	P paterni için aktivasyon
b	Bias
C	Proteinin sarım ikincil yapısı
c	Yarışmalı fonksiyon
D_j	Uzaklık
E	Toplam karesel hata
E	Proteinin tabaka ikincil yapısı
E_e	Birinci katmandaki düğümlerin çıkışı
$F'(a_i^p)$	P paterni için aktivasyonun türevi
H	Proteinin sarmal ikincil yapısı
IW	Giriş ağırlıkları
k	İterasyon
K	Giriş datalarının sınıf sayısı ve 2. katmandaki nöron sayısı
LW	Katman ağırlıkları
M	Çıkış katmanı nöron sayısı
m_j	Ağırlık
n	Verilen örnek boyutu
n	Net girişler
P	Giriş
P	Pencere genişliği
Q	Katmanlardaki nöron sayısı ve Giriş-hedef çiftlerinin sayısı
R	Giriş vektörlerinin eleman sayısı
R	Komşuluk boyutu
R	Amino asitte yan zincir grubu
R_i	Penceredeki i.amino asidin ikincil yapısı
S	Katmanlardaki nöron sayısı
T_{ec}	Sadece 0 veya 1 değerini alan OR'lama işlemini gerçekleyen ikinci katmana ait bağlantı katsayıları
t_i^p	P paterni için hedef değer
W_{ij}	Saklı katman ve çıkış katman arasında ağırlıklı bağlantı
W_{ij}	Ağırlık
W_{hi}	Giriş katman-saklı katmanı arasında ağırlıklı bağlantı
x_i	Giriş vektörü
y_i^p	P paterni için gerçek çıkış
ΔW_{ij}	Saklı katman ve çıkış katman ağırlıklı bağlantılar için ağırlık yenileme işareti
ΔW_{hi}	Giriş katman-saklı katmanı arasında ağırlıklı bağlantılar için ağırlık yenileme işareti
β_i	Öğrenme oranı
λ	Kazanç terimi
α	Öğrenme oranı
δ	Delta hata ifadesi
δ_i^p	P paterni için hata işaret terimi
θ	Eşik değeri

KISALTMA LİSTESİ

BIN	Binary – Çift - Bit
DNA	Deoksiribo Nükleik Asit
GRNN	Generalized Regression Neural Network - Genelleştirilmiş Regresyon Yapay Sinir Ağları
MLP	Multilayer Perceptron - Çok Katmanlı Algılayıcı
NN	Neural Network – Yapay Sinir Ağı
PDB	Protein Data Bank – Protein Data Bankası
PNN	Probabilistic Neural Network - Olasılıklı Yapay Sinir Ağları
RNA	Ribo Nükleik Asit
SOGR	Self Organizing Global Ranking Algorithm - Kendi Kendine Organize Olan Global Sıralamalı Sınıflama Algoritması
SOM	Self Organizing Maps - Kendi Kendine Organize Olan Haritalar
YSA	Yapay Sinir Ağları

ŞEKİL LİSTESİ

Şekil 2.1 Qian ve Sejnowski YSA yapısı	5
Şekil 2.2 Holley ve Karplus YSA yapısı	6
Şekil 3.1 Amino asitlerin genel formu	8
Şekil 3.2 Peptid bağının oluşumu	9
Şekil 3.3 Amino asitlerin moleküler yapısı	11
Şekil 3.4 Rezidü oluşumu	11
Şekil 3.5 Proteinin primer (birincil) yapısı	13
Şekil 3.6 Proteinin sarmal (helix) ikincil yapısı	14
Şekil 3.7 Proteinin tabaka (sheet) ikincil yapısı	15
Şekil 3.8 Proteinin tabaka (sheet) ikincil yapısının ayrıntılı görünüşü	15
Şekil 3.9 Proteinin sarmal (helix), tabaka (sheet) ve sarım (coil) ikincil yapılarının ayrıntılı görünüşleri	16
Şekil 3.10 Proteinin tersiyer yapısı	18
Şekil 3.11 DNA'nın yapısı	19
Şekil 3.12 Protein sentezi	21
Şekil 3.13 Genin yapısı	23
Şekil 4.1 Biyolojik nöron	27
Şekil 4.2 Miyelinli bir nöron yapısı	28
Şekil 4.3 Sinaps	29
Şekil 4.4 İşlem elemanın yapısı	30
Şekil 4.5 Denetimli öğrenme	32
Şekil 4.6 Denetimsiz öğrenme	33
Şekil 4.7 Takviyeli öğrenme	33
Şekil 4.8 İleri beslemeli ağ	34
Şekil 4.9 Geri beslemeli ağ	35
Şekil 4.10 Eşik fonksiyonu	36
Şekil 4.11 Parçalı lineer fonksiyonu	36
Şekil 4.12 Signum fonksiyonu	37
Şekil 4.13 Sigmoid fonksiyonu	37
Şekil 4.14 Geri yayılım MLP yapısı	39
Şekil 4.15 Radyal temelli nöron modeli	42
Şekil 4.16 Radyal temelli transfer fonksiyonu	42
Şekil 4.17a PNN matlab yapısı	43
Şekil 4.17b PNN yapısı	44
Şekil 4.18 GRNN yapısı	45
Şekil 4.19 SOM yapısı	46
Şekil 4.20 Yarışmalı katman	47
Şekil 4.21 SOM mimarisi	48
Şekil 4.22 SOM sınıflaması	49
Şekil 5.1a Proteinlerin ikincil yapılarının YSA ile kestirimi	52
Şekil 5.1b Proteinlerin ikincil yapılarının YSA ile kestirimi	53
Şekil 5.1c Proteinlerin ikincil yapılarının YSA ile kestirimi	53
Şekil 5.2 Dokuz kayan pencereyi iki çıkışlı örnek bir YSA yapısı	54
Şekil 5.3 On üç kayan pencereyi üç çıkışlı örnek bir YSA yapısı	54
Şekil 5.4 Proteinlerin yapılarının elde edilmesi	58

ÇİZELGE LİSTESİ

Çizelge 2.1 Yirmi tane amino asit kodlaması.....	7
Çizelge 2.2 Üç sınıf ikincil yapı kodlaması.....	7
Çizelge 3.1 Amino asitlerin açılımı.....	9
Çizelge 4.1 Bilgisayar ile beynin karşılaştırılması	27
Çizelge 4.2 Sinir sistemi ile YSA'nın benzerlikleri	29
Çizelge 5.1 Tezde kullanılan proteinlerin özellikleri	56
Çizelge 5.2 Proteinlerin ikincil yapılarının sınıflaması	58
Çizelge 5.3 Giriş katmanı için 20 amino asitin BIN-20 yöntemi ile sayısallaştırılması	59
Çizelge 5.4 Üç durumlu çıkış katmanının sayısallaştırılması	60
Çizelge 5.5 İki durumlu çıkış katmanının sayısallaştırılması.....	60
Çizelge 6.1 Hemoglobin-141 proteini ve 11- pencere boyutu için eğitim sonuçları.....	63
Çizelge 6.2 Hemoglobin-141 proteini ve 11- pencere boyutu için test sonuçları	63
Çizelge 6.3 Hemoglobin-141 proteini ve 13- pencere boyutu için eğitim sonuçları.....	64
Çizelge 6.4 Hemoglobin-141 proteini ve 13- pencere boyutu için test sonuçları	64
Çizelge 6.5 Hemoglobin-141 proteini ve 15- pencere boyutu için eğitim sonuçları.....	65
Çizelge 6.6 Hemoglobin-141 proteini ve 15- pencere boyutu için test sonuçları	65
Çizelge 6.7 Hemoglobin-141 proteini ve 17- pencere boyutu için eğitim sonuçları.....	66
Çizelge 6.8 Hemoglobin-141 proteini ve 17- pencere boyutu için test sonuçları	66
Çizelge 6.9 Hemoglobin-141 proteini ve 19- pencere boyutu için eğitim sonuçları.....	67
Çizelge 6.10 Hemoglobin-141 proteini ve 19- pencere boyutu için test sonuçları	67
Çizelge 6.11 Hemoglobin-141 proteini ve 21- pencere boyutu için eğitim sonuçları.....	68
Çizelge 6.12 Hemoglobin-141 proteini ve 21- pencere boyutu için test sonuçları	68
Çizelge 6.13 Hemoglobin 141 proteini ve 23- pencere boyutu için eğitim sonuçları	69
Çizelge 6.14 Hemoglobin-141 proteini ve 23- pencere boyutu için test sonuçları	69
Çizelge 6.15 Hemoglobin-146 proteini ve 11- pencere boyutu için eğitim sonuçları.....	70
Çizelge 6.16 Hemoglobin-146 proteini ve 11- pencere boyutu için test sonuçları	70
Çizelge 6.17 Hemoglobin-146 proteini ve 13- pencere boyutu için eğitim sonuçları.....	71
Çizelge 6.18 Hemoglobin-146 proteini ve 13- pencere boyutu için test sonuçları	71
Çizelge 6.19 Hemoglobin-146 proteini ve 15- pencere boyutu için eğitim sonuçları.....	72
Çizelge 6.20 Hemoglobin-146 proteini ve 15- pencere boyutu için test sonuçları	72
Çizelge 6.21 Hemoglobin-146 proteini ve 17- pencere boyutu için eğitim sonuçları.....	73
Çizelge 6.22 Hemoglobin-146 proteini ve 17- pencere boyutu için test sonuçları	73
Çizelge 6.23 Hemoglobin-146 proteini ve 19- pencere boyutu için eğitim sonuçları.....	74
Çizelge 6.24 Hemoglobin-146 proteini ve 19- pencere boyutu için test sonuçları	74
Çizelge 6.25 Hemoglobin-146 proteini ve 21- pencere boyutu için eğitim sonuçları.....	75
Çizelge 6.26 Hemoglobin-146 proteini ve 21- pencere boyutu için test sonuçları	75
Çizelge 6.27 Hemoglobin-146 proteini ve 23- pencere boyutu için eğitim sonuçları.....	76
Çizelge 6.28 Hemoglobin-146 proteini ve 23- pencere boyutu için test sonuçları	76
Çizelge 6.29 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 11- pencere boyutu için eğitim sonuçları (tabaka sınıfı çıkarılmış)	77
Çizelge 6.30 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 11- pencere boyutu için test sonuçları (tabaka sınıfı çıkarılmış).....	77
Çizelge 6.31 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 15- pencere boyutu için eğitim sonuçları (tabaka sınıfı çıkarılmış)	78
Çizelge 6.32 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 19- pencere boyutu için eğitim sonuçları (tabaka sınıfı çıkarılmış)	78

Çizelge 6.33	2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 19- pencere boyutu için eğitim sonuçları (tabaka sınıfı çıkarılmış)	79
Çizelge 6.34	2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 19- pencere boyutu için test sonuçları (tabaka sınıfı çıkarılmış).....	79
Çizelge 6.35	2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 11- pencere boyutu için eğitim sonuçları	80
Çizelge 6.36	2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 11- pencere boyutu için test sonuçları	80
Çizelge 6.37	2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 15- pencere boyutu için eğitim sonuçları	81
Çizelge 6.38	2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 15- pencere boyutu için test sonuçları	81
Çizelge 6.39	2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 19- pencere boyutu için eğitim sonuçları	82
Çizelge 6.40	2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 19- pencere boyutu için test sonuçları	82

ÖNSÖZ

Canlılar için çok önemli olan proteinin fonksiyonelliğinin bilinmesi gereklidir. Bu da ancak proteinlerin ikincil yapılarının bilinmesi ile mümkündür. Deneysel olarak bu olay çok zor, çok pahalı ve çok zaman almaktadır. Protein ikincil yapılarının Yapay Sinir Ağları (YSA) ile kestirimi konusunda bilim adına çok faydası olacağına inandığımız bu çalışmayı hazırladık. Bu çalışmamız sonucunda diğer yapılan çalışmalara göre çok daha yüksek doğruluk oranları elde edilmiştir ve çok daha farklı YSA metodları denenmiştir. Ayrıca yepyeni bir algoritma olan Kendi Kendine Organize Olan Global Sıralamalı Sınıflama Algoritması'nın (SOGR) ne kadar iyi sonuç verdiği görülmüştür.

Bu tez çalışması sırasında beni yönlendiren, yakın ilgi ve desteklerini esirgemeyen tez danışmanlarım değerli hocalarım Sayın Y.Doç.Dr. Lale Özyılmaz'a ve Prof.Dr. Okan Ersoy'a, ayrıca diğer hocalarıma, günlere gelmemde büyük emeği ve katkısı bulunan aileme ve bana her konuda büyük destek veren nişanlıma teşekkürlerimi sunarım.

ÖZET

Proteinler, önemli görevlerinden dolayı organizmaların en önemli bölümlerindendir. Proteinlerin biyolojik fonksiyonu ile proteinlerin birincil ve özellikle de ikincil yapısı çok yakın ilişkilidir ve bu nedenle proteinlerin ikincil yapılarının bilinmesi gereklidir. Proteinlerin ikincil yapıları birincil yapılarına göre çok daha karmaşık ve çok daha fonksiyoneldir. Protein ikincil yapılarının deneysel olarak bulunması hem çok zor olmakta, hem çok zaman almakta hem de çok pahalı olmaktadır. Bunun tersi yapay sinir ağları (YSA), proteinlerin ikincil yapılarının kestirimi için çok kullanışlı bir yöntemdir.

Bu çalışmada kullanılan harfsel bütün datalar PDB'den (Protein Data Bankası) (<http://www.rcsb.org/pdb/>) elde edilmiş daha sonra harfsel datalar Matlab 7.0'da yazılan bir programla yapay sinir ağlarına protein ikincil yapı kestirimi yapılabilmesi için uygun sayısal formatlı datalara dönüştürülmüştür. Yirmi farklı tip data ve beş ayrı tip YSA algoritması oluşturulmuştur. Bu yirmi farklı data Hemoglobin-141 proteini yedi (11-13-15-17-19-21-23) farklı pencere boyutuyla, Hemoglobin-146 proteini yedi (11-13-15-17-19-21-23) farklı pencere boyutuyla, ayrıca 2003 yılında yapılan Protein Yapı Konferansı'nda sunulan on beş yeni proteinin üç (11-15-19) farklı pencere boyutuyla ve son olarak bu on beş proteinin tabaka yapısı çıkarılmış olarak üç (11-15-19) farklı pencere boyutuyla oluşturulmuş datalardır. Bu beş farklı YSA algoritması ise Kendi Kendine Organize Olan Global Sıralamalı Sınıflama Algoritması(SOGR), Çok Katmanlı Algılayıcı(MLP), Olasılıklı YSA(PNN), Genelleştirilmiş Regresyon YSA(GRNN), Kendi Kendine Organize Olan Haritalar(SOMs) şeklindeki algoritmalar. SOGR algoritması ise SOM'a karşı 2004 yılının sonunda tamamlanmış yepyeni bir algoritmadır. Genel olarak sonuçlar çok iyi çıkmıştır ve test başarımları en yüksek SOGR algoritmasında görülmüştür.

Anahtar kelimeler: Protein ikincil yapı kestirimi, Protein Data Bankası, Kendi Kendine Organize Olan Global Sıralamalı Sınıflama Algoritması, Çok Katmanlı Algılayıcı, Olasılıklı YSA, Genelleştirilmiş Regresyon YSA, Kendi Kendine Organize Olan Haritalar.

ABSTRACT

Proteins are one of the most important parts of an organism because of their vital tasks. Consequently, it is necessary to know both the primary and secondary structures of proteins as closely related to its biological function. Secondary structures of proteins are much more complicated and much more functional than primary structures of proteins. Discovery of proteins secondary structures is both very difficult and very expensive. Besides it requires a lot of time. On the contrary, Artificial Neural Networks (ANNs) are a useful methodology for secondary structure prediction of proteins as a more automated, cheaper, and less labor intensive approach.

In this study, all of the data were obtained from PDB (Protein Data Bank) (<http://www.rcsb.org/pdb/>). Then, the letter data were converted to numerical data and processed with ANNs by a Matlab program. Twenty different types of data and five different types of ANN algorithms were used. The twenty different type data were Hemoglobin-141 protein seven different kinds of sliding windows (11-13-15-17-19-21-23), Hemoglobin-146 protein seven different kinds of sliding windows (11-13-15-17-19-21-23), fifteen new protein presented at the Fifth Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction Asilomar Conference Center in 2003 three with kinds of sliding windows (11-15-19) and the same proteins three different kinds of sliding windows (11-15-19) without the sheet structures. The five different types ANNs were Self Organizing Global Ranking Algorithm (SOGR), Multilayer Perceptron (MLP), Probabilistic Neural Network (PNN), Generalized Regression Neural Network (GRNN) and Self Organizing Maps (SOMs). In general, results were very satisfactory, and the SOGR had the highest testing accuracies.

Keywords: Protein secondary structure prediction, Protein Data Bank, Self Organizing Global Ranking Algorithm, Multilayer Perceptron, Probabilistic Neural Network, Generalized Regression Neural Network, Self Organizing Maps.

1. GİRİŞ

Tüm yaşayan hücrelerin esas unsuru proteinlerdir. Proteinler, amino asitlere parçalanırlar ve organizmada da amino asitlerden sentezlenirler [1]. Proteinler, kimyasal reaksiyonları katalize eden (enzim), gen ekspresyonunu* düzenleyen (düzenleyici protein), hücre, doku ve virüslerin birçok yapısal özelliklerini katalize kontrol eden moleküllerdir. Protein, birbirine bağlı olan birçok amino asitten oluşmuştur (Jones ve Karp, 1990). Örneğin saç, tırnak ve tüylerde bulunan sert yapıyı oluşturan keratin isimli madde bir proteindir. Bazı proteinler, kasları kemiğe bağlayan tendonlarda bulunan dayanıklı naylon benzeri bir maddeyi oluştururlar. Ayrıca hücreye gelen mesajları getirenler de, mesajları alan ve değerlendirenler de proteinlerdir. Hücrenin içine giriş çıkışları kontrol eden kapılar ve pompa sistemleri de proteinlerdir. Kimyasal reaksiyonları hızlandıranlar yine proteinlerdir. ‘ Hemoglobin ’ adındaki protein kandaki oksijeni dokulara taşır. Transferin isimli protein ise kanda bulunan demiri taşır. İmmunoglobülinler bakteri ve virüslere karşı vücudu savunan proteinlerdir. Fibrinojen ve trombin ise kanın pıhtılaşmasını sağlar. İnsülin, vücuttaki şeker metabolizmasını düzenleyen bir protein çeşididir. Derinin pürüzsüz elastikiyetini ve kemiklerin dayanıklılığını sağlayan ise kolajen isimli bir başka proteindir. Atardamarları çevreleyen kauçuk benzeri elastik maddeyi oluşturan da yine başka bir proteindir. Retinaya ışık çarptığında görme etkisini başlatan ise rodopsin isimli proteindir. Bu arada başka proteinler de gözün lensini oluşturan saydam maddeyi yaparlar. Hücrelerin içine moleküllerin giriş çıkışında yine özel taşıyıcı proteinler görev yapar. Tüm canlılığın bilgisini taşıyan DNA molekülü proteinler olmadan kopyalanamaz ve bilgi üretemez, hücre bölünmesini sağlayamaz (Demirsoy, 1995).

Proteinler, normal olarak yirmi (20) çeşit amino asitten oluşurlar. Doğada 200'ün üzerinde amino asit bulunmasına rağmen proteinler bu amino asitlerin sadece 20 tanesinden oluşur. Çünkü sadece bu amino asitlerin ikincil yapılarında sarmal, tabaka ve sarım yapıları vardır ki bu yapı sayesinde protein üç boyutlu şeklini alabilmektedir ve bunlar, bu proteinin işlevini görebilmesi için gerekli olan yapılardır. (Demirsoy, 1995).

* Kalıtsal materyalde, genetik bilgiyi taşıyan üniteler gen olarak bilinmektedir. Hücrelerde bulunan genlerin taşıdıkları bilginin, canlının yaşantısında gerekli olan moleküllere dönüşmesi gen ekspresyonu olarak bilinmektedir. Genin ekspresyonu, DNA baz dizilişinden başlayıp RNA'ya kopyalanıp protein molekülü oluşturmak üzere amino asitleri bir araya getirme olayıdır (Jones ve Karp, 1990).

Proteinler, amino asit isimli moleküllerden oluşurlar. Bu amino asitler birbirlerine peptit bağlarıyla bağlanırlar. Amino asit zincirleri polipeptid olarak isimlendirilir [1]. Amino asitler proteinlere göre daha küçük moleküller olmalarına rağmen, oldukça kompleks bir yapıları vardır. Amino asitleri oluşturan atomlar üç ayrı grup halinde bulunurlar; amino grubu, karboksil grubu ve yan zincir grubu (ya da radikal grup). Bütün amino asitlerde amino ve karboksil grupları aynıdır. Bir amino asiti diğerlerinden farklı kılan tek özellik, moleküle bir ucundan bağlanan yan zincir grubudur. Bu yan zincir gruplarının her amino asitte farklı olması sayesinde her amino asit birbirinden çok farklı özelliklere sahip olur. En basit R grupları glycine (-H) ve alanin(-CH_3) dir (Jones ve Karp, 1990).

Polipeptid zincirleri bir amino asitin karboksil grubu diğer bir amino asitin amino grubuna bağlandığı zaman peptid bağ olarak isimlendirilen düzenli bağ oluşturmaktadır. Bu şekilde proteinin temel ünitesi polipeptid olup karbon atomları iskeleti oluşturmak üzere peptid gruplarında birbirlerine alternatifli olarak bağlanırlar (Jones ve Karp, 1990).

Protein “ilk element” (Yunancada proteios) anlamına gelir, zira proteinler tüm bitki ve hayvan hücrelerinin temel parçasıdır. Bir insan 100.000 değişik proteinden oluşur. Proteinler lineer (doğrusal) “zincirlerdir”. Moleküler “bağları” amino asit moleküllerinden oluşmaktadır. Yirmi amino asitin her biri kendine has bir şekle sahiptir. Bir zincir içinde sıralandıklarında ortaya çıkan proteinler, üç boyutlu “tel heykellere” dönüşürler. Proteinin heykel kalıbı amino asit bağlantılarının sıralamasına bağlıdır. Protein zincirinin boyunca bulunan elektromanyetik yükün dengelenmesi heykelin nihai şeklini kontrol etmekte hizmet verir. Bir protein heykelinin benzersiz, kendine has şekline onun formasyonu diye adlandırılır. Bir anahtar ve kilit gibi protein heykelleri de çevrelerindeki moleküllerin (başka proteinler içeren) şekillerini tamamlar.

Proteinler kimyasal olarak diğer moleküllerle eşleştiklerinde proteindeki elektromanyetik yüklerin dağılımı da değişir. Elektromanyetik yükteki değişim proteinin şeklinin değişmesine neden olur. Bu nedenle kimyasalların eşleşmesi sırasında bir protein bir formasyondan bir diğer formasyona şekil değiştirir. Protein şekil değiştirirken aynı zamanda da bir hareket üretir. Bir proteinin hareketi iş yaratır. Belirli bir fonksiyonu yerine getirmek için, bir arada çalışarak interaktif guruplar oluşturan proteinlere patikalar adı verilir. Belirli protein patikalarının aktiviteleri ise canlı organizmalar tarafından gerçekleştirilen sindirim, soluma, üreme gibi adlar altında toplayabileceğimiz tüm fizyolojik fonksiyonları temin eder (Lipton, 2001).

Her polipeptid molekülünün iki sonu farklı olup biri serbest NH_2 grubu şeklinde amino terimi olarak ikinci tarafı ise COOH şeklinde karboksil terimi olarak isimlendirilir. Polipeptidler, zincirin karboksil tarafına bağlanarak sentez edilirler. Polipeptid zincirinin amino asitleri amino terimi tarafından başlayarak sayılır (Jones ve Karp, 1990).

Çoğu polipeptid zincirleri yüksek oranda katlanırlar. Yan zincirler arasındaki interaksiyonlar vasıtasıyla katlanma şekli amino asitlerin dizilişi ile ilgilidir. Bu şekilde her bir polipeptid zinciri sentezine bağlı olarak üç boyutlu yapıya katlanır. Bazı durumlarda protein katlanması hücre içerisinde diğer proteinlerle interaksiyon halinde oluşur. İşlem esnasında su molekülleri serbest bırakılır, bu arada peptid bağı oluşur. Proteinlerin doğru olarak görevlerini yapmaları ancak doğru dizildikleri zaman gerçekleşir. Amino asitlerin zincirden kaybı veya yanlış düzeni enzim aktivitesinde kayıp veya azalmaya neden olmaktadır (Jones ve Karp, 1990).

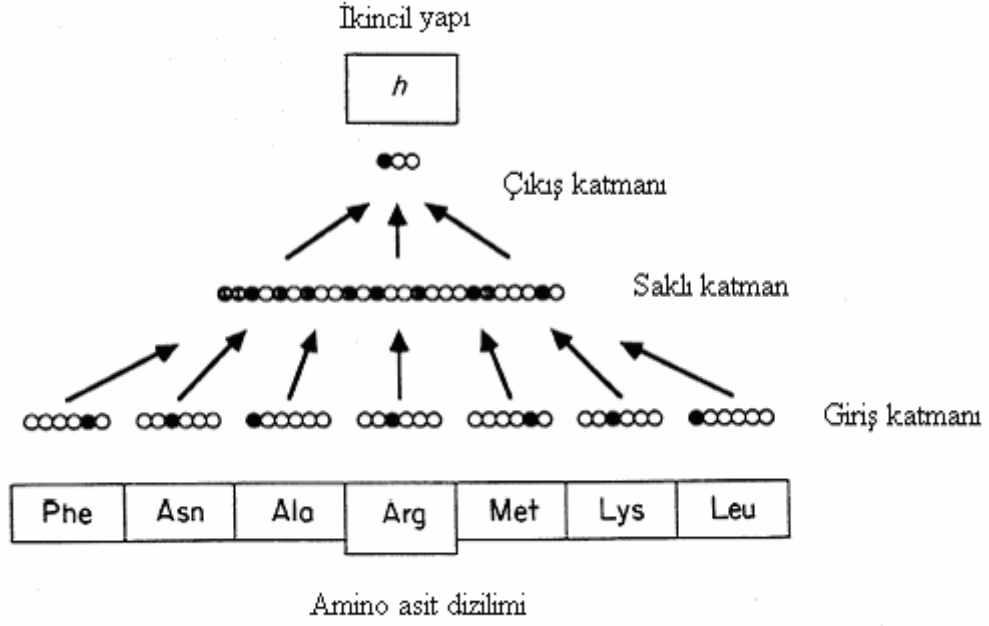
Bu çalışmada deneysel olarak protein ikincil yapısının bulunması çok zor, çok zaman alması ve çok pahalı olması sebebiyle YSA ile protein ikincil yapısının kestirimi yapılmıştır. YSA olarak SOM, MLP, PNN, GRNN ve orijinallik açısından 2004 yılının sonunda geliştirilen yeni bir algoritma olarak SOGR kullanılmıştır ve data, Hemoglobin-141 proteini Hemoglobin-146 proteini ile 2003 yılındaki Protein İkincil Yapı Konferansı'nda sunulan 15 yeni protein ve tabaka sınıfı çıkarılmış bu 15 yeni proteinden oluşturulmuştur.

Bu tezin genel içeriğinde, ikinci bölümde, önceki yapılan çalışmalar hakkında bilgi verilmiştir. Üçüncü bölümde, proteinler hakkında bilgi verilmiştir. Dördüncü bölümde, yapay sinir ağları anlatılmıştır. Beşinci bölümde de bu tez için yapılan çalışmanın ayrıntıları anlatılmaktadır. Altıncı bölümde ise sonuçlar, tartışma ve öneriler yer almaktadır.

2. ÖNCEKİ ÇALIŞMALAR

(Rost ve Sander, 1993) ve (Rost, 1996); protein ikincil yapı tahmini için yaptıkları çalışmada girişte 13 boyutlu pencere varsa BIN20 kodlama için bir amino asit 20 binary ile temsil edildiği için toplam $13 \times 20 = 260$ birim gereklidir. Protein dizilimleri, sabit uzunluklu bölümlerde kayan pencerelerde işlenmiştir. Genellikle 11 -19 boyutlu pencere ile amino asit dizilimleri merkez rezidüye göre tahmin yapılır. Örneğin; üç durumlu tahmin için 3-kodlu ise 100-sarmal, 010-tabaka, 001-sarım, iki durumlu tahmin için ise 10-sarmal ,01-sarmal olmayan yapı kullanılmıştır. Perceptron ağırlık matrisleri yapı ile amino asit pozisyonu arasında ilişki vardır. Bu çalışmalarda giriş ve çıkış kodlama arasında korelasyon bilgisi kullanılmıştır. Doğruluk ortalama %65 civarındadır. Back-propagation yapay sinir ağlarının kaskat bağlanması durumunda; NN1(dizilim-dizilim) , NN2 (yapı-yapı); birinci seviye NN1 'de 13 amino asit penceresi ve tek merkez rezidü için eşleme bilgisi elde edilmiştir. Her rezidü de giriş birimlerin yirmi değeri kullanılmıştır. Giriş vektörü, dizilim profili için 20 birimli giriş vektörü içerir. İkinci seviye amino asit penceresi ve ikincil yapı arasındaki korelasyon içindir. Çıkış üç durumludur. Birinci yapının çıkışı ikincil yapının girişi olmuştur (Wu ve McLarty, 2000).

(Qian ve Sejnowski, 1988) protein ikincil yapı tahmininde için lineer olmayan bir YSA modeli geliştirmişlerdir. Datalar PDB'ten elde edilmiştir. Toplam 106 protein kullanılmıştır. Ağ modeli var olan protein yapısını, amino asitlerin yerel dizilimlerinde ikincil yapısını öğrenir. Ortalama başarı oranı % 33 - % 61 arasında olmuştur. Tahminin doğruluğu ilk yirmi beş rezidü için iyidir. Proteinin ikincil yapısının tahmin problemi benzerdir, giriş semboller analog harfler, yani amino asitler ve çıkış sembolleri de analog ikincil yapıdır. Metodun amacı, datalardaki bilinen ikincil yapı datalarını kullanarak, protein ikincil yapılarını tahmin etmektir. Bilinen yapılar amino asitler ve onların etkileşimini bilgisini içermektedir. Bilinen yapı dataları PDB'den elde edilmiştir. Yüz altı proteinin bir kısmı eğitim, bir kısmı da test için kullanılmıştır. YSA için eğitici ileri beslemeli back-propagation öğrenme algoritması kullanılmıştır. Gizli katmandaki nöron sayısı değişimleri ile doğruluk oranı değişmesi çok az olmuştur. En iyi sonuç 13-pencere boyutunda olmuştur. En kötü tahmin tabaka sınıfında olmuştur.

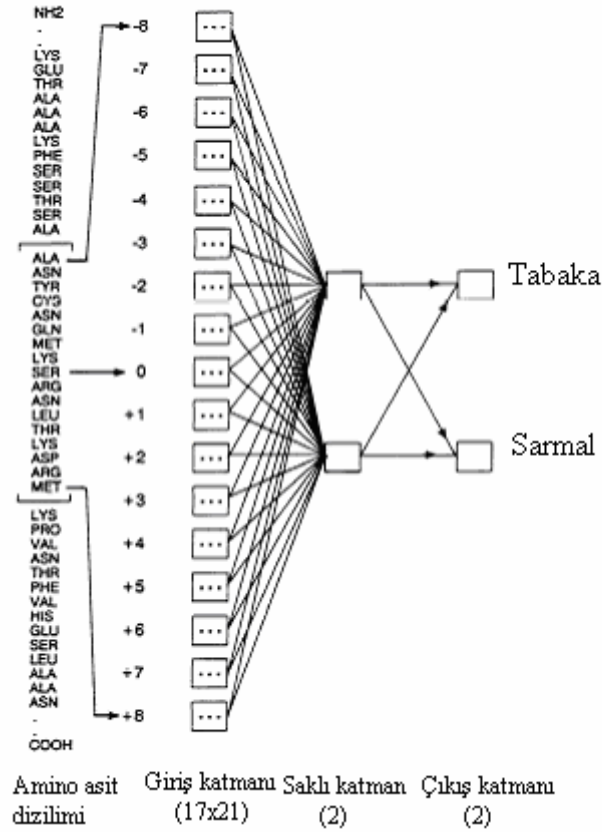


Şekil 2.1 Qian ve Sejnowski YSA yapısı (Qian ve Sejnowski, 1988)

Kayan pencere on üç boyutundadır. Merkez rezidüye göre tahmin yapılmıştır. Giriş katmanı on üç gruba ayrılmıştır. Her grup, yirmi birim içerir. Her birim bir amino asit sunar. Yirmi bitin içinde sadece bir tane bir, diğerleri sıfır olur. Buna lokal kodlama denir, çünkü her birim sadece birini kodlar. İlk önce ağırlıklar random olarak seçilmiştir (-0.3 - 0.3 arasında). Back-propagasyon öğrenme algoritmasına göre eğitmede ağırlıklar güncelleştirilir ve hata minimize edilmiştir. Ağın performansı birçok etkene bağlıdır; eğitim örneklerine, gizli birim sayısına, benzerliklere vb. Amino asitlerin dizilimine göre ikincil yapıları atanmıştır. Girişler yirmi birime sahip ve kırk saklı birim ve üç çıkış birimi YSA ile eğitilmiştir. Ağırlık değişimlerine de bakıldığında en etkili ağırlık merkez rezidü de olduğu görülmüştür (Qian ve Sejnowski, 1988).

(Holley ve Karplus, 1989) yaptıkları çalışmada; proteinin ikincil yapısını iki sınıfa ayırmışlardır (sarmal ve tabaka). Data seti olarak eğitim için 8315 rezidü, test için ise 2441 rezidü kullanılmıştır. Genel olarak doğruluk oranı % 60 civarında olmuştur. Datalar Brookhaven Protein Data Bank'tan (PDB) sağlanmıştır. Bir giriş katmanı, bir saklı katman (iki birim), bir çıkış katmanı(iki birim) kullanılmıştır. Giriş katmanında bir hareketli pencere kullanılmıştır. Bu pencere on yedi boyutunda ve merkeze göre her iki yönde sekizer olarak bölünmüş istatistiksel korelasyon tahmin noktası hesaplanmıştır. Binary kodlama , BIN21

kullanılmıştır. Tahmin merkez rezidüye göre yapılmıştır. İlk önce ağırlıklar rastgele seçilmiştir. İleri beslemeli Perceptron kullanılmıştır. 1000 adımda eğitim gerçekleştirilmiştir. Birimler ve pencere boyutu deneme-yanılma yoluyla bulunmuştur. Saklı katmandaki birim sayısı artarsa eğitim doğruluğu artmıştır, fakat test doğruluğu düşmüştür. Sarmal (Helix) dışındaki tahminlerin doğruluğu, sarmal (helix) tahminine göre daha düşüktür (Holley ve Karplus, 1989).



Şekil 2.2 Holley ve Karplus YSA yapısı (Holley ve Karplus, 1989)

(Tang, 1995) yaptığı çalışmada geriye yayılım algoritması kullanılmıştır. Datalar PDB'ten alınmıştır. Ağ yirmi birimli bir giriş katmanı, bir saklı katman ve bir çıkış katmanından oluşmuştur. Pencere on bir boyutludur. Merkez rezidünün sağında ve solunda beşer rezidü vardır.

Çizelge 2.1 Yirmi tane amino asit kodlaması (Tang, 1995)

Giriş	A	C	D	E	...	T	V	W	Y
Çıkış	0.00	0.05	0.10	0.15	...	0.85	0.90	0.95	1.00

Çizelge 2.2 Üç sınıf ikincil yapı kodlaması (Tang, 1995)

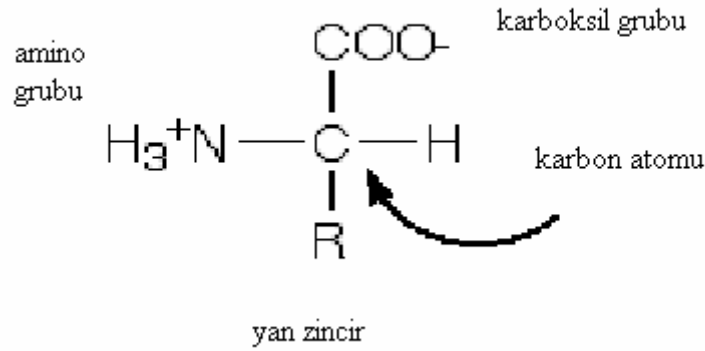
Giriş	Sarmal	Tabaka	Sarm
Çıkış	0.20	0.40	0.60 0.80

Ağ, 100 adımda eğitim işlemini tamamlamıştır. Ortalama doğruluk değeri % 45 civarında olmuştur (Tang, 1995).

3. PROTEİNLER

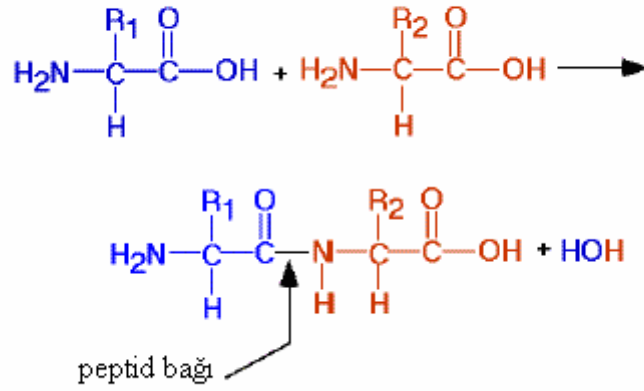
3.1 Amino Asitler

Proteinler, normal olarak yirmi (20) çeşit amino asitten oluşurlar. Amino asitlerden sadece 20 tanesinin kullanılmasının nedeni ise; yaşam için gerekli olan proteinler görevlerini yerine getirebilmek için belirli özelliklere sahip olmalıdırlar ve onlara bu özelliklerini sağlayan en önemli unsurlardan biri amino asitlerdir. Proteinlerin ikincil yapılarının sarmal, tabaka ve sarım özelliğine sahip olmalıdırlar. Ayrıca yan zincirler çok büyük olmamalıdır, yoksa onları proteinin içine paket ederek yerleştirmek imkansızlaşır (Demirsoy, 1995). Amino asitler iki karakteristik fonksiyonel grup içerirler: Amino (NH_3) grubu ve karboksil (COOH) grubu. Aşağıdaki şekilde amino asitlerin genel formülü verilmiştir. Amino asitlerde karbon atomuna bağlı bir amino, bir karboksil grubu ve bir H atomu vardır. Farklılığı sağlayan R(yan) grubu ise bir karbon iskeletidir [1]. Amino asitlerin yan zincir gruplarındaki atomların şekli, sayısı ve sıralamaları, elektrik yükleri, hidrojen bağı kapasitelerinin farklı farklı olması, amino asitlere çeşitlilik kazandırır ve bu çok çeşitlilik de yine çok çeşitli protein üretilir. Örneğin yan zincir gruplarının (+) veya (-) elektrik yükünün olması veya yüksüz olması amino asit molekülünün suda eriyip erimemesini sağlar (Demirsoy, 1995). Proteinlerin bütün fonksiyonel özellikleri amino asit diziliminden türemiştir (Karey ve Hanley, 1997).



Şekil 3.1 Amino asitlerin genel formu

peptid bağ formasyonu



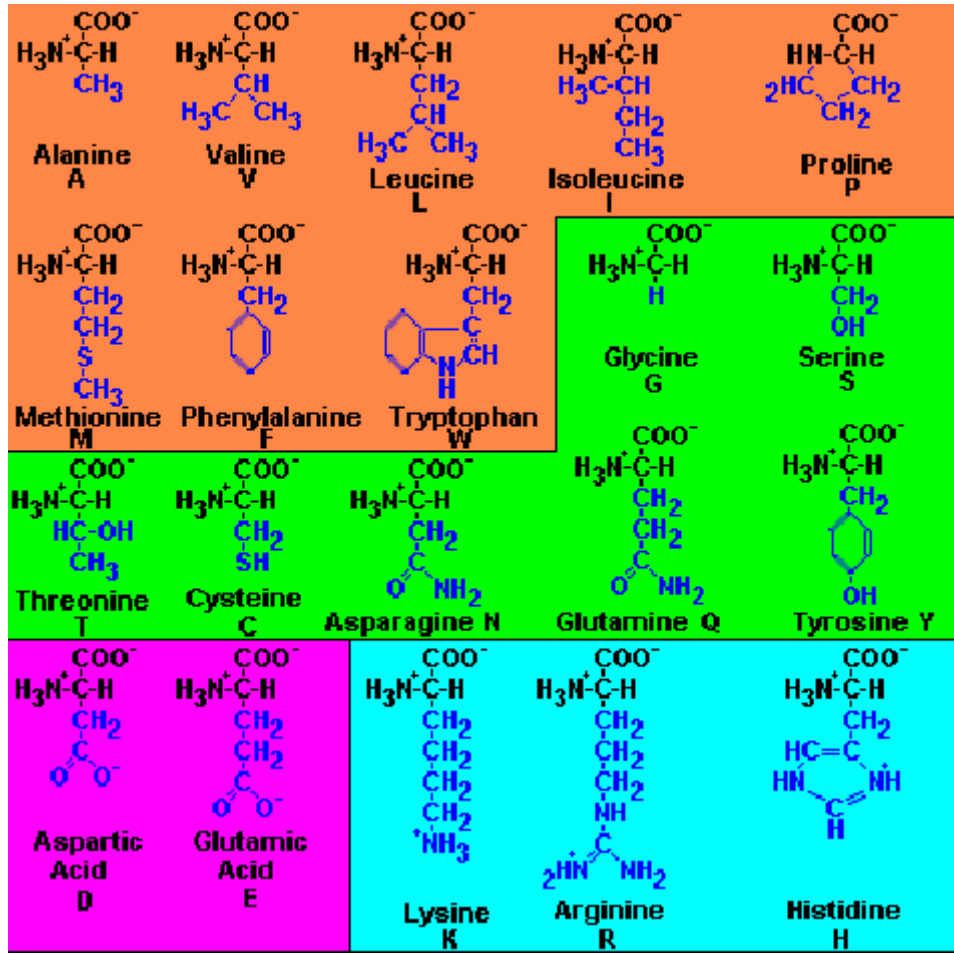
Şekil 3.2 Peptid bağının oluşumu

Çizelge 3.1 Amino asitlerin açılımı

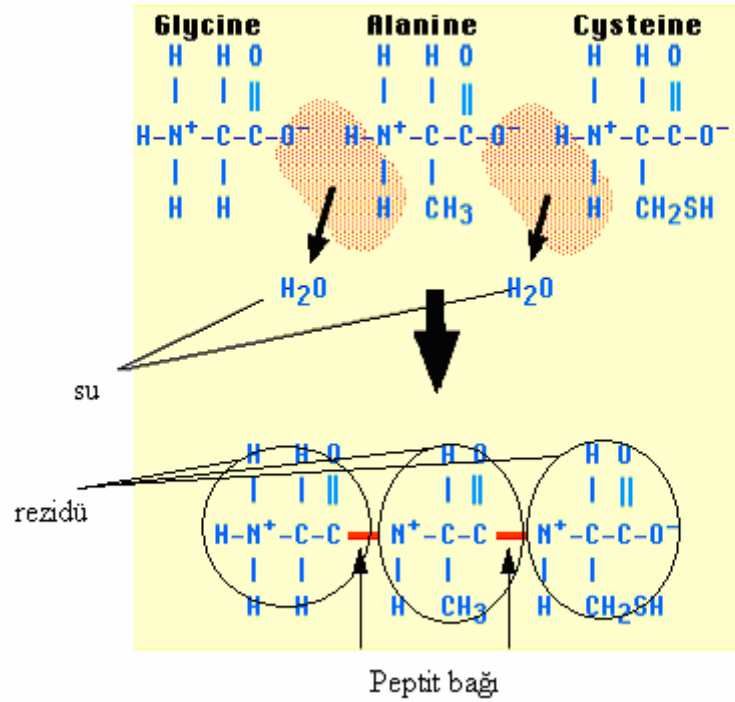
Amino asit tablosu		
Bir harfli sembol	Üç harfli sembol	İsim
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Proteinlerin oluşması için gereken koşullardan bir diğeri de, doğru amino asitlerin, doğru sıralamada bulunmalarının yanısıra, doğru bağ ile birbirlerine bağlanmalarıdır. Amino asitler arasındaki bu bağ adeta bir köprü gibidir. Bu köprüde, amino asitlerin birbirine bağlanma açıları, yönleri, içlerindeki atomların çeşidi ve sayısı her bir protein için özel olarak hesaplanmıştır. Örneğin bir protein zincirindeki iki amino asidin arasındaki birleşme açısının olması gerekenden farklı olması bu köprünün kurulmasını, dolayısıyla proteinin oluşumunu engelleyecektir. Sonuçta işe yaramayan farklı bir molekül elde edilecektir. İşte amino asitlerin birleşmelerindeki bu özel köprülere "peptid bağları" adı verilir. Canlıların kimyasını inceleyen bilim adamları canlıların yapısında bulunan moleküllerdeki atomların hemen hemen hepsinin kovalent bağ* denilen bir tür bağ ile birleştiklerini biliyorlardı. Fakat yapılan incelemelerde, proteinleri oluşturmak için biraraya gelen amino asitlerin aralarında, daha önce tanımlanmamış özel bir bağ kurdukları anlaşıldı. Bütün proteinler için bu değişmez bir kuraldı. Proteinlerin oluşmasında bu bağların önemi ilk kez 1902 yılında Hofmeister ve Fisher tarafından ortaya çıkarıldı. Bu iki araştırmacı bu özel bağın varlığını ortaya çıkarmak bir test uyguladılar. Bu testin sonucunda proteinlerde görev alan özel bir bağın(peptid bağı) varlığını tespit ettiler. Peptid bağını diğer bağlardan ayıran en önemli özellik, ısıtılarak veya buna benzer yollarla çok çabuk çözülmemesidir. Peptid bağ ancak yüksek ısıda, uzun süre kuvvetli asit ya da bazlara maruz kaldığında çözülebilir. Proteinlerin sağlam ve dayanıklı olmalarını da işte bu peptid bağ sağlar. Bu özel bağın kurulabilmesi için bir amino asitteki karboksil grubunun (yani içinde karbon, oksijen, hidrojen atomlarının bulunduğu özel molekül) diğer amino asitteki amino grubuyla (içinde nitrojen ve hidrojen atomları bulunan özel bir molekül) birleşmesi gereklidir. Bu şekilde protein zinciri boyunca bağlantı yerlerinde önemli bir denge kurulmuş olur. Protein moleküllerinin oluşturan bu bağın meydana gelmesi sırasında su açığa çıkar, bu geriye kalan dizilime 'rezidü' denir (Gözükara, 1997). Yani proteinin birincil yapısındaki her bir amino aside rezidü denir.

* Kovalent bağ : Aynı türden ya da farklı atomların bir veya iki elektronu paylaşmaları ile oluşan bağa kovalent bağ denmektedir. N₂, O₂, H₂O, SiC,.....gibi moleküllerde ya da kristallerde atomlar kovalent bağ ile bağlanırlar. İki atomlu bir molekül olan H₂ nin elektron spinleri birbirine zıt yönelmiştir. Hidrojen sadece bir elektrona sahip olduğu için son derece basit bir atomdur ve kararlı olabilmesi için bu tek elektronu ikiye tamamlamaya çalışır. Bunun için hidrojen, çeşitli atomlarla bağ kurar. Atmosferde bulunan hidrojen gazı da iki hidrojen atomunun kovalent bağ ile birleşmiş halinden başka bir şey değildir. Yarı iletkenler kovalent bağı ile birbirine bağlıdır (Zengin ve Aygün , 1992).



Şekil 3.3 Amino asitlerin moleküler yapısı



Şekil 3.4 Rezidü oluşumu

Amino asitler bir arada bulunduklarında aralarında oluşturdıkları bağların sadece yaklaşık olarak %50'si peptid bağı ile olmakta, diğerlerinde ise farklı bağlarla birbirlerine bağlanmaktadır. Bu farklı bağlarla bağlandıklarında ise ortaya protein molekülü çıkmamaktadır (Temussi, 1976).

3.2 Yaşam İçin Özel Hızlandırıcılar: Enzimler

Her canlı hücrede, her biri kendi özel işini yapan, örneğin DNA kopyalanmasına yardımcı olan, besin maddelerini parçalayan, besinlerden enerji üreten, basit moleküllerden zincir yapılmasını sağlayan ve bunlar gibi sayısız işler gören binlerce enzim vardır. Enzimler, proteinlerden yapılmışlardır ve doğal olarak yalnız canlılar tarafından sentezlenirler. Hücre içerisinde meydana gelen binlerce tepkimenin hızını ve özgüllüğünü düzenlerler.

Eğer bu enzimler olmasaydı, en basitinden en karmaşığına kadar hemen hiçbir fonksiyonunuz çalışmaz ya da neredeyse duracak kadar yavaşlardı. Sonuç her iki halde de durum değişmezdi ve ölüm olurdu. Nefes alamaz, bir şey yiyemez, sindiremez, göremez, konuşamaz kısacası yaşayamazdık.

Enzimlerin en önemli görevleri vücuttaki birtakım kimyasal reaksiyonları başlatıp durdurmak ve onları hızlandırmaktır. Vücuttaki hücreler görevlerini yerine getirirken, içerdikleri kimyasalların reaksiyona girmeleri gerekir. Kimyasal reaksiyonların başlaması içinse yüksek derecede ısı gereklidir. Bu yüksek ısı ise canlı hücrelerin hayatları için tehlikeli bir durumdur; hücrelerin ölümüne neden olur. İşte bu sorunu çözenler enzimlerdir. Yüksek ısıya gerek kalmadan, enzimler kimyasal reaksiyonları başlatır veya hızlandırırlar, ancak kendileri reaksiyona girmezler (Gözükara, 1997).

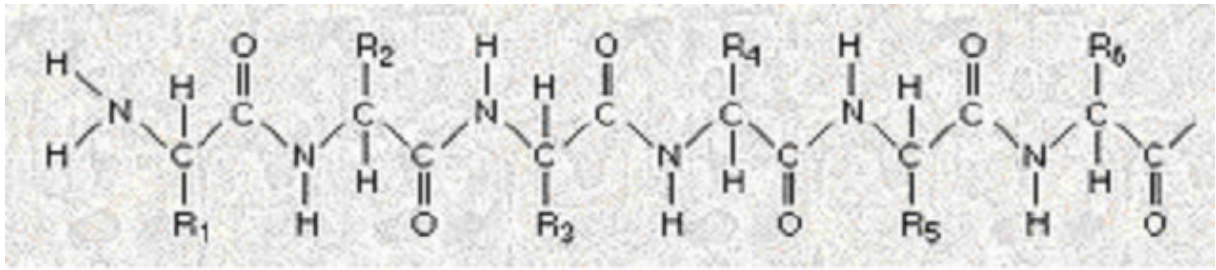
3.3 Proteinlerin Yapısı

Proteinlerin fiziksel, kimyasal ve biyolojik özelliklerini ve bu özellikler sayesinde yerine getirecekleri görevlerini, yapılarında yer alan amino asitlerin türü, sıralanışı ve bu amino asitlerin yan zincirindeki düzenlemeler belirler. Bir proteinin yapısı peptid zincirinin kendi üzerinde katlanması ile oluşturulur ve bazı durumlarda birden fazla peptid zinciri de bu işe karışabilir. Proteinler genel olarak;

1. primer (birincil)
2. sekonder (ikincil)
- 3.tersiyer (üçüncül) olmak üzere üç farklı yapıda bulunur (Karey ve Hanley, 1997) [2],[3],[4],[5].

3.3.1 Proteinlerin Primer (Birincil) Yapısı: Amino Asit Dizilimi

Primer (birincil) yapı, düz amino asit zincirlerinden (rezidülerden) meydana gelir. Protein primer yapısında fonksiyonel değildir. Ancak sekonder ve tersiyer yapılardan birine katılınca birtakım işlemlerde rol alabilir. Belirli sayı, şekil ve düzendeki amino asitler bir zincir oluştururlar. Düz amino asit zincirlerinden meydana gelir. Proteinlerin canlılık için son derece önemli olan şekillerinin en önemli belirleyicisi proteinleri oluşturan amino asitlerin sıralamasıdır. Amino asit diziliminin anormal olması, birçok genetik hastalığın da nedenidir. Bu yüzden proteinlerin birincil yapısı yani amino asitlerin doğru dizilimi son derece önemlidir (Karey ve Hanley, 1997) [2],[3],[4],[5],[6].

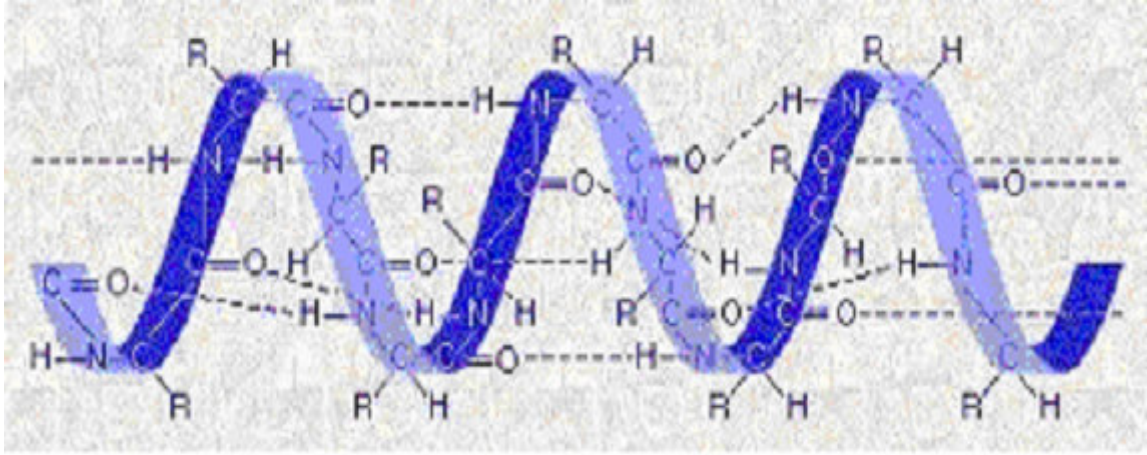


Şekil 3.5 Proteinin primer (birincil) yapısı [6]

3.3.2 Proteinlerin Sekonder (İkincil) Yapısı: Sarmal, Tabaka ve Sarım

Proteindeki tam amino asit sekansıdır; on yıllarca önce polipeptid (peptid zincirlerinin birleşmesi) zincirlerin düzenli yapılar şeklinde bazen katlandıkları bulunmuştur (yani aynı polipeptidler için şekilce farklı olan yapılar). Üç tane ikincil yapı vardır: Sarmal, Tabaka ve Sarım. Birinci yapı helikstir (sarmal) ve alfa heliks olarak ifade edilir. Her bir amino asit bir yanındaki amino asit ile oluşturduğu peptid bağın dışında hidrojen bağları da oluşturur. Bu bağların oluşma şekli amino asitlerin dizilimleri boyunca alacağı şekli ve pozisyonu belirler. Örneğin bazı durumlarda amino asit, içinde bulunduğu zincirde hidrojen bağları yaptığında sarmal bir yapı oluşturur. Amino asitler, içinde bulundukları zincirin dışından bir amino asit ile zayıf bağlar kurduklarında ise sarmal gibi tabakalı yapılar meydana gelir. Zincirleri sarmal şeklinde olan proteinler telefon kordonuna benzerler. Aynı bir telefon kordonu gibi bir eksen etrafında bir hat boyunca kıvrılırlar. Saçtaki proteinler ve bir kas proteini olan miyosin bu sarmal yapıdadır ve bunun sonucu olarak elastiktirler. Çünkü hidrojen bağları kırılabilir ve kolaylıkla tekrar oluşabilir. Günlük hayatta hidrojen bağlarının vücut proteinleri üzerinde etkisinin öğrenilmesi sayesinde çeşitli imkanlar doğmuştur. Örneğin kıvrıcık saçları düzleştirmek ya da düz saçları kıvrıcıklaştırmak için saç proteinlerindeki amino asitler

arasındaki hidrojen bağları bozulur ve yeni bağlar kurulur (Barnes, 1985) (Karey ve Hanley, 1997) [2],[3],[4],[5],[6].

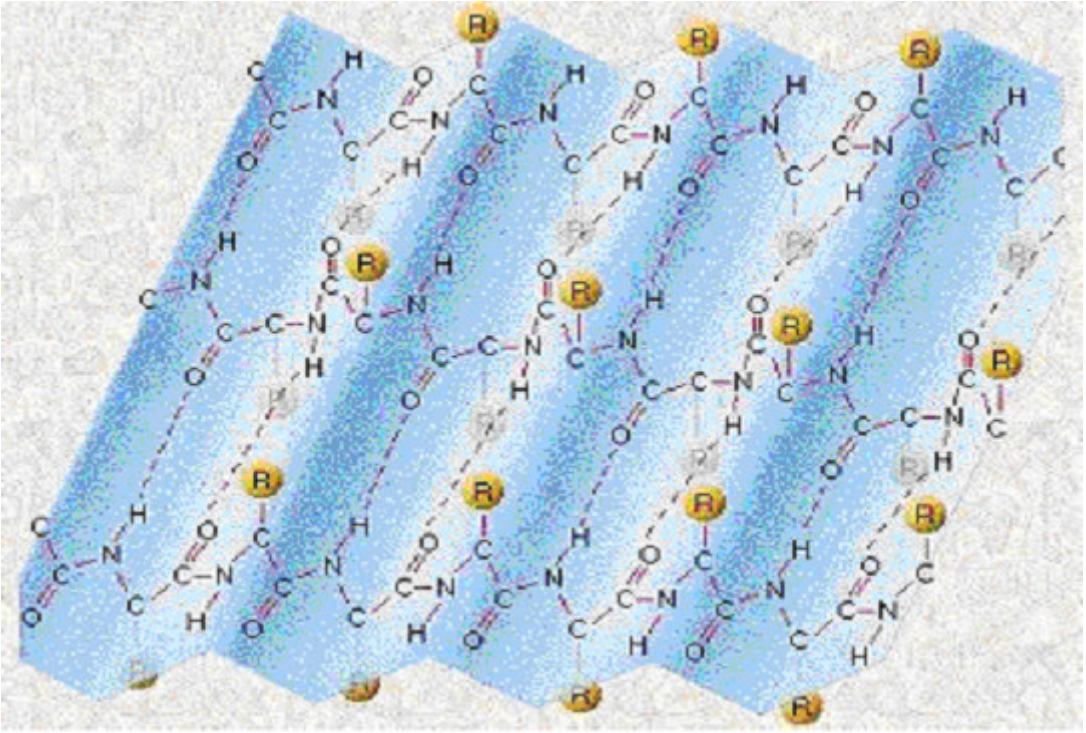


Şekil 3.6 Proteinin sarmal (helix) ikincil yapısı [6]

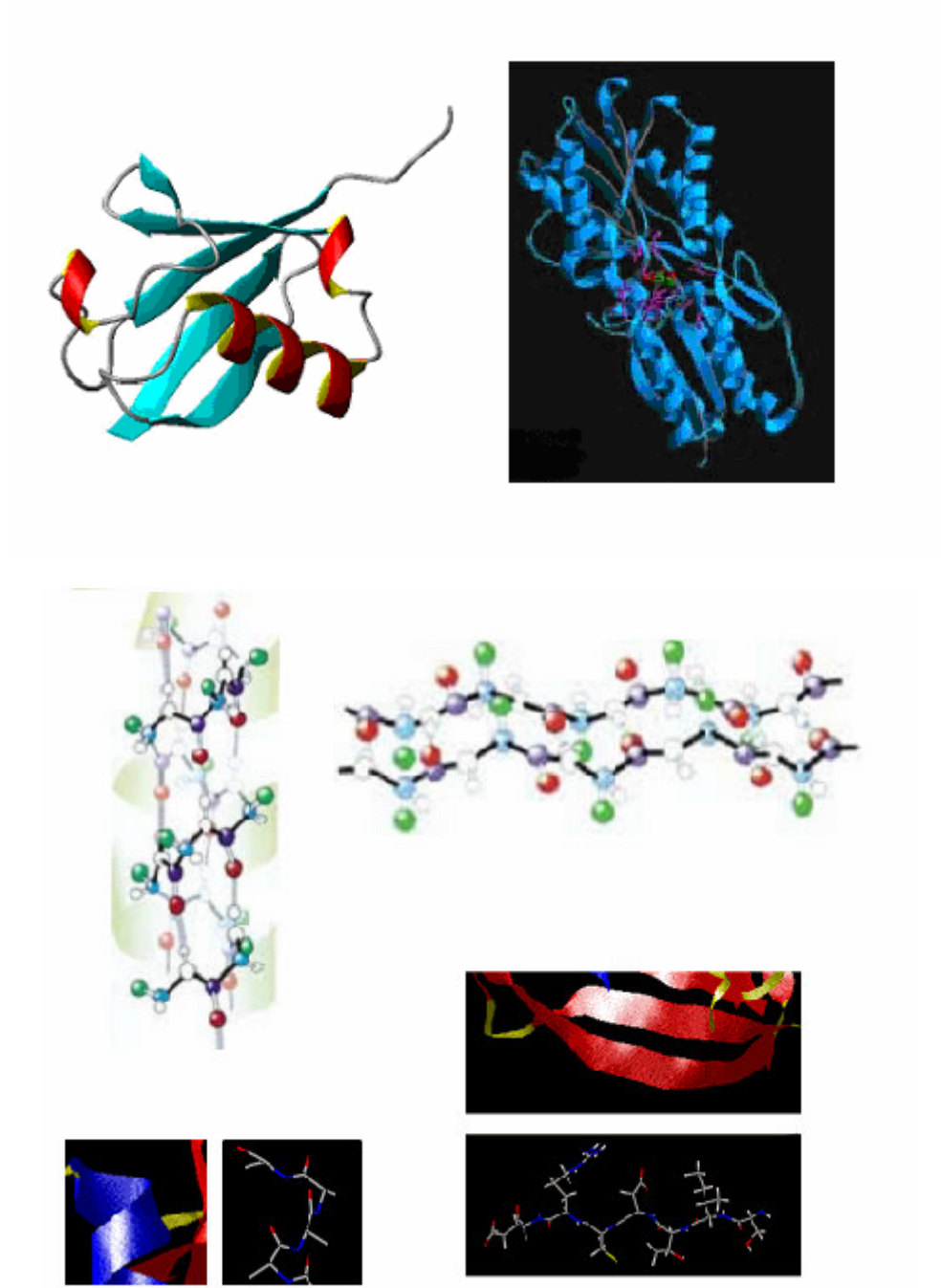
Veya başka bir şekil olarak, polipeptid zinciri ileri ve geri katlanarak levha şeklinde bir yüzey meydana getirir. Bu yapı beta sheet (tabaka) olarak adlandırılır. Sekonder yapısı merdiven şeklinde tabaka halinde olan proteinler ise sarmal yapıya sahip olanlar gibi esnek olmazlar. Bu proteinlerin omurgası bir örgü modeli gibi aşağı yukarı kıvrılır. Bunun nedeni peptid atomlarının protein zincirine dik olarak bağlanmasıdır. Bu sayede bu modele sahip olan proteinler elastik olmak yerine düz ve bükülgebilirler. Ayrıca ikincil yapısı sarmal ve tabaka olmayan çevrim veya sarım (coil) sınıfına girerler. İkincil yapısı belli olmayanlar da sarım sınıfına girerler.(Barnes, 1985) (Karey ve Hanley, 1997) [2],[3],[4],[5][6].



Şekil 3.7 Proteinin tabaka (sheet) ikincil yapısı [6]



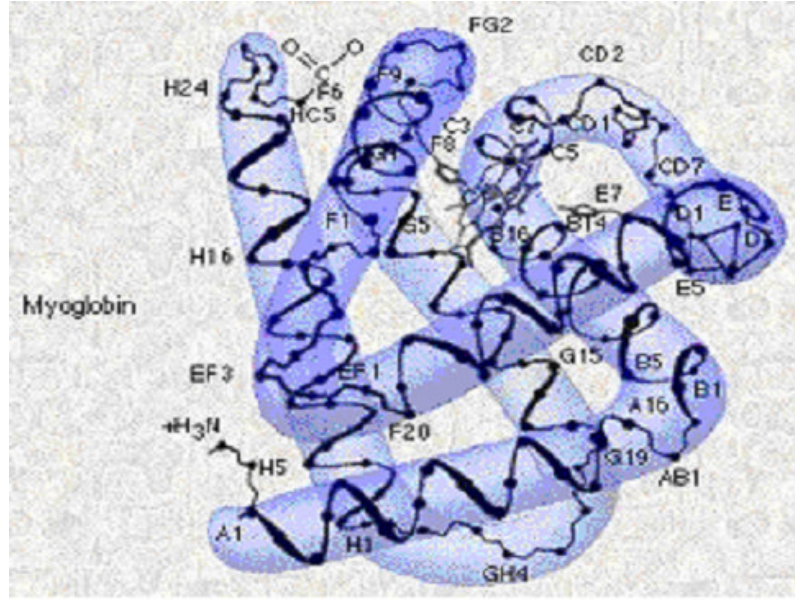
Şekil 3.8 Proteinin tabaka (sheet) ikincil yapısının ayrıntılı görünüşü [6]



Şekil 3.9 Proteinin sarmal (helix), tabaka (sheet) ve sarım (coil) ikincil yapılarının ayrıntılı görüşleri [8][3]

3.3.3 Proteinlerin Tersiyer (Üçüncül) Yapısı

İkincil yapı tiplerine eklenecek örnekler vardır ki bu örneklerde bir polipeptid katlanabilir veya bazen de bu gibi düzenli yapılar şeklinde katlanmayabilir. Proteinler sekonder yapılarında aldıkları şekilden sonra birbirlerine yaklaşan veya uzaklaşan amino asitlerin etkisiyle bükülmeye, katlanmaya ve bazen de ani dönüşler yaparak yepyeni şekiller almaya başlarlar. Ve bu şekilde, proteinin işlevi için son derece önemli olan üç boyutlu şekli meydana getirirler. Bu bükülme ve katlanmanın nedeni amino asitlerin yan zincirlerinin arasındaki etkileşimlerdir. Proteinlerdeki amino asitlerin yan zincirleri bazı etkiler sonucu birbirlerini çekerler veya iterler. Bu olaylar sayesinde, amino asitlerin bazı bölümleri birbirlerine yaklaşır, amino asit zinciri kendi üzerine katlanır, proteinlerin belirli zamanlama ve açılarla bükülmeleri sağlanır, protein molekülünün üç boyutlu yapısı kararlı kalır ve hücre dışındaki ortamda çözülmesi engellenir. Daha önce bahsedilen yapılar, tersiyer yapıya göre daha düzenlidir. Öncekilerden daha kompleks olan proteinin tersiyer yapısıdır. Gerçek bir polipeptid zincirine baktığınız zaman son şeklin ikincil özellikler tarafından şekillendirilmiştir. Bu, her şey dahil yapıya üçüncül yapı (fonsiyonel yapı) denir. (Barnes, 1985) (Karey ve Hanley, 1997) [2],[3],[4],[5][6]. Protein yapısının anlaşılmasının önemi birlikte çalışan iki faktörden gelir. Bunlardan birincisi proteinin fonksiyonunun tümüyle yapısına bağlı olduğudur. Gerçekten proteinlerin fonksiyonlarını kaybetmelerinin en genel yolu yapılarının bozulmasıdır (mesela sıcaklık ve mekanik stres ile). Sadece tam ve düzenli bir şekilde katlanmış proteinler çalışabilir. İkinci faktör ise bir proteinin yapısının deneysel olarak tanımlanmasının aşırı derecede zor olmasıdır. Bugüne kadar pek çok sekansın primer yapısı (yaklaşık 30,000) tanımlanmıştır. Ancak çok az üçüncül yapı (yaklaşık 500) tanımlanmıştır. Genel olarak proteinin ikinci yapısından, proteinin fonksiyonu hakkında bilgi edinilebilir. Belli ki üçüncül yapının tanımlanması oldukça değerlidir. Bu, protein yapısını tam olarak tahmin edebilmek için bir abartma değildir. Protein fonksiyonu; tıp, farmakoloji, kimya ve ekoloji alanlarında devrim yapacaktır [6].

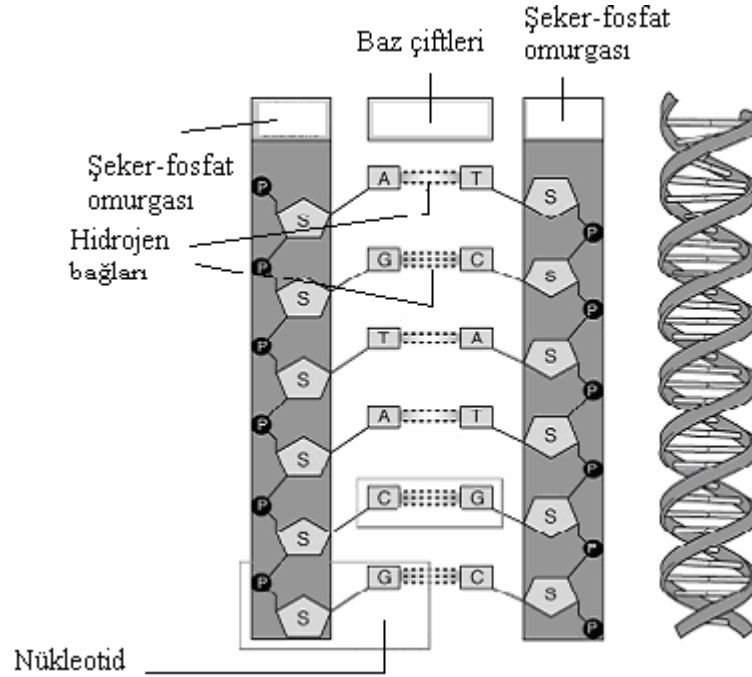


Şekil 3.10 Proteinin tersiyer yapısı [6]

3.4 DNA

DNA, ebeveynlerden çocuğa fiziksel özellikleri taşıyan genetik materyaldir. RNA ise, DNA'dan bilgiyi alıp protein sentezinde kullanılmasını sağlar. DNA ve RNA sarmallardır. DNA, dört organik bazdan, bir fosfat grubundan ve bir şeker grubundan oluşur. DNA çift zincirden oluşmuştur. Yapısındaki dört organik bazın (adenin, timin, guanin, sitozin) diziliş şekilleri ve sayıları ile genetik bilgiler belirlenir. Ebeveynlerden çocuğa aktarılabilecek genetik bilgiyi taşır. RNA protein sentezi için bilgiye ihtiyaç duyar. Eğer bu bilgi yoksa RNA sentez yapamaz. O yüzden DNA, RNA açısından çok önemli bir materyaldir. RNA, dört organik bazdan oluşmuştur. Fakat bu materyal tek zincirden oluşmuştur. Yapısında bir baza sadece bir fosfat-şeker grubu denk gelmektedir. DNA'da ise bazlar çift halinde bulunur ve bir çift baza bir fosfat-şeker grubu denk gelmektedir. RNA, DNA'dan bilgiyi alır. Bunu işleyerek protein sentezi işleminde kullanır. Çekirdekten bilgiyi alıp, sitoplazmaya taşır. Çekirdekteki DNA molekülü kromozom adlı özel kılıflarda paketlenir. Tek hücrede bulunan kromozomlarda paketlenen DNA molekülünün toplam uzunluğu bir metreyi bulur. Kromozomun toplam kalınlığı ise bir nanometre kadardır. DNA molekülü kromozom paketleri aslında çok daha küçük özel ambalaj sistemlerinden oluşur. DNA molekülü önce adeta bir ipin makaraya sarılması gibi sıkı sıkıya histon adlı özel proteinlere sarılır. Bu histon makaralara sarılmış DNA bölümleri nükleozom olarak adlandırılır. Bu nükleozom bölümleri DNA'nın korunması ve zarar görmemesi için özel olarak dizayn edilmiştir. Nükleozomlar ucuca eklendiğinde kromatinleri oluştururlar. Kromatinde iyice birbirine sarılıp kıvrılarak yoğun yumaklar

meydana getirirler. Ve böylece DNA molekülü kendi uzunluğunun milyarda biri kadar küçük olan bir yere sığdırılmış olur [7].



Şekil 3.11 DNA'nın yapısı [7]

3.4.1 Genetik Bilginin Değerlendirilmesi

Amino asitlerin proteinlere dönüşmesi sitoplazmada ve çekirdeğin dışında yer almaktadır. Bu sebeple genetik bilgi hücre içerisinde depolandığı yerden değerlendirildiği yere transfer olmaktadır. Bu transfer ve değerlendirme işlemi RNA tarafından sağlanmaktadır. RNA molekülleri hücre ve dokularda fazla miktarda bulunmakta ve proteinlerin oluşumunda görev yapmaktadır. Ribonükleik asit; RNA, DNA'ya kimyasal kompozisyon ve yapısı oldukça benzeyen nükleik asittir. DNA'dan farkı; RNA tek iplikli moleküldür. Fakat çeşitli kompleks formlara katlanabilir. RNA'da timin yerine urasil bulunmaktadır. Hücre içerisinde farklı tipte RNA molekülleri bulunmaktadır. Her biri proteinlerin sentezinde anahtar rol oynamaktadır. Bunlar elçi RNA (mRNA), transfer RNA (tRNA) ve ribosomal RNA (rRNA) 'dır.

mRNA : Bu RNA, çekirdekteki DNA ile sitoplazmada proteinin sentez edildiği yere genetik mesajları(kodon) taşımaktadır. DNA'dan bilginin alınması için çift heliks ilgili genin bölgesini ters istikamette açmakta ve açılan kollardan tek ipliği mRNA ipliğinin sentezi için kalıp görevi yapmaktadır. mRNA'nın DNA'dan sentez edilme işlemi transkripsiyon

(kopyalama) olarak bilinmektedir. Bu işlem RNA tarafından kontrol edilmektedir. Bu şekilde genetik bilgi çekirdekten sitoplazmaya geçmektedir.

tRNA : Sitoplazmada bulunmaktadır. Görevi amino asitleri toplamak ve protein sentezi kısmı olan ribozomlara taşımaktır. En az 20 farklı tRNA molekülü bulunmaktadır. Bunların hepsinin temel yapıları aynıdır. Tek iplikli RNA yaklaşık 80 baz uzunluğunda olup üçgüçl yaprağı şeklinde ve kendi üzerindeki bazlar arasındaki eşlenmeye göre katlanmaktadır. Bunlar diğer bazlarla hidrojen bağları oluşturmaz ve molekül içerisinde eşlenmemiş halkalar şeklinde bulunurlar. Her bir tRNA'nın eşlenmemiş son kısmı üçlü bir yapı içerir bu yapı antikodon olarak bilinmektedir. Antikodon, mRNA'da taşınan kodonların bir veya daha fazlasına eşlenecek özelliğe sahiptir. Diğer eşlenmemiş son kısmı ise özel amino asite bağlanır. Her bir tRNA mRNA yapısıyla antikodon yapısını karşılaştırarak amino asitleri kendine toplar.

rRNA : Ribozomların bir komponentidir. Hücre içerisinde protein sentezinin yer aldığı yapılardır. Her bir alt ünite RNA ve proteinin yaklaşık eşit büyüklükteki kısımlarından oluşmuştur. Küçük alt ünite 1500 baz uzunluktaki bir RNA molekülüne sahiptir. Geniş olan ünitesi ise 3000 ve 100 baz uzunluğundaki iki RNA'dan oluşmuştur.

Anti-kodon: t-RNA'nın m-RNA kodonlarına uygun gelen ve her biri bir amino asiti taşıyan üçlü nükleotid grubudur.

Kodon-Antikodon (Anahtar-Kilit) Metodu : Protein sentezi sırasında DNA dilinin protein bilgilerine dönüşmesi için, mRNA ile tRNA, anahtar ve kilit gibi karşı karşıya gelirler. mRNA'daki her üç harf bir kilit(kodon) sayılır. tRNA'nın bu kilidi açabilecek özellikte olan alt ucu da bir anahtar olarak tam karşısına geçer.

Transkripsiyon: DNA'daki şifrenin m-RNA'ya kopyalanması.

Translasyon: Amino asitlerin uygun bir şekilde birbirine bağlanması ile protein oluşması sürecine denir (Jones ve Karp, 1990).

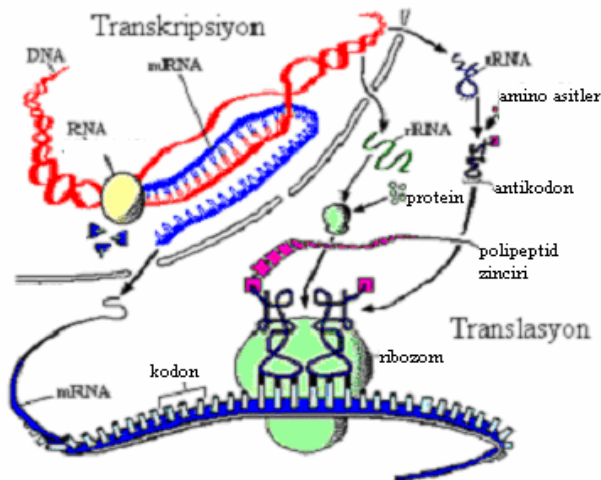
3.5 Protein Sentezi

Hücrelerde canlılığın sürdürülebilmesi için her an birçok proteine ihtiyaç duyulur. Protein üretimi sırasında ise birçok protein aynı anda faaliyet gösterir. Hücrelerin içinde protein üretimi için gereken bütün parçalar eksiksiz biçimde bir arada çalışırlar. Seksenin üzerinde ribozom proteini, yirminin üzerinde amino asit habercisi olan moleküller, bir düzinenin

üzerinde yardımcı enzim, yüzün üzerinde son işlemleri gerçekleştiren enzimler, kırkın üzerinde RNA molekülü olmak üzere yaklaşık üç yüz makro molekül, bir koordinasyon halinde protein sentezinde rol alır. Polipeptidleri oluşturmak için amino asitlerin bağlanmasıdır. Bu işlem ribozomlarda yer almaktadır. Ribozomların fonksiyonu, küçük alt ünite vasıtasıyla mRNA'ya bağlanmak, tRNA'da karşılık gelen antikodonlarla eşleşen ve tanınan kodonlardan oluşan mesajları tutmaktır. Ribozom, peptid bağıyla bağlanan amino asitlerin birleştirir. Bağ olduğu zaman ribozom mesaj boyunca devamlı hareket eder. tRNA tekrar kullanılmak üzere serbest bırakılır. tRNA ile beraber hareket eden amino asit sağ taraftan ribozoma gelir ve sol taraftan gider iken amino asiti peptid bağıyla bağlar ve serbest kalır. mRNA molekülü ribozomların bu hareketi ile aşağıya çekilirken saniyede 15 amino asit bağlar. mRNA kodon ve tRNA antikodonları arasında eşleşen bazların karşılaştırmalı yapısı, alınan DNA dizilişinin kopyasının mRNA daki mesajı tam olarak deşifre edip, polipeptid ürününde amino asitlerin doğru dizilişine çevirir.

mRNA'da AUG kodonu için transkripsiyon olur ve antikodonlu tRNA tarafından tanınır. Bir polipeptidin uzunluğu yaklaşık 333 amino asittir. Bu diziliş, ribozomlarda üç özel stop sinyali olan UAA, UGA ya da UAG olduğu zaman tamamlanır. Kodonlarıyla eşleşen tRNA'lar artık o amino asiti taşımazlar.

mRNA 'nın baz dizilişinde kopyalanan bilgi translasyon olarak bilinen polipeptid zincirinde amino asit dizilişi üretmek üzere kullanılır. Polipeptid sentez edildiğinde ribozomlar uzaklaşır ve sitoplazmada serbest kalır. Daha sonra translasyonel modifikasyon adı verilen ve diğer peptidlerle birleşme ya da katlanma şeklinde fonksiyonel proteini oluşturur.



Şekil 3.12 Protein sentezi

mRNA ipliğinin devamlı olarak ribozomlarla translasyon olduğu düşünülüyor. Amino asit peptid bağlarıyla bağlanır bağlanmaz tRNA'lar serbest bırakılır mRNA ipliğini başı serbest bırakılır ve diğer bir ribozom tarafından kullanılır. Bu ikincisi başlangıç ipliği boyunda hareket eder ve tekrar serbest kalır. Üçüncü bir ribozom devreye girer. Bir çok polipeptid birbirine benzer olup aynı zamanda bir mRNA molekülünden sentez edilebilir (Jones ve Karp, 1990).

Protein Sentezi özetlenirse; DNA dört harfli bir alfabeden (dört çeşit nükleotid(baz)) (Adenin, Guanin, Citosin ve Timin) oluşurken protein yirmi harfli bir alfabeden (20 çeşit amino asit) oluşur. Protein sentezi DNA sayesinde olur. Bunun için bu iki alfabe arasında çeviri yapılması gerekir. Bu olay kısaca şu şekilde olmaktadır. DNA'nın bir genlik kısmı bir enzim aracılığıyla RNA'ya kopyalanır. DNA bir çok genden oluşur. RNA ise bir genden oluşur. Daha sonra mRNA ribozomlara yapışır ve ribozom bu RNA'daki kod dizilimini okumaya başlar, ribozomun çevresinde tRNA'lar vardır, bu tRNA'lara ise çeşitli amino asitler bağlanmıştır. mRNA'dan hangi baz dizilimi okunduysa ona ait amino asitin bağlandığı tRNA gelir ve ribozoma yapışır, bu işlem bu şekilde devam eder ve bir protein oluşur.

3.6 Gen

Gen, kalıtım ünitesi olarak isimlendirilen özel bir yapı olarak bilinmektedir. Kromozomları ve genlerdeki bilgiyi taşıyan madde DNA'dır. Gen, polipeptid ya da RNA için kodlanan DNA molekülü boyunca nükleotid çiftlerinin dizilişidir. Gen, bir canlının hücrelerinde bulunan DNA'ların toplamıdır.

Çift heliks yapının her ikisi geni oluşturur. Sadece, bir iplik transkripsiyon olan kodlu iplik, bilgiyi içermekte ve RNA ya da amino asit olarak değerlendirmektedir. Diğer iplik taranskripsiyona uğramıyor ve karşılık olarak replikasyon (DNA'nın kendini eşlemesi) esnasında görev yapmaktadır. Kodlanan ipliğin kromozom boyunca aynı iplik olması zorunlu değildir. Bazen değişebilir. Genin üzerinde bulunup da protein yapısına girmeyecek bölgeler vardır. Gen yapısında bulunup proteine dönüşmeyen bu kısımlara intron denir. Proteine dönüşen bölgelere ise (protein sentez kodu içeren) ekson denir. Bu bölgeler birbirinden özel dizilerle ayrılır [7].



Şekil 3.13 Genin yapısı [7]

Poliipeptid kodlarından oluşan genler, iki ana gruba ayrılmaktadır. Bunlar; fonksiyonel proteinler için kodlanan yapısal genler (enzim, hormon, hücre yapı komponentleri, depo proteinleri vs) ikincisi düzenleyici genler bu genler diğer genlerin aktivasyonlarını kontrol ederler. Poliipeptid zincirleri farklı uzunlukta olup genetik koda bağlı olarak uzunluğu değişmektedir. Ortalama poliipeptid zincirinin uzunluğu 333 amino asit olup yapısal bir genin ortalama büyüklüğü yaklaşık 1000 nükleotid çifti DNA dizisidir.

Transfer ve ribosomal RNA molekülleri genler tarafından kodlanır fakat bunlar proteinler gibi aynı yolla sentez edilmezler. tRNA ve rRNA doğrudan DNA 'dan transkripsiyonla oluşturulurlar mRNA 'da aynı yolla ortaya çıkmaktadır. tRNA molekülleri 80 nükleotid uzunluğunda ve bu genler DNA da ilgili uzunluğa göre kodlanırlar. rRNA üç farklı büyüklükte olup 100, 1500 ve 3000 baz olmak üzere DNA'dan kodlanır. rRNA için kodlanan genler çok fazla kopyası yapılır (Jones ve Karp, 1990).

3.6.1 Genetik Kodun Özellikleri

Belli amino asitler (methionin ve triptophan) sadece bir kodon tarafından kodlanmakta fakat çoğu birkaçı tarafından kodlanmaktadır. Bundan dolayı kodun farklı olduğu söylenebilir çünkü ihtiyaçtan daha fazla bulunmakta ve amino asitten fazla kodon bulunmaktadır. Bu farklılık kodonun üçüncü bazındaki varyasyonlarla izah edilmektedir. Örneğin GCU, GCC, GCA ve GCG kodlarının hepsi alanin içindir. tRNA'nın alanin taşımakta ve bundan dolayı mRNA'daki ilk iki bazı tanımlamaktadır. Üçüncü olanı ise nispeten önemsizdir. Bazı kodonlar herhangi bir amino asidi kodlamaz, bunlar stop kodonları veya zinciri sınırlandıran kodonlardır. UAA, UAG, ve UGA bunlara örnektir. Bunlar translasyon olayında sinyal olarak rol oynarlar ve poliipeptid zincirinin oluşumunun sonunu belirler. Stop kodonları bir geni

komşularından ayıran bir noktalama işareti olarak düşünülüyor. Kod devamlı olup üst üste gelmez. Şayet üst üste gelirse o zaman bir baz çifti değişikliği olur yani mutasyon gerçekleşir. Mutasyon polipeptid zincirinde oluşursa bir amino asitten fazlasında da oluşabilir. Ancak bu her zaman böyle değildir. Kodlar genel olarak her canlıda bulunurlar. Aynı kodonlar bütün yaşayan organizmalarda aynı amino asidi oluşturmak üzere kullanılır (Jones ve Karp, 1990).

4. YAPAY SİNİR AĞLARI

4.1 Yapay Sinir Ağlarının Özellikleri

Yapay sinir ağları, temelde tamamen insan beyni örneklenerek geliştirilmiş bir teknolojidir. Bilindiği gibi; öğrenme, hatırlama, düşünme gibi tüm insan davranışlarının temelinde sinir hücreleri bulunmaktadır. İnsan beyninde tahminen 10^{11} adet sinir hücresi olduğu düşünülmektedir ve bu sinir hücreleri arasında sonsuz diyebileceğimiz sayıda sinaptik birleşme denilen sinirler arası bağ vardır (Ergezer, Dikmen ve Özdemir, 2003), [9]. YSA, eğitmeye / örnek bilgiye dayanan bir fonksiyon geliştirir. Tasarımdan ziyade eğitmeyle hesaplama yapıları sağlayabilir (Yıldırım, 2003).

Yapay sinir ağları temelde aşağıdaki özelliklere sahiptir:

YSA' nın hesaplama ve bilgi işleme gücünü, paralel dağılmış yapısından, öğrenebilme ve genelleme yeteneğinden aldığı söylenebilir. YSA biyolojik sinir ağlarından esinlenerek geliştirilmiş, bilgi işlem sistemleridir. Bir sinir ağı, bilgiyi depolamak için doğal eğilimi olan basit birimlerden oluşan paralel dağıtılmış bir işlemcidir. Bu ağlar sinaptik ağırlıklar olarak bilinen nöronlar arası bağlantı kuvvetlerini, bilgiyi depolamak için kullanır. Benzer olarak YSA da birçok işlemciden oluşur ve farklı biçimlerde ifade edilebilen nümerik verileri taşıyan bağlantı yada ağırlıklarla birbirine bağlı yapılardır (Sağıroğlu, Beşdok ve Erler, 2003).

Genelleme, eğitim ya da öğrenme sürecinde karşılaşılmayan girişler için de YSA' nın uygun tepkileri üretmesi olarak tanımlanır. Bu üstün özellikleri, YSA' nın karmaşık problemleri çözebilme yeteneğini gösterir. Günümüzde birçok bilim alanında YSA, aşağıdaki özellikleri nedeniyle etkin olmuş ve uygulama yeri bulmuştur.

4.1.1 Doğrusal Olmama

YSA' nın temel işlem elemanı olan hücre, doğrusal değildir. Dolayısıyla hücrelerin birleşmesinden meydana gelen YSA da doğrusal değildir ve bu özellik bütün ağa yayılmış durumdadır. Bu özelliği ile YSA, doğrusal olmayan karmaşık problemlerin çözümünde en önemli araç olmuştur.

4.1.2 Öğrenme

YSA' nın arzu edilen davranışı gösterebilmesi için amaca uygun olarak ayarlanması gerekir. Bu, hücreler arasında doğru bağlantıların yapılması ve bağlantıların uygun ağırlıklara sahip olması gerektiğini ifade eder. YSA' nın karmaşık yapısı nedeniyle bağlantılar ve ağırlıklar önceden ayarlı olarak verilemez ya da tasarlanamaz. Bu nedenle YSA, istenen davranışı gösterecek şekilde ilgilendiği problemten aldığı eğitim örneklerini kullanarak problemi öğrenmelidir.

4.1.3 Genelleme

YSA, ilgilendiği problemi öğrendikten sonra eğitim sırasında karşılaşmadığı test örnekleri için de arzu edilen tepkiyi üretebilir. Örneğin, karakter tanıma amacıyla eğitilmiş bir YSA, bozuk karakter girişlerinde de doğru karakterleri verebilir ya da bir sistemin eğitilmiş YSA modeli, eğitim sürecinde verilmeyen giriş sinyalleri için de sistemle aynı davranışı gösterebilir.

4.1.4 Uyarlanabilirlik

YSA, ilgilendiği problemdeki değişikliklere göre ağırlıklarını ayarlar. Yani, belirli bir problemi çözmek amacıyla eğitilen YSA, problemdeki değişimlere göre tekrar eğitilebilir ve değişimler devamlı ise gerçek zamanda da eğitime devam edilebilir. Bu özelliği ile YSA, uyarlamalı örnek tanıma, sinyal işleme, sistem tanılama ve denetim gibi alanlarda etkin olarak kullanılır.

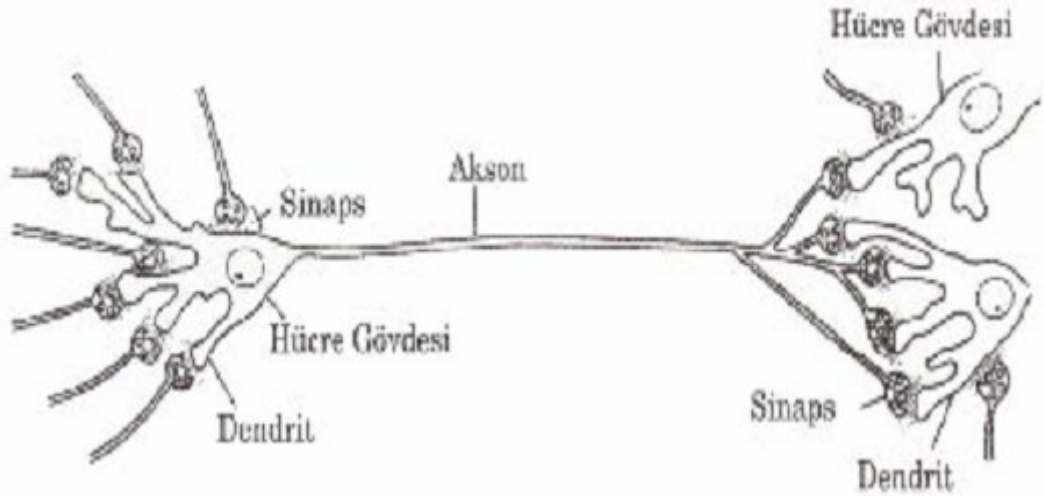
4.1.5 Hata Toleransı

YSA, çok sayıda hücrenin çeşitli şekillerde bağlanmasıyla oluştuğu için paralel dağılmış bir yapıya sahiptir ve ağına sahip olduğu bilgi, ağıdaki bütün bağlantılar üzerine dağılmış durumdadır. Bu nedenle, eğitilmiş bir YSA'nın bazı bağlantılarının hatta bazı hücrelerinin etkisiz hale gelmesi, ağına doğru bilgi üretmesini önemli ölçüde etkilemez. Bu nedenle, geleneksel yöntemlere göre hatayı tolere etme yetenekleri son derece yüksektir (Ergezer, Dikmen ve Özdemir, 2003), [9].

4.2 Yapay Sinir Ağları ve Biyolojik Beyinler

Çizelge 4.1 Bilgisayar ile beynin karşılaştırılması

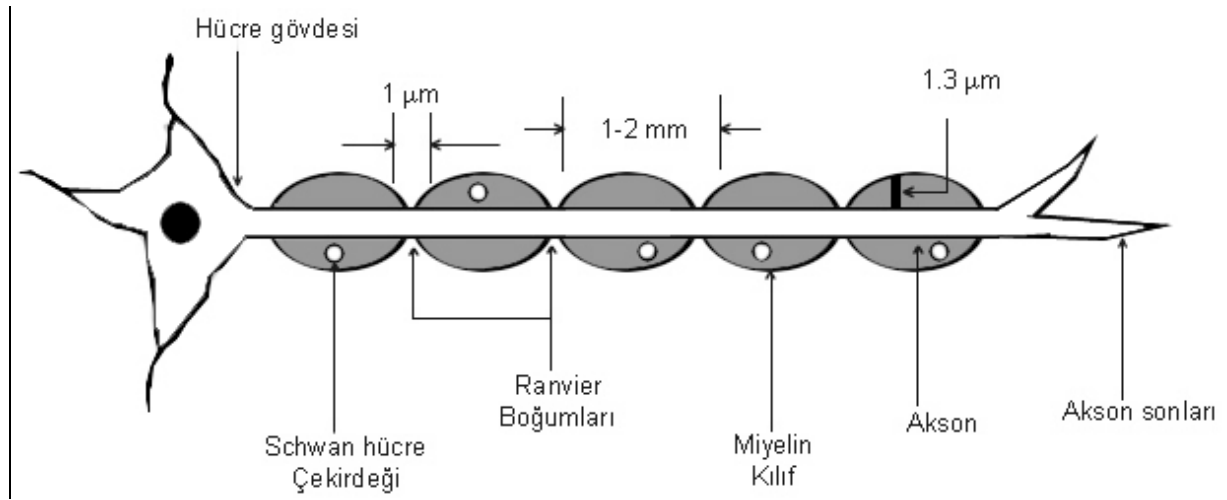
Bilgisayar	Beyin
Seri çalışma	Paralel çalışma
Hızlı birimler	Yavaş birimler
Sayısal	Analog
Tam tanımlı komut dizini	Eksik veya bozulmuş bilgiye adapte olabilir
Giriş verilerine duyarlı	Giriş verilerine duyarsız
Her adım önemli	Birimlerin yapılan ana işe katkısı azdır



Şekil 4.1 Biyolojik nöron

Akson : Hücre çıktısını göndermeye yarayan uzantısıdır. Bir hücrenin tek bir akson uzantısı bulunur. Ancak bu akson uzantıdan çıkan çok sayıda uzantı ve bunların ucunda sinaptik bağlantılar bulunur. Uzunlukları birkaç mikrondan, 1-2 metreye kadar değişebilir. Her nöronun yalnızca bir aksonu vardır. Aksonlar akson kesecikleri denilen bir çok yapı ile sonlanırlar. Bu keseciklerde sinaptik geçişte rol alan kimyasal taşıyıcılar (nörotransmitter) bulunur. Aksonlar özel bir örtüye sahip olmalarına göre miyelinli yada miyelinsiz olarak

sınıflandırılabilirler. Akson üzerini örten miyelin kılıfın, yalıtım ve darbe hızını arttırmak gibi iki önemli görevi vardır. Sonraki sayfada miyelinli bir nöronun yapısı gösterilmiştir. Sinir hücreleri, dış kısmı pozitif, iç kısmı negatif yüklü ve ortasında yalıtılmış, kutuplanmış bir kondansatöre benzer. Bir uyarıcı ile bu kondansatör nötrlenebilir hatta anlık olarak ters kutuplanabilir. Zarın belirli bir bölgesinde 1ms kadar süren bu potansiyel değişiminden sonra zar bu bölgede eski haline dönerken, potansiyel değişikliği zar boyunca yayılmaya devam eder. Akson boyunca iletilen bilginin temel birim olan bu potansiyel değişimine aksiyon potansiyeli denir.

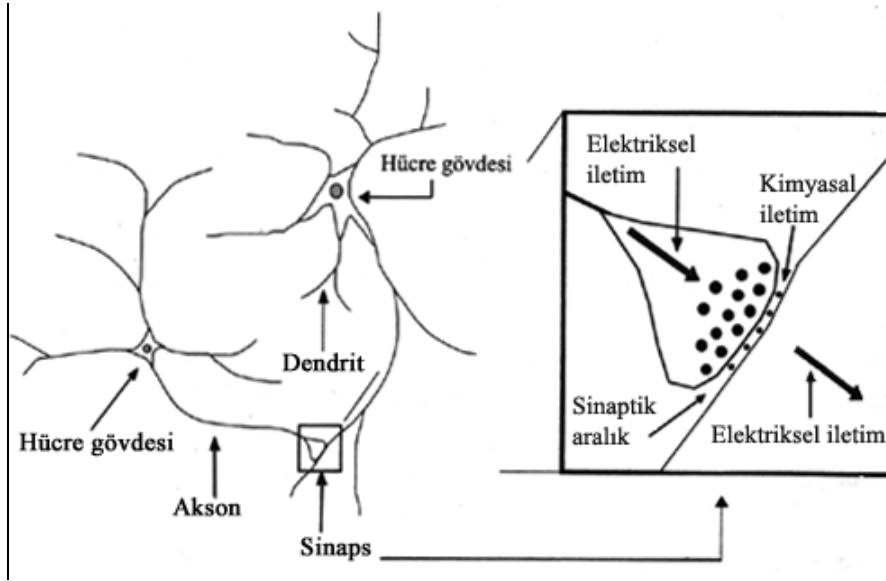


Şekil 4.2 Miyelinli bir nöron yapısı [15]

Dendrit : Nöronun ağaç köküne benzeyen, görevi hücreye girdilerin sağlanması olan uzantılardır.

Sinaps : Sinapslar, sinir hücrelerindeki aksonların, diğer sinir hücreleri ve/veya onların dendritleri üzerinde sonlanan özelleşmiş bağlantı noktalarıdır. Bu bağlantı noktalarının görevi aksondaki elektriksel iletinin diğer hücelere aktarılmasıdır. Bu bağlantı noktalarında iletiler elektro-kimyasal süreçlerle diğer hücelere geçirilir. Sinapslar bağlandıkları dendrite veya nöronda bölgesel olarak elektrik kuvvetini pozitif veya negatif yönde etkileyebilme yeteneğine sahiptirler. Böylelikle bir nöronun diğerini etkileyebilmesi söz konusu olmaktadır. Sinir hücreleri arasında iletişimin gerçekleştiği, yapısal ve fonksiyonel olarak özelleşmiş bölgelere sinaps adı verilir. Mesajı gönderen ve presinaptik hücre olarak adlandırılan hücre ile mesajı alan ve postsinaptik hücre bu kavşak bölgelerinde bir birlerine oldukça yaklaşırlar

(Cichocki ve Unbehaven, 1993).



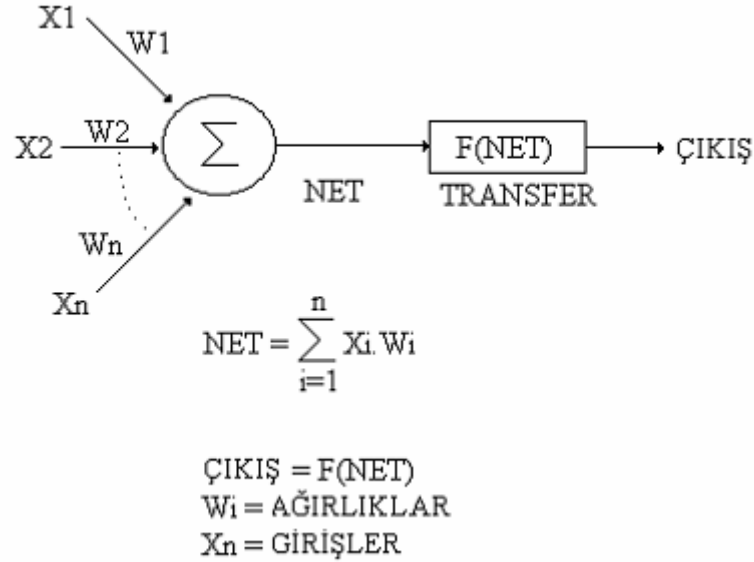
Şekil 4.3 Sinaps [15]

Soma : Bir nöronun gövdesine soma adı verilir. Soma hücre çekirdeğini içermektedir. Hücrenin yaşamasını sağlayan işlevleri görür. Sinapslar aracılığıyla dendritlere geçirilen iletiler birleşerek akson üzerinde elektriksel bir çıktı oluştururlar. Bu çıktının olup olmayacağı veya çıktının elektriksel olarak yoğunluğu, sinapsların etkileri sonucu hücreye gelen tüm girdilerin, toplam değeri tarafından belirlenmektedir. Somaya gelen girdilerin ağırlıklı toplamı akson üzerinde çıktı oluşturacak değere ulaştığında, bu değere "eşik değeri" adı verilmektedir ve nöron ateşlendi (fired) olarak ifade edilmektedir. Bu şekilde girdiler nöron tarafından değerlendirilerek çıktıya dönüştürülmüş olur (De Wilde, 1997).

Çizelge 4.2 Sinir sistemi ile YSA'nın benzerlikleri

Sinir Sistemi	YSA Sistemi
Nöron	İşlem elemanı
Dendrit	Toplama fonksiyonu
Hücre gövdesi	Transfer fonksiyonu
Aksonlar	Eleman çıkışı
Sinapslar	Ağırlıklar

4.3 Yapay Sinir Ağı Modeli



Şekil 4.4 İşlem elemanın yapısı

Tek yönlü işaret kanalları ile birbirine bağlanan işlemsel elemanlardan oluşur. Çıkış işareti tek olup kopyalanabilir. Daha önce öğrendiği bilgiyi eksik veya bozuk giriş verildiği zaman bile yeniden üretebilir. Lineer olmayan bir karakteristiğe sahip olmaları nedeniyle gerçek dünya problemlerine daha doğru çözümler getirebilirler (Yıldırım, 2003).

Yapay sinir ağında ilk katman girdi katmanıdır ve dışarıdan verilerin yapay sinir ağına alınmasını sağlar. Diğer katman ise çıktı katmanıdır. Bu katmanda bilgilerin dışarıya iletildiği katmandır. Girdi ile çıktı katmanları arasında katman varsa bu katmanlara gizli katman (hidden layer) adı verilir. Bir yapay sinir ağında gizli katman olması gerekmediği gibi, birden fazla gizli katman da bulunabilir. Hemen hemen tüm çalışmalarda girdi katmanı, girdilerin ağına girmesinde bir kapı işlevi görmesi nedeniyle katman olarak sayılmamaktadır. Bir yapay sinir ağındaki en önemli unsurlardan biri de bağlantılardır. Nöronların bir birilerine veri aktarımını sağlayan bağlantılar aynı zamanda bir ağırlık (weight) değerine sahiptirler. Ağırlıklar bir nöronda girdi olarak kullanılacak değerlerin göreceli kuvvetini (matematiksel katsayısını) gösterir. Yapay sinir ağı içinde girdilerin nöronlar arasında iletimini sağlayan tüm bağlantıların farklı ağırlık değerleri bulunmaktadır. Böylelikle ağırlıklar her işlem

elemanının her girdisi üzerinde etki yapmaktadır.

Yapay sinir ağı öğrenme sürecinde, gerçek hayattaki problem alanına ilişkin veri ve sonuçlardan, bir başka deyişle örneklerden yararlanır. Gerçek hayattaki problem alanına ilişkin değişkenler yapay sinir ağının girdi dizisini, bu değişkenlerle elde edilmiş gerçek hayata ilişkin sonuçlar ise yapay sinir ağının ulaşması gereken hedef çıktıların dizisini oluşturur. Bu eğitim setindeki girdi ile çıktı arasındaki ilişkiyi belirleyen örüntü, yapay sinir ağının öğrenmesi istenen örüntüdür ve yapay sinir ağının ağırlık değerleri bu örüntüyü yansıtır. Yapay sinir ağının eğitilebilmesi için çok sayıda girdi ve girdilere ilişkin çıktı dizisine gereksinim duyulur. Yapay sinir ağının eğitilmesinde kullanılan girdi ve çıktı dizileri çiftinden oluşan verilerin tümüne "eğitim seti" adı verilir. Yapay sinir ağının öğrenme sürecinde yapılan temel iş, ağırlık değerlerinin değiştirilmesidir. Amaç tüm girdi dizileri için çıktı dizisinin doğrulukla üretilebileceği şekilde yapay sinir ağının ağırlıklarının düzenlenmesidir (Vemeri, 1992).

Bunu nörona gelen girdilerin katsayılarının düzenlenmesi olarak görmek mümkündür. Böylelikle yapay sinir ağı, kullanılan girdi ve çıktılara bağlı olarak, gerçek hayattaki örüntüyü temsil eder duruma gelmektedir. Öğrenme sürecinde yapay sinir ağının istenen çıktıları üretecek şekilde ağ içindeki ağırlıkların düzenlenmesini sağlayan mekanizma "öğrenme algoritması" (learning algorithm) veya "öğrenme kuralı" (learning rule) olarak adlandırılır (Rojas, 1996).

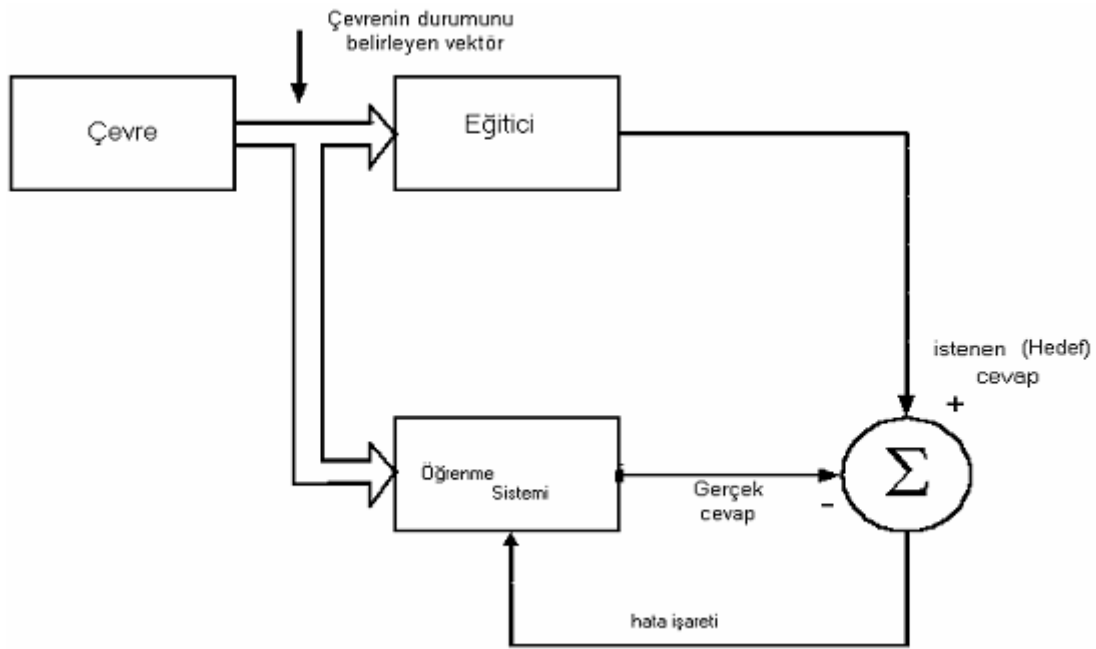
Öğrenme sürecinin başlangıcında yapay sinir ağının ağırlıkları rastgele atanmış durumdadır. Girdiler, girdi katmanından başlayarak gizli katmanlara ve çıktı katmanına işlenerek geçirilir. Böylelikle yapay sinir ağı, ağırlıklar ile toplam ve geçiş fonksiyonlarının etkisi altında bir çıktı dizisi üretmiş olur. Bu çıktılar ile hedef çıktılar arasında hesaplanan fark "hata" olarak adlandırılır. Bu hata, yapay sinir ağının ağırlıkları ile istenen çıktılar arasındaki farkın giderilmesi için ağ içinde kullanılmaktadır (Başbuğ, 1994).

Eğitim sırasında yapay sinir ağının ağırlıklarının derece, derece değiştirilmesi ve yapay sinir ağının istenen çıktıları üretecek şekilde uygun ağırlıkların bulunması süreci, çok sayıda tekrarlanan döngüler gerektirir (Rojas, 1996). Eğitim setinin bu döngü içinde yapay sinir ağından bir defa geçmesi devir (epoch) olarak adlandırılır (Başbuğ, 1994). Bir yapay sinir ağının öğrenme sürecinin tamamlanması için gerekli devir sayısı oldukça değişkendir. Tanınması gereken örüntü karmaşıklaştıkça gerekli devir sayısı artar, ancak kullanılan yapay sinir ağı mimarisi ve öğrenme algoritması da devir sayısını belirleyen önemli bir etkidir.

4.4 Yapay Sinir Ağlarında Kullanılan Öğrenme Algoritmaları

4.4.1 Öğreticili-Eğiticili-Denetimli Öğrenme (Supervised Learning)

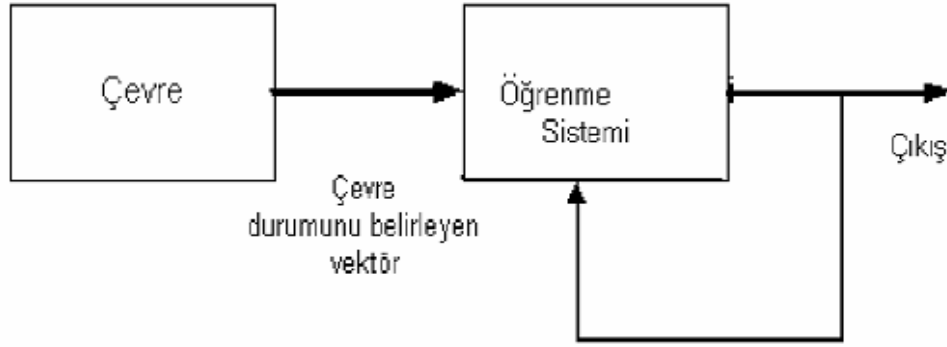
Öğreticili öğrenmede bir dış öğreticinin ağı her bir girdi için ağın üretmesi beklenen "doğru" sonuçları hedef çıktı olarak verdiği düşünülür. Bu öğrenmede ağın ürettiği çıktılar ile hedef çıktılar arasındaki fark hata olarak ele alınır ve bu hata minimize edilmeye çalışılır (Hertz ve Krogh, 1991).



Şekil 4.5 Denetimli öğrenme

4.4.2 Öğreticisiz-Eğiticisiz-Denetimsiz Öğrenme (Unsupervised Learning)

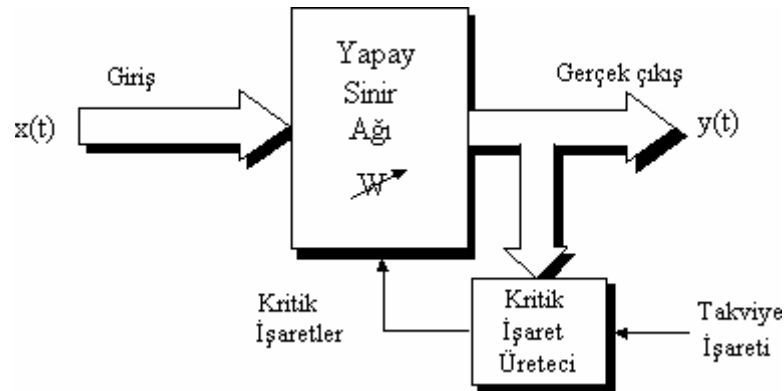
Bu tür öğrenmede ağı sadece girdiler verilir. Ağın ulaşması gereken hedef çıktılar bulunmaz. Ağ bu girdiler arasındaki ilişkiyi bir başka ifadeyle girdi setindeki örüntüyü öğrenmeye çalışır. Böylelikle ağ kendini girdi setindeki örüntüye uyumlaştırır.



Şekil 4.6 Denetimsiz öğrenme

4.4.3 Takviyeli Öğrenme (Reinforcement Training)

Öğreticili öğrenme yöntemine benzemekle birlikte, ağa hedef çıktılar yerine, ağın çıktılarının ne ölçüde doğru olduğunu belirten bir skor veya derece bildirilir (Hecht-Nielsen, 1989).



Şekil 4.7 Takviyeli öğrenme [15]

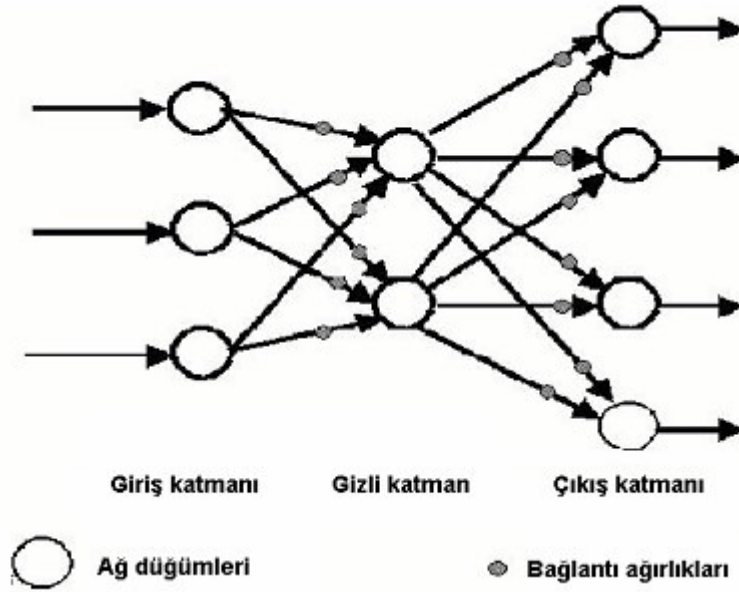
4.5 YSA Ağ Yapıları

Yapay sinir ağlarında, nöronlar arası bağlantı farklılıklarından kaynaklanan, değişik yapay sinir ağı mimarileri bulunmaktadır. Bu mimariler arasında performans ve karakteristik özellikleri bakımından farklar bulunmaktadır. Mimari farklılıklar özellikle ağın modelleme yeteneğini belirledikleri için oldukça önemlidirler. Yapay sinir ağının tasarımı aşamasında bu mimariler arasından uygulamaya en elverişli olanı seçilir.

Yapay sinir ağları yapı olarak iki ana gruba ayrılır.

4.5.1 İleri Beslemeli Ağlar

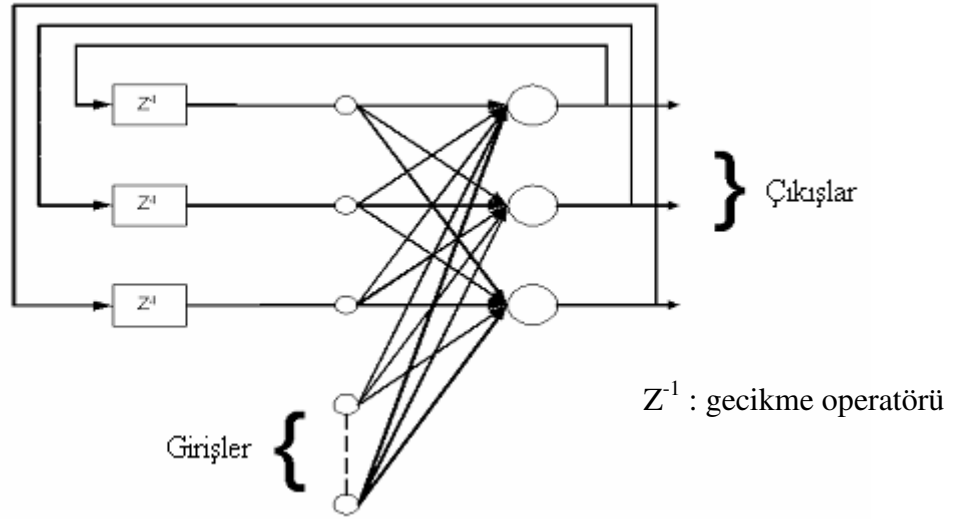
İleri beslemeli ağlarda, işlem elemanları birbiri ardı sıra sıralanırlar. Girdiler dışarıdan ya da bir önceki nörondan alınır; nöronların çıktıları da bir sonraki nörona veya dışarıya verilir. Bu ağlarda, nöronun önceki veya aynı katmandaki herhangi bir nörona bağlantısı ve veri iletmesi söz konusu değildir. Bu mimarilerde ağın çıktısı, tamamen ağa giren girdilere bağlıdır. İleri beslemeli ağlar herhangi bir dinamiklik özelliği taşımazlar ve gösterdikleri özellik bakımından doğrusal ve doğrusal olmayan kararlı problem alanlarına uygulanabilirler.



Şekil 4.8 İleri beslemeli ağ

4.5.2 Geri Beslemeli Ağlar

Geri beslemeli ağlarda işlem elemanlarının çıktıları daha gerilerdeki işlem elemanlarına girdi olarak kabul edilir. Geri beslemeli ağları ileri beslemeli ağlardan ayıran temel özellik, bu tip mimariyle oluşturulan ağlar dinamik doğrusal olmayan sistemleri, özellikle doğrusal olmayan diferansiyel denklemleri modelleyebilmeleridir (Cichocki ve Unbehaven, 1993).

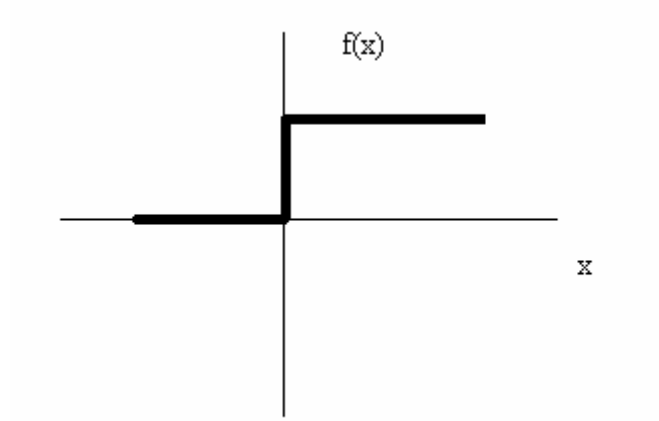


Şekil 4.9 Geri beslemeli ağ

4.6 Aktivasyon Fonksiyonları

Hücre modellerinde, hücrenin gerçekleştireceği işleve göre çeşitli tipte aktivasyon fonksiyonları kullanılabilir. Aktivasyon fonksiyonları sabit parametrelili yada uyarlanabilir parametrelili seçilebilir. Sonraki sayfalarda, hücre modellerinde yaygın olarak kullanılan çeşitli aktivasyon fonksiyonları tanıtılmıştır. Çıkışta bir nöronun ne düzeyde uyarılacağını belirleyen bir fonksiyondur. Bir nöronun hangi uyarılma düzeyinden sonra çıktı üreteceği ve çıktının uyarılma düzeyi ile ilişkisi ise, aktivasyon fonksiyonu tarafından belirlenmektedir. Bu ilişki çıktının hangi aralıklarda olacağını belirlemesi açısından da önemlidir. Çünkü toplam fonksiyonu tarafından üretilen değer oldukça büyük olabilir ve değer çıktı için kullanıldığında, birkaç katman sonra çıktı değeri çok büyük değerlere ulaşacağı için çıktının 0, 1 veya bazen -1, +1 aralığına dönüştürülmesi gerekmektedir (Trippi ve Turban, 1996). Bazı Aktivasyon Fonksiyonları;

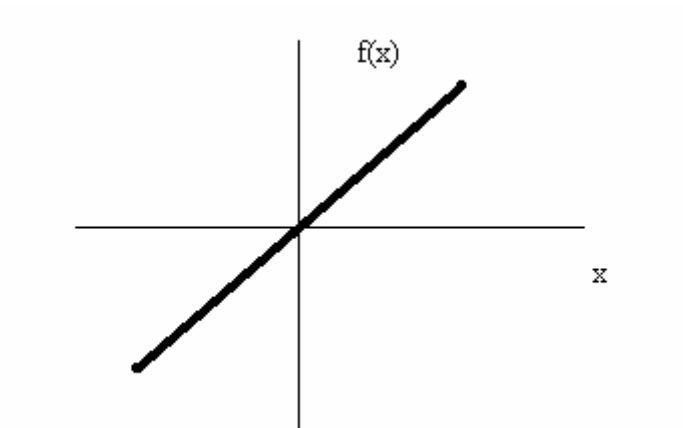
4.6.1 Eşik (Threshold)



Şekil 4.10 Eşik fonksiyonu

$$f(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (4.1)$$

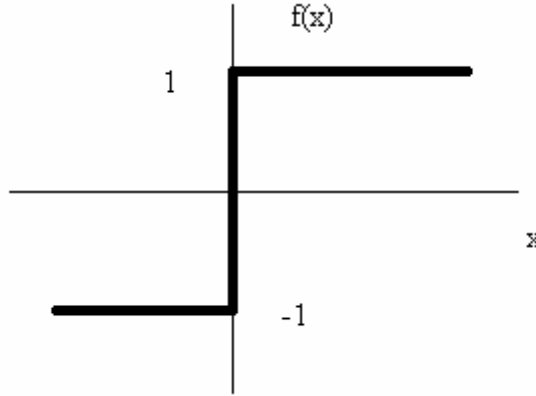
4.6.2 Parçalı lineer (Purelin)



Şekil 4.11 Parçalı lineer fonksiyonu

$$f(x) = x \quad (4.2)$$

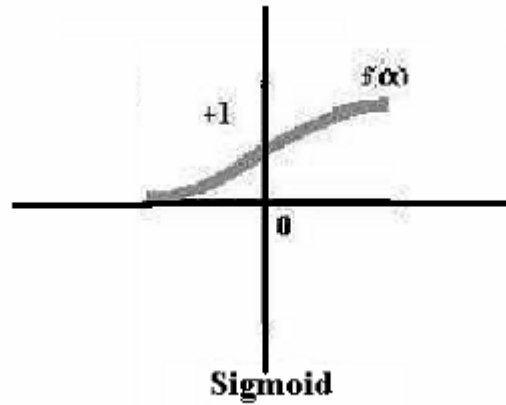
4.6.3 Signum (Sign)



Şekil 4.12 Signum fonksiyonu

$$f(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases} \quad (4.3)$$

4.6.4 Sigmoid (Logsig)



Şekil 4.13 Sigmoid fonksiyonu

$$f(x) = \frac{1}{1 + \exp(-\lambda x)} \quad (4.4)$$

4.6.5 Yarışmalı (Competitive)

$$f(x)=\begin{cases} 1 & \text{maksimum } x' \text{ e sahip nöron} \\ 0 & \text{tüm diğer nöronlar} \end{cases} \quad (4.5)$$

4.7 Ağ Modelleri

4.7.1 Statik Ağlar

Hafızasız düğüm denklemleriyle karakterize edilen ağlardır. Yani çıkışların geçmiş veya gelecek giriş ve çıkışların değil, sadece o anki girişin bir fonksiyonudur.

4.7.2 Dinamik Ağlar

Hafızalı sistemlerdir. Düğüm denklemleri tipik olarak diferansiyel veya fark denklemleriyle tanımlanırlar. Üç farklı gruba ayrılırlar. Bunlar;

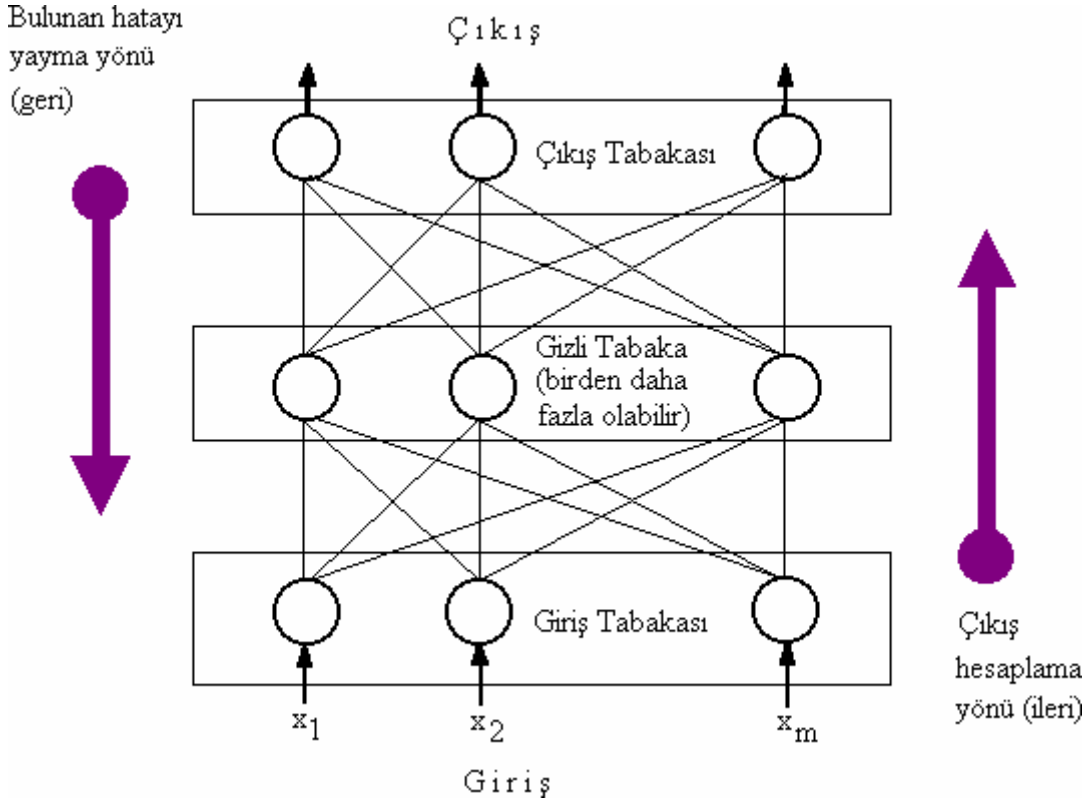
- i) İleri beslemeli dinamiğe sahip ağlar
- ii) Çıkış geri beslemeli ağlar
- iii) Durum geri beslemeli ağlar (Yıldırım, 2003).

4.8 Öğrenme Kuralları

4.8.1 Çok Katmanlı Algılayıcı (MLP: Multilayer Perceptron)

Bu ağ modeli özellikle mühendislik uygulamalarında en çok kullanılan sinir ağı modeli olmuştur. Bir çok öğretim algoritmasının bu ağı eğitmede kullanılabilir olması, bu modelin yaygın kullanılmasının sebebidir. Bir MLP modeli, bir giriş, bir veya daha fazla ara ve bir de çıkış katmanından oluşur. Bir katmandaki bütün işlem elemanları bir üst katmandaki bütün işlem elemanlarına bağlıdır. Bilgi akışı ileri doğru olup geri besleme yoktur. Bunun için ileri beslemeli sinir ağı modeli olarak adlandırılır. Giriş katmanında herhangi bir bilgi işleme yapılmaz. Buradaki işlem elemanı sayısı tamamen uygulanan problemler giriş sayısına bağlıdır. Ara katman sayısı ve ara katmanlardaki işlem elemanı sayısı ise, deneme-yanılma yolu ile bulunur. Çıkış katmanındaki eleman sayısı ise yine uygulanan probleme dayanılarak

belirlenir. MLP ağlarında, ağa bir örnek gösterilir ve örnek neticesinde nasıl bir sonuç üreteceği de bildirilir (danışmanlı öğrenme). Örnekler giriş katmanına uygulanır, ara katmanlarda işlenir ve çıkış katmanından da çıkışlar elde edilir. Kullanılan eğitime algoritmasına göre, ağın çıkışı ile arzu edilen çıkış arasındaki hata tekrar geriye doğru yayılarak hata minimuma düşünceye kadar ağın ağırlıkları değiştirilir. İleri beslemeli ağlar, en genel anlamıyla giriş uzayıyla çıkış uzayı arasında statik haritalama yapar. Bir andaki çıkış, sadece o andaki girişin bir fonksiyonudur (Haykin, 1994),(Yüçetürk, 2000).



Şekil 4.14 Geri yayılım MLP yapısı

4.8.1.1 Geri Yayılım Algoritması

MLP tipi YSA'ların eğitiminde oldukça yaygın olarak kullanılan Geri Yayılım algoritması, giriş katman saklı katman ve saklı katman-çıkış katmanları arasında mevcut olan ağırlıklı bağlantıların, ağ çıkışında oluşan hataya bağlı olarak optimize edilmesini amaçlar. MLP, nonlinear aktivasyon fonksiyonlarında lineer dönüşüm yapar. Giriş ve ilgili çıkış vektörü ağı eğitmek için kullanılmaktadır. Örüntüyü belirleyen giriş vektörü ağın ilk katmanına verildikten sonra aradaki gizli katmanlardan geçerek en sondaki çıkış katmanına ağırlıklar yardımı ile ulaşır. Ağdaki her bir nöron, kendisinde sonlanan ağırlık vektörlerinin aritmatiksel toplamını aldıktan sonra sonucu kendinden sonraki katmanın tüm nöronlarına kullanılan

aktivasyon fonksiyonuna bağılı olarak aktarılır. Geriye yayılım algoritması için ilk önce uygun bir ağ konfigürasyonu ile ağ başlatılır, ağa eğitim örnekleri sunulur, çıkış paterni üretilir, bu patern ile istenilen çıkış karşılaştırılır, aradaki fark hatadır, çıkış nöronlarına hata işaretini vererek bunun ağ boyunca geri yönde yayılarak çıkış ile hedef arasındaki hatayı azaltacak şekilde ağırlık ve eşik değerleri güncelleştirilir ve bu işlem ağ parametreleri kararlı hale gelinceye kadar ve ortalama karesel hata minimum olacak şekilde tekrarlanır. Katmanlar arasında sigmoid aktivasyon fonksiyonu kullanıldığı kabulü altında ağırlıkların yenilenmesi için kullanılan ifadeler şu şekilde elde edilebilir:

$$y_i^p = F(a_i^p) \quad (4.6)$$

$$a_i^p = \sum_j W_{ij} x_j^p + \theta_i \quad (4.7)$$

Çıkış birimlerindeki hatanın toplam ölçümü;

$$E = \frac{1}{2} \sum_p \sum_i (t_i^p - y_i^p)^2 \quad (4.8)$$

Hataların karesi alınarak beklenen değerden uzak olan çıkış değerlerinin toplam hatayı oluşturması sağlanmaktadır. Hatayı minimum yapmak amacı ile hatanın ağırlıklara olan bağımlılığı hesaplanır ve gradiente bağılı olarak ağırlıklar hatayı düşürecek şekilde yenilir.

$$\frac{\partial E^p}{\partial W_{ij}} = \frac{\partial E^p}{\partial a_i^p} \cdot \frac{\partial a_i^p}{\partial W_{ij}} \quad (4.9)$$

zincir kuralı kullanarak diferansiyel denklem çözümü aşağıdaki şekilde elde edilir.

$$\frac{\partial a_i^p}{\partial W_{ij}} = x_j^p \quad (4.10)$$

Hata işareti terimi;

$$\delta_i^p = -\frac{\partial E^p}{\partial a_i^p} = -\frac{\partial E^p}{\partial y_i^p} \cdot \frac{\partial y_i^p}{\partial a_i^p} \quad (4.11)$$

$$\frac{\partial y_i^p}{\partial a_i^p} = F'(a_i^p) \quad (4.12)$$

$$\frac{\partial E^p}{\partial y_i^p} = -(t_i^p - y_i^p) \quad (4.13)$$

$$\delta_i^p = (t_i^p - y_i^p) F'(a_i^p) \quad (4.14)$$

$$\Delta_p W_{ij} = \alpha \delta_i^p x_j^p \quad (4.15)$$

$$y_i^p = F(a_i^p) = \frac{1}{1 + e^{-\lambda a_i^p}} \Rightarrow F'(a_i^p) = y_i^p (1 - y_i^p) \quad (4.16)$$

Çıkış birimi için hata işareti

$$\delta_i^p = (t_i^p - y_i^p) y_i^p (1 - y_i^p) \quad (4.17)$$

Saklı birim için hata işareti

$$\delta_i^p = F'(a_i^p) \sum_h \delta_h^p W_{hi} = y_i^p (1 - y_i^p) \sum_h \delta_h^p W_{hi} \quad (4.18)$$

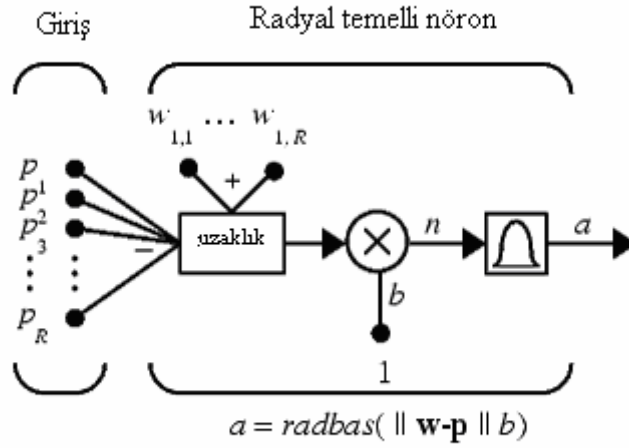
Elde edilen ifadelerde yapay sinir ağı eğitiminde data giriş seti ile ağırlıkların aktivasyona, aktivasyonun çıkışa, çıkışın da hataya etkileri görülmektedir (Altun, Eminoğlu ve Tezekici, 1999). Tez için oluşturulan MLP'nin özellikleri;

```
net = newff([minmax(data)],[5 4 1],{'logsig','logsig','purelin'},'trainlm');
```

```
net.trainParam.lr= 0.3; net.trainParam.mc= 0.2; net.trainParam.goal= 0.001;
```

4.8.2 Radyal Temelli Fonksiyon

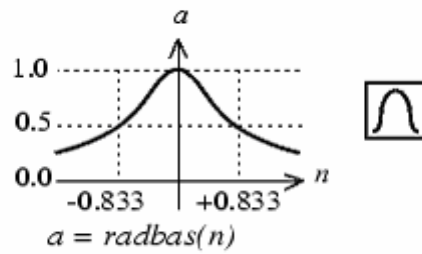
Radyal temelli ağ standart ileri-beslemeli geriye yayılım ağlarından daha fazla nöron gerektirebilir. Eğitim boyutu büyük olursa çok iyi sonuç verir [17].



Şekil 4.15 Radyal temelli nöron modeli [17]

$$\text{radbas}(n) = e^{-n^2} \quad (4.19)$$

Radyal Temelli ağlarda ağırlıklı girişlerin farkı alınıp, bir bias değeri ile çarpıldıktan sonra transfer fonksiyonundan (radbas) geçirildikten sonra çıkış elde edilir.

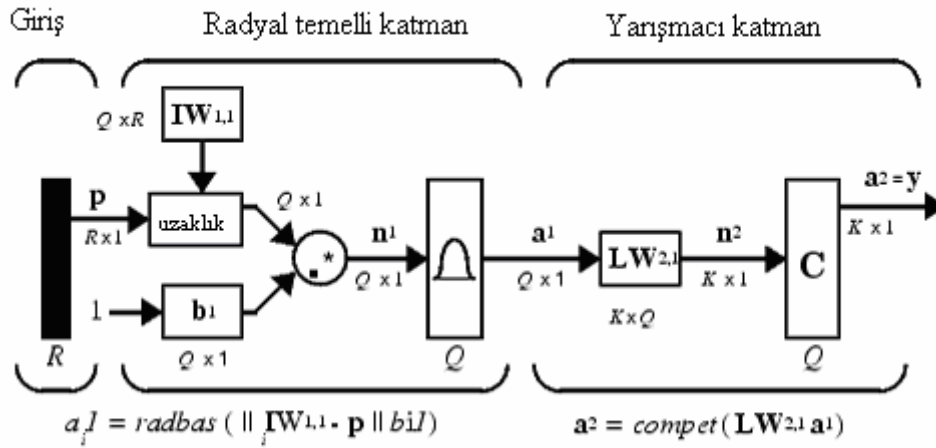


Radyal temelli fonksiyon

Şekil 4.16 Radyal temelli transfer fonksiyonu [17]

4.8.2.1 Olasılıklı Yapay Sinir Ağları (PNN: Probabilistic Neural Network):

PNN bir sınıflamadır. Eğiticiisizdir. 1960’larda geliştirilmiş olmasına rağmen geçen on yıla kadar pek sık kullanılmamıştır. Sınıflandırıcı, yapılacak işlemi kolay çözülebilecek birçok basit parçaya böler. Bundan sonra çok katmanlı bir ağ yapısında bu parçalar paralel olarak işlenir. Rastgele değişkenlerin olasılık yoğunluk fonksiyonunun tahmini PNN’in temelini oluşturur (Specht, 1990). Yapay sinir ağları içerisinde sınıflandırma problemine en uygun olanı olasılıklı yapay sinir ağlarıdır. PNN gerek çalışma hızı, gerekse doğru sınıflandırma oranı yönüyle birçok çalışmada tercih sebebi olmuştur (Specht, 1999).

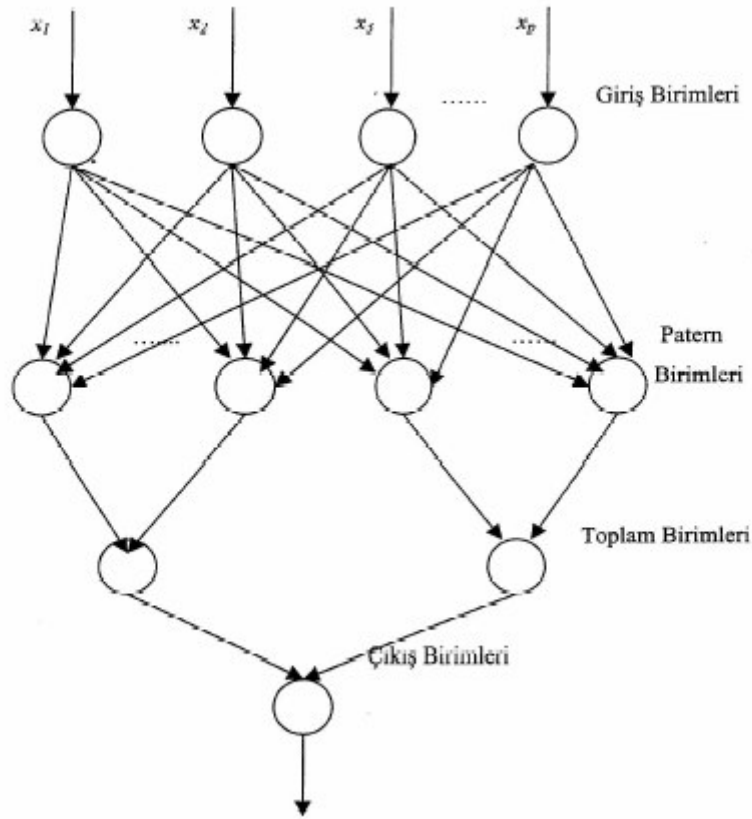


Şekil 4.17a PNN matlab yapısı [17]

Q : 1. katmanın giriş-hedef çiftleri (1.katmandaki nöron sayısı)

K : Giriş datalarının sınıf sayısı (2. katmandaki nöron sayısı)

R: Giriş vektörlerindeki eleman sayısı



Şekil 4.17b PNN yapısı

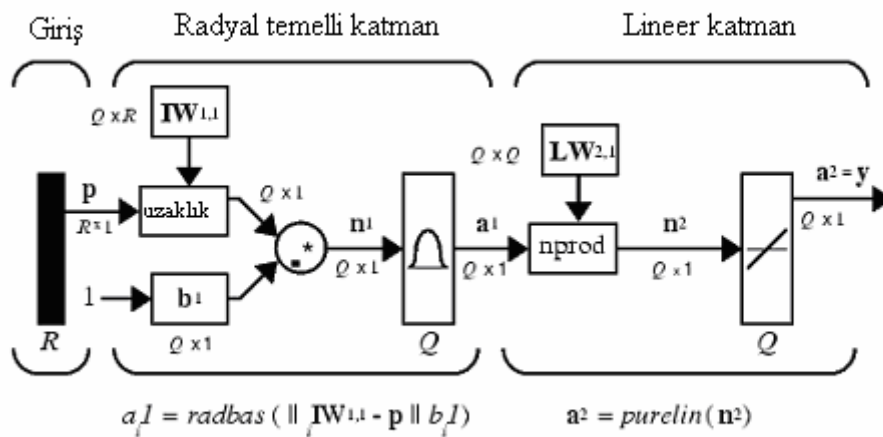
PNN’ de birinci katman, giriş vektörü ile eğitim giriş vektörleri arasındaki mesafeyi hesaplar ve bir vektör üretir. Bu vektör giriş vektörünün eğitim giriş vektörüne ne kadar yakın olduğunu gösterir. İkinci katman ise sınıfları toplama birimidir. Sonuçta bir yarışmalı transfer fonksiyonu maksimum çıkışı belirler.

Parametrik olmayan yöntemlerle olasılık yoğunluk tahminine dayanan sınıflandırma istemleri ile ileri beslemeli yapay sinir ağları arasında çok büyük benzerlikler gözlenmiştir. Şekil 4.17b’ de X giriş örneğinin iki farklı sınıftan birine atamakta kullanılan bir yapay sinir ağı gösterilmiştir. Şekil 4.17b’ deki giriş birimleri sadece giriş değerlerini tüm örüntü birimlerine dağıtmakta kullanılmaktadır. Her bir örüntü birimi giriş vektörü ile ağırlık vektörünün iç çarpımını hesaplamakta ve toplam çıkış üzerinde doğrusal olmayan bir aktivasyon hesaplamaktadır. Toplama birimleri örüntü birimlerinden gelen girişlerin toplanması işlemini gerçekleştirir. Çıkış ya da karar birimleri iki girişli nöronlardır. Bu birimler ikili çıkış üretir ve sadece bir ağırlığı vardır. Ağı eğitilmesi için her örüntü biriminin ağırlık vektörü (W) öğrenme kümesindeki bir X vektörüne atanır ve örüntü biriminin çıkışı

ilgili toplam birimine bağlanır. Her bir öğrenme vektörü için ayrı bir nörona ihtiyaç vardır. PNN' nin son katmanında ise test sonucunda en büyük değere sahip nöron kazanan nöron olarak adlandırılır ve test sonucu belirlenir. (Masters, 1993).

4.8.2.2 Genelleştirilmiş Regresyon Yapay Sinir Ağları (GRNN: Generalized Regression Neural Network)

Yeni geliştirilen algoritmalarından biridir, genelde fonksiyon yaklaştırma için kullanılır. MLP gibi iteratif bir eğitim gerektirmez. Bu sebeple çok hızlı sonuca ulaşabilir. Eğitim datasına bağlı olarak girişle çıkış arasında bir fonksiyon kullanılır. Eğitim datasının boyutu arttıkça hata sıfıra doğru azalma gösterir. GRNN, standart tanıma tekniklerinde olduğu gibi sürekli değişken verilerin tahmini için ya da sınıflandırma işlemleri için kullanılabilir (Hagan, Demuth, ve Beale, 1996). GRNN'de sık sık yaklaşıklık fonksiyonu kullanılır. GRNN iki katmanlıdır. Birinci katmanda radyal temelli nöronlar yer alır. Birinci katmanda giriş vektörü ile ağırlıklar arasında bir öklit mesafesi hesaplanır. Radyal katman bilinen eğitim datalarına göre kümeleme yapar. Tanıma katmanı ise lineerdir. GRNN hafıza temelli ileri beslemeli bir ağ yapısıdır. GRNN'de bağımlı bir y değişkeni ile bağımsız bir x değişkeni ve eğitim setiyle y değeri tahmin edilir. Tanıma değeri tahmin edilen y değerini üretir ve hatayı minimize eder. GRNN, verilen bir eğitim setinden x ve y değerlerinin olasılık yoğunluk fonksiyonunu tahmin eder. (Specht, 1991).



Şekil 4.18 GRNN yapısı [17]

Q : 1. ve 2. katmandaki nöron sayıları ve giriş-hedef çiftlerinin sayısı

R : Giriş vektörlerinin eleman sayısı

nprod ise normalize edilmiş matrissel çarpma işlemi yapar.

Örneğin

LW= [1 -2;3 4;5 6];

a = [7; -8]; ise

aout = normprod(LW,a)

aout =

-23

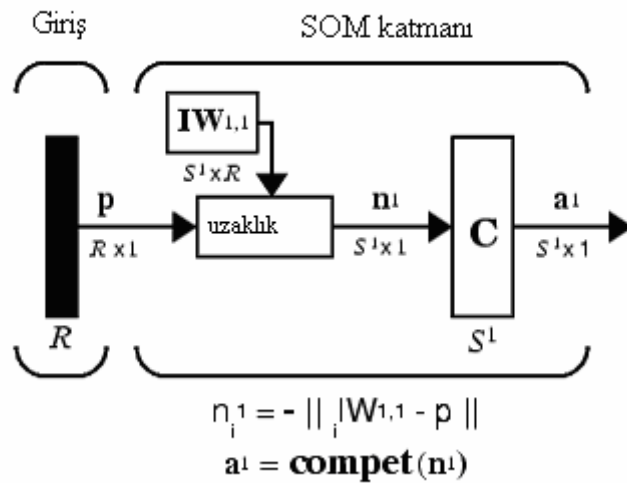
11

13

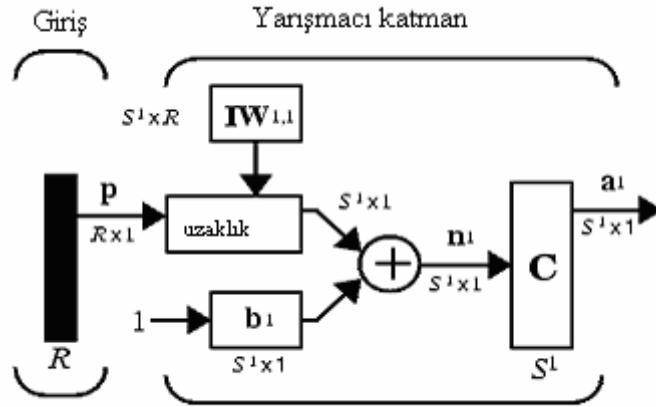
GRNN' de birinci katmanda giriş vektörü ile ağırlık vektörü arasında mesafe ölçümü yapılır.

İkinci katmanda ise nprod işlemi ve parçalı lineer fonksiyonu işlemi yapılır.

4.8.3 Kendi Kendine Organize Olan Haritalar (SOMs: Self Organizing Maps)



Şekil 4.19 SOM yapısı [17]

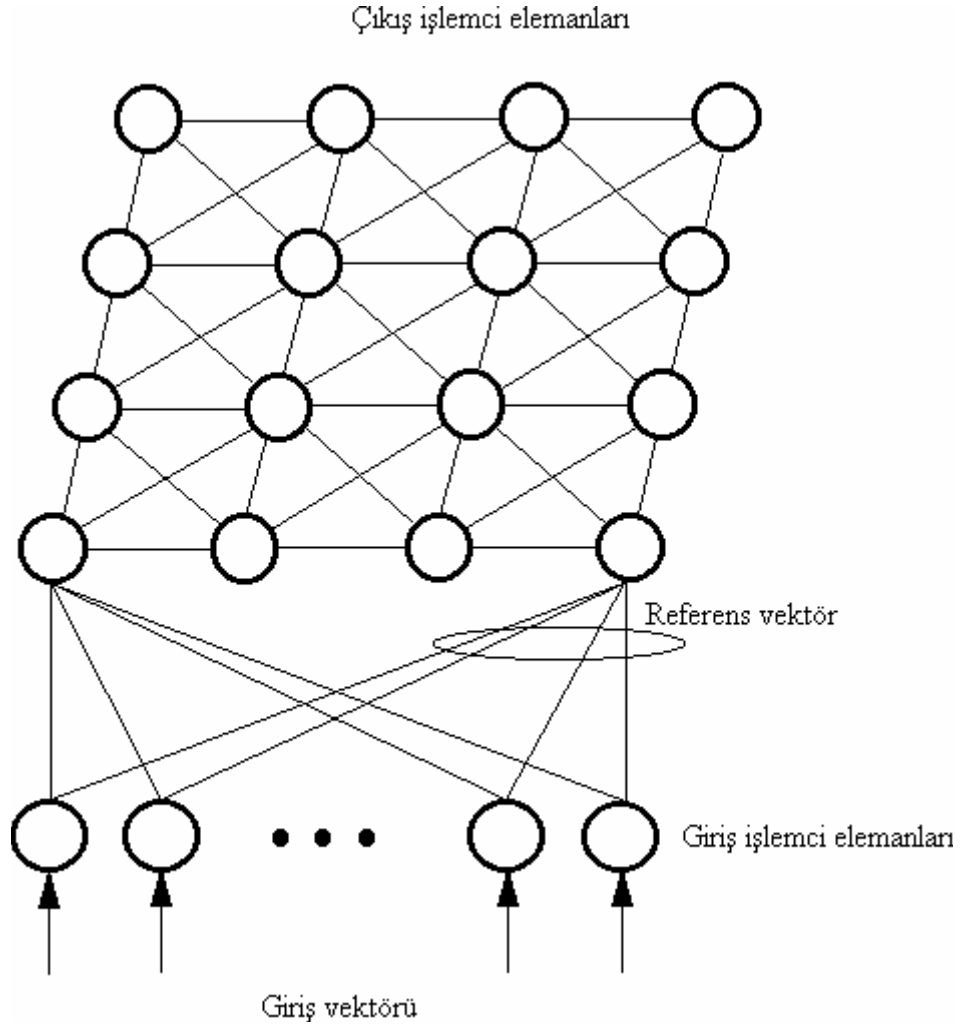


Şekil 4.20 Yarışmalı katman [17]

Beyindeki hücrelerin yerleşimi belli bir dizilişe göredir ve bu diziliş alınan dış uyarının fiziksel özelliğine bağlıdır. Örneğin beynin ses ile ilgili bölgesinde hücrelerin dizilişi ses dizgesinin sıklık yanıtını yansıtır. Alçak sıklıktaki işaretler bölgenin bir ucunda, yüksek sıklıktakiler öteki ucunda etkiye yol açarlar. Eğitim algoritması ile n -boyutlu uzaydaki sınıf dağılımı iki-boyutlu uzaya taşınmaktadır. Eğitim tamamlanıp düğüm ağırlıkları uzayda belirli pozisyonlara yerleştikten sonra bir etiketleme yöntemi kullanılarak düğümlerin sınıfları belirlenir.

SOM, ağırlık girişleri ile onların gelecekteki cevapları arasında korelasyon kurabilir. Yarışmalı nöron sayesinde girişleri benzer gruplarda kümeler. SOM, giriş uzayında birbirine yakın olan benzer verileri gruplar ve iki boyutlu haritada yakın birimlere atar. Böyle bir algoritmanın özellikleri şunlardır :

- (1) Bir ya da iki boyutlu kafes yapısındaki nöronlar, girişten alınan bilgilerin basit fonksiyonlarını hesaplar.
- (2) Bu fonksiyonları karşılaştırılır ve en büyük değerli fonksiyonuna sahip olan nöron seçilir.
- (3) Seçilen nöronlar ve komşuları aktive edilir.
- (4) Aktive olan nöronlar giriş işaretine göre fonksiyonlarını arttırırlar (Kohonen, 1997).



Şekil 4.21 SOM mimarisi

Ağın içindeki düğümlerin yapısı aşağıdaki denklemlerle ifade edilir.

$$D_j = \sqrt{\left(\sum_{i=1}^n (x_i - w_{ji}(k))^2\right)} \quad (4.20)$$

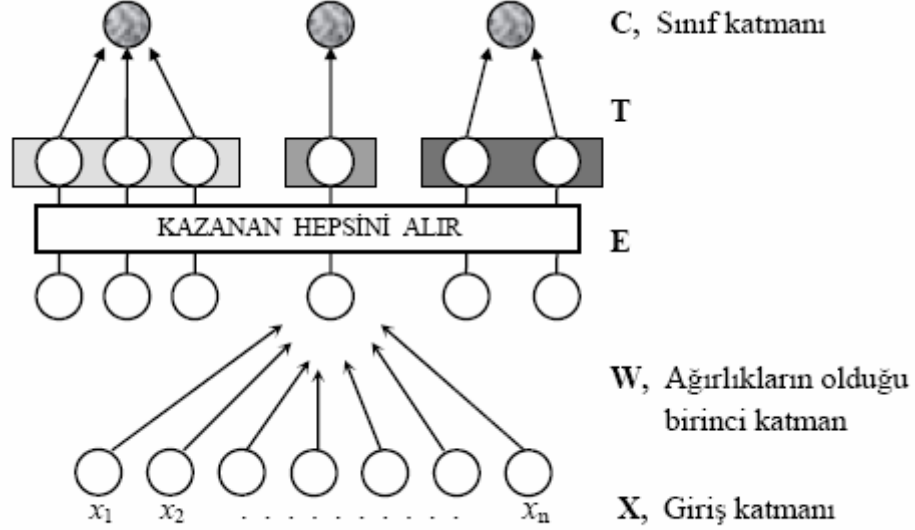
$$E_e = \begin{cases} 1, \dots, D_e = \min\{D_j\} \\ 0, \dots, \text{aksi_takdirde} \end{cases} \quad (4.21)$$

$$\text{Çıkış katmanı } C_c = \sum_e E_e \cdot T_{ec} \quad (4.22)$$

E_e , birinci katmandaki düğümlerin çıkışını, T_{ec} ise sadece 0 veya 1 değerini alan OR'lama işlemini gerçekleyen ikinci katmana ait bağlantı katsayılarını göstermektedir.

İlk katman, düğüm ağırlıkları ile giriş vektörü arasında minimum mesafeyi bulmada

kullanılırken, ikinci katman, ağdaki düğümlerin ait oldukları sınıfı tanımlamak için kullanılır. İkinci katmandaki ağırlıklar eğitim tamamlandıktan sonra bir etiketleme işlemi ile belirlenir. İkinci katman, aynı sınıfa ait çıkış düğümlerini lojik olarak OR'lamak için kullanılır.



Şekil 4.22 SOM sınıflaması

Düğümlemler komşularına meksika şapkası şeklinde bir bağlantı biçimi ile bağlıdır. Bu bağlantı sayesinde yarışmalı öğrenmede bir düğümün kazanması temin edilir. Kohonen ağıının eğitiminde sadece kazanan düğümün ağırlığı değil, o düğümün çevresinde seçilen belirli bir komşuluk içindeki düğümlemlerin ağırlıkları da değiştirilir. Aşağıda Kohonen ağıının eğitim algoritması verilmiştir. Eğitim başlatılmadan önce, çıkış düğüm sayısı, harita boyutu ve iterasyon sayısı belirlenir. Öğrenme oranı zamanla azalır. Ağıın tüm düğümlemlerinin ağırlıklarına 0 ile 1 arasında rastgele değerler veya eğitim setinden rastgele değerler verilir.

Adım 1: Eğitim kümesinden rastgele bir vektörü ağa giriş olarak ver.

Adım 2: Giriş vektörü ile ağıın düğümlemleri arasındaki mesafeleri aşağıdaki ifadeyi

kullanarak hesapla.

$$D_j = \sqrt{\left(\sum_{i=1}^n (x_i - w_{ji}(k))^2\right)} \quad (4.23)$$

x_i giriş vektörünü, w_{ji} ise j -inci çıkış düğümünün ağırlık vektörünü temsil etmektedir.

Adım 3: Minimum uzaklığa sahip J^* çıkış düğümünü bul.

Adım 4: J^* çıkış düğümünün ve komşularının ağırlıklarını aşağıdaki ifadeleri kullanarak değiştir.

$$w_{ji}(k+1) = w_{ji}(k) + \eta(t)(x_i - w_{ji}(k)) \quad (4.24)$$

Kazanç terimi olarak ifade edilen $\eta(t)$, 0 ile 1 arasındadır ve zamanla azalır. k , iterasyon değişkenidir.

Adım 5: İterasyon sayısını azalt. İterasyon sayısı sıfıra eşit değil ise *adım 1*'e git. Aksi takdirde eğitim algoritmasını durdur.

Eğitim tamamlandıktan sonra çıkış düğümlerine sınıf etiketleri vermek için her çıkış düğümüne sınıf sayısı kadar bellek (sınıf etiketi gözü) atanır. Eğitim kümesi içindeki herbir vektörün hangi çıkış düğümüne yakın olduğu araştırılır. İncelen giriş vektörüne en yakın ağ düğümünün sınıf etiket gözü, giriş vektörünün sınıfı ile ilişkili olarak artırılır. Eğitim kümesindeki tüm vektörler için bu işlem tekrarlanır. Eğitim tamamlandıktan sonra her düğüm için en yüksek değere sahip etiket gözü belirlenir. Çıkış düğümünün sınıfı bu gözün etiketi ile belirlenir (Wu ve McLarty, 2000) (Kangas, Kohonen ve Laaksonen, 1990) (Kohonen, 1997).

4.8.4 Kendi Kendine Organize Olan Global Sıralamalı Sınıflama Algoritması (SOGR: Self Organizing Global Ranking Algorithm)

Yeni sınıflama algoritmasıdır. SOGR, yalnızca vektörler arasında mesafe ölçümü gerektirir, örneğin; öklit mesafesi gibi. Başlangıçta SOGR ağırlık vektörü rastgele üretilir. Eğitim esnasında, onlar sınıfları sunmak için uyarlanırlar. SOGR, algoritmanın benzer grupları içinde kümeleme içerir, SOM gibi. Komşuluk kavramı, SOM'dan tamamen farklıdır. SOGR'da bir komşu ağırlık vektörü geometriksel olarak şimdiki (geçerli) ağırlık vektörüne yakın olmak zorunda değildir. Yani, komşu ağırlık vektörleri genele dağıtılmıştır. Denemeler gösterir ki; SOGR sonuçları çok iyidir. Aynı zamanda SOGR basit ve hızlıdır. SOM sık sık 2-D ızgara ile sunulur. Eğitim sırasında, kazanan düğümün ağırlığı ve onun genel komşu düğümlerinin ağırlıkları güncelleştirilir. SOM'da datanın topolojisi korunabilir. Çünkü benzer ağırlık vektörleri lokal olarak dağıtılmıştır. SOM datanın topolojisini kaydedebilir, fakat o gerekli olmayan ağırlık vektörlerini kullanabilir. Bu yüzden, eğitim ve test zamanı artabilir. Bu da optimizasyon problemi doğurur. Önerilen topoloji, genel komşuluğu sıralama yaparak yeni bir kavram üretir. Önemli avantajı, düşük hesaplama yüküdür. SOGR, basit bir tekrarlayan

algoritmadır. SOM'la ağırlık vektörlerinin güncellenmesi benzerdir, fakat komşuluk tanımı SOM'dan tamamen farklıdır. SOGR'da komşu ağırlık vektörleri genel temelli, benzerlik sıralaması dikkate alınır ve bu yüzden onlar her yerde vektörlerin uzayı ile ilişkilidir. SOM'da komşuluk geometriktir ve çok hesaplama gerektirir. Bu SOGR'un neden daha basit ve daha hızlı olduğunu açıklamaktadır. İterasyon sırasında, komşuluk ağırlık vektörleri, global benzerlik sıralamasına göre seçilir. Bu yüzden, bir uzaklık ölçümü gereklidir. Bu eğitim giriş dataları ve düğümlerin ağırlık vektörleri arasında bir öklit mesafe ölçümü olabilir. Başlangıçtaki ağırlık vektörlerini giriş datalarından rastgele seçilir. Bir iterasyon sırasında bir eğitim giriş vektörü rastgele seçilir ve ondan sonra bütün düğümlerin ağırlık vektörleri ile bu eğitim giriş vektörü arasındaki uzaklık hesaplanır. R; komşuluk boyutu(harita boyutu) diye farzedilirse globalde en küçük mesafedeki düğümün ağırlık vektörü kazanan seçilir. Sonra globalde tekrar kazanan düğümün ağırlık vektörleri güncelleştirilir.

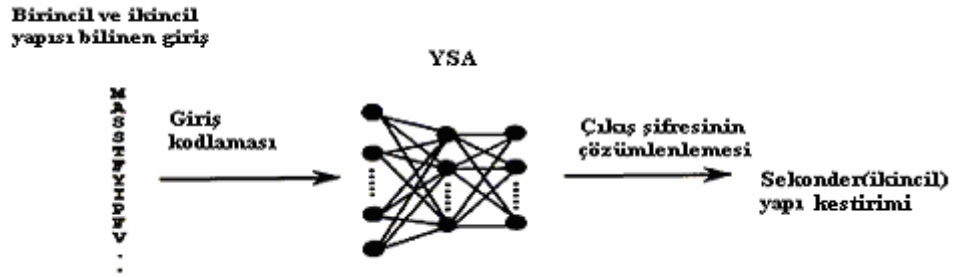
$$m_j(t+1) = \begin{cases} m_j(t) + \beta[x(t) - m_j(t)] & j \in R \text{ ve } j \text{ kazanan düğüm ise} \\ m_j(t) & \text{değilse} \end{cases} \quad (4.25)$$

β_i ; öğrenme oranı, zamanla azalan şekilde seçilir. Onu her komşuluk boyutu(R) için farklıca seçmek uygundur. Büyük R için, öğrenme oranı biraz küçük seçilmelidir. Bu değişiklikler global minimum için ve lokal minimumdan sakınmak için yardımcı olur. SOGR'u eğiticili öğrenmede kullanmak için ağırlık vektörlerini uygun sınıflarla etiketlemek gereklidir. Etiketleme ve sınıflama SOM'a benzerdir. Ağırlık vektörlerinin güncellemesi bitirildikten sonra, her eğitim vektörü ve bütün düğümlerin ağırlık vektörleri arasındaki uzaklıklar hesaplanır. Sonra en yakın ağırlık vektörü "kazanan" olur ve o sınıfla ilgili sayıcı bir artar. Bu işlem bütün eğitim vektörleri için tekrarlanır. Bu işlemden sonra sayıcıdaki değerlere göre, yani sayıcıdaki en fazla değerlere göre etiketleme olur ve o sınıf çıkış için kazanır. Sonra eğitim biter. Test işlemi de eğitim işlemi gibidir. İlk önce birinci test vektörü ile başlanır. Bu vektör ile bütün düğümlerin ağırlık vektörleri arasındaki uzaklık hesaplanır. Eğitimde olduğu gibi en yakın mesafe için en yüksek sayacın sahip olduğu sınıf kazanır ve o sınıfla etiketlenir. Bu işlem bütün test pikselleri için yapılır (Ersoy, Erer ve Sağlam, 2004).

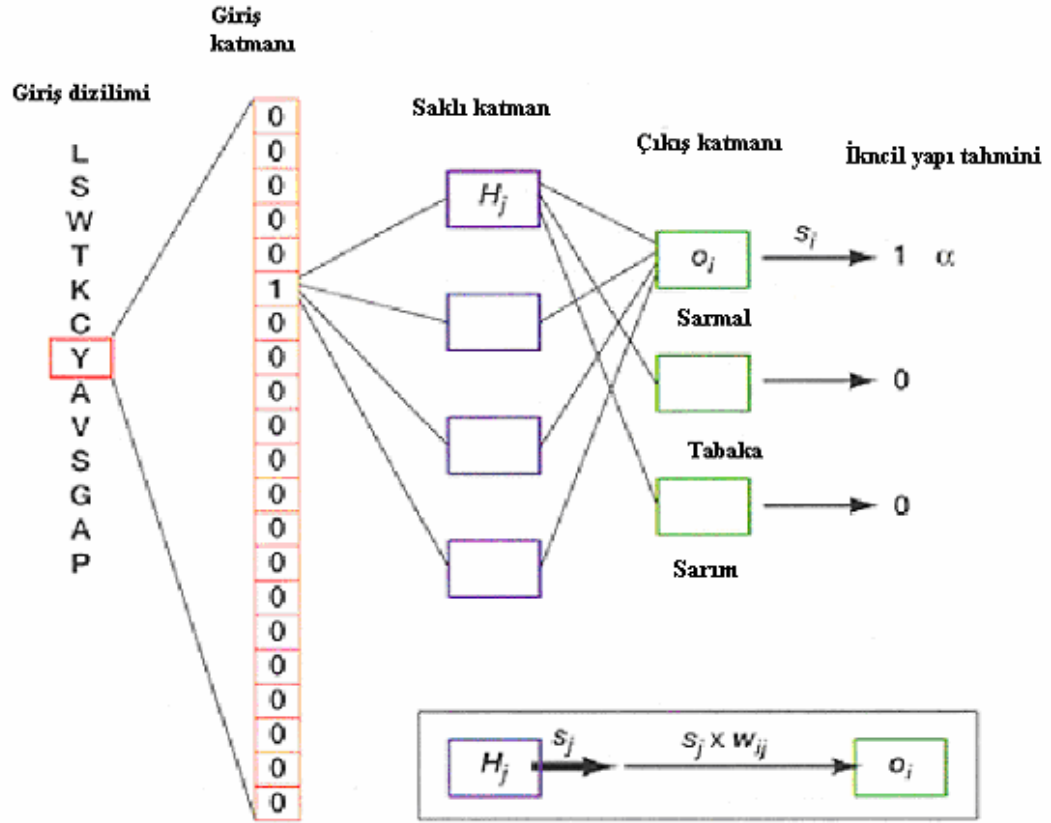
5. YSA ile PROTEİNLERİN İKİNCİL YAPILARININ KESTİRİMİ

5.1 Özellik Sunma

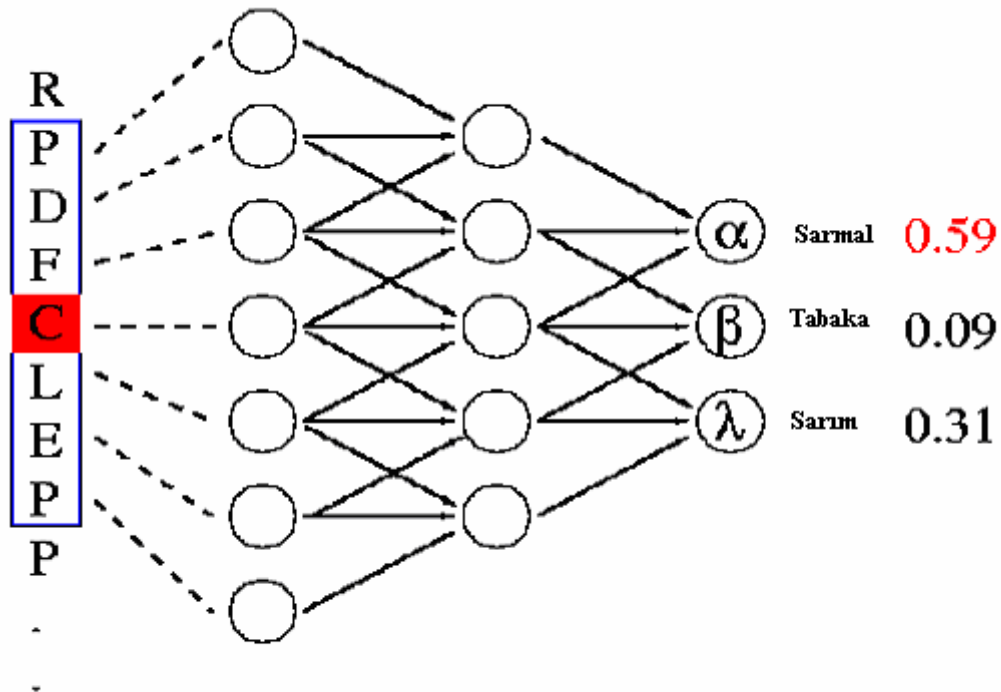
Birinci adım problem tanımına uygun bir yapay sinir ağı geliştirebilmek, giriş-çıkış eşlemelerine (kodlamalarına) ve YSA modeline karar vermek gerekir. Ayrıca data önışlem ve data son işlem kararı gerekir. Data önışlemi herhangi bir YSA'nın önemli bir parçasıdır. Ön işlem hem özellik sunma (bilgi seçip çıkarma) ve giriş kodlamadır. Uygun özellik seçimi ve kodlama metodu seçimi, son sistemin kararı için çok önemlidir. Çünkü YSA'ya hangi bilgi sunulacağına o karar verir. Özellik sunma, moleküler dizi analiz için iki sebepten dolayı çok önemlidir. Birincisi; moleküler yapı ve fonksiyon hakkında bilgi doğruluğu için; ikincisi ise verilen yapının karakteristiğinin öğrenilmesi için. Giriş kodlaması için genellikle BIN-20 metodu kullanılır. Çıkış ise iki veya üç durumludur (Mount, 2001), (Wu ve McLarty, 2000).



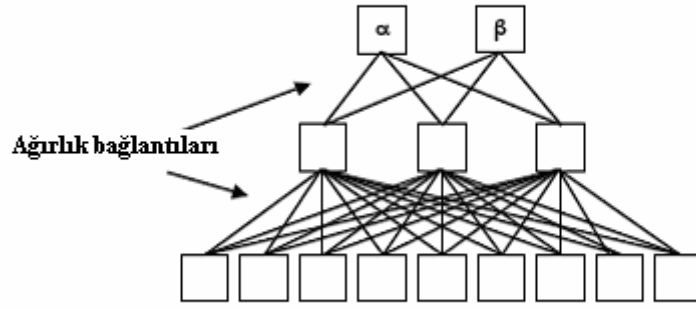
Şekil 5.1a Proteinlerin ikincil yapılarının YSA ile kestirimi



Şekil 5.1b Proteinlerin ikincil yapılarının YSA ile kestirimi



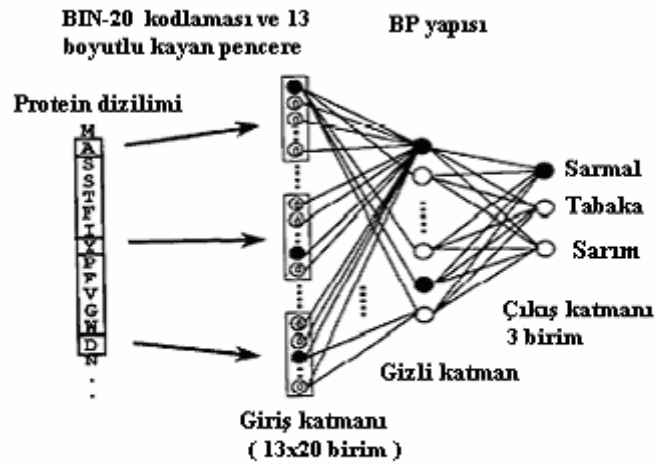
Şekil 5.1c Proteinlerin ikincil yapılarının YSA ile kestirimi



Şekil 5.2 Dokuz kayan pencereci iki çıkışlı örnek bir YSA yapısı

5.2 Proteinlerin İkincil Yapıları Kestiriminin Önemi

Giriş katmanı dizilimdir ve giriş katmanı binaridir. Bu yüzden her amino asit rezidüsü için çoklu girişler olmalıdır. Eğitim ve test seti bağımsız olmalıdır [10].



Şekil 5.3 On üç kayan pencereci üç çıkışlı örnek bir YSA yapısı

Proteinin fonksiyonel, yani özelliğinin ve görevinin anlaşılabilmesi için mutlaka proteinin ikincil yapısının bilinmesi gereklidir. Yapılan tahmin çalışmalarında doğruluk oranı %45 - %70 arasındadır [11][12][13]. Protein ikincil yapı tahmininde Sarmal'ın (heliks) doğruluk oranı Tabaka'dan (sheet) daha yüksektir. Ayrıca en yüksek doğruluk oranları BIN-20 de çıkmıştır [11]. Protein ikincil yapısının deneylerle bulunması çok pahalı, çok zor ve çok fazla zaman almaktadır. Proteinin ikincil yapısının bilinmesi ile proteinin vücut içindeki fonksiyonu anlaşılabilmekte ve böylece insanın şifresi çözülebilmektedir. Proteinin birincil ve

ikincil yapılarına PDB (Protein Data Bank)(www.rcsb.org/pdb/) sayfasından ulaşılabilir. (Tang J., 1995), (Holley L.H. ve Karplus M., 1989), (Qian ve Sejnowski, 1988), (Wu ve McLarty, 2000), [1],[6],[12].

5.3 Proteinlerin İkincil Yapıları Kestirimine Genel Bir Bakış

Giriş için kodlama işlemi yapılır. Yirmi tane amino asit BIN-20 metoduna göre kodlanır (BIN-20: Yani bir amino asit için bir satırda bir tane bir on dokuz tane sıfır olması). Merkez rezidüye göre tahmin yapılır. Eğitim boyunca ağırlıklar güncellenir. Dezavantajı ise çok fazla giriş gerektirmesidir. Amino asit dizilimi (protein birincil yapısı ve birincil yapının merkez rezidüye ikincil yapısı) pencerelere ayrılır. Genellikle yarışmalı öğrenme vardır ve kazanan hepsini alır prensibi kullanılır (Şekil 5.1a, şekil 5.1b ve şekil 5.1c) [11][12][13](Rost ve Sander, 1994).

5.4 Yapılan Çalışmanın Ayrıntıları

5.4.1 Data Oluşturma

Datalar; “ Fifth Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction Asilomar Conference Center, 2003 ” konferansında sunulan 15 yeni proteinden ve ayrıca Hemoglobin-141 ve Hemoglobin-146 proteinlerinden oluşturulmuştur. Bu konferansta sadece protein isimleri sunulmuştur. Dataların harfsel bilgileri PDB internet sayfasından temin edilerek Matlab 7.0’ da yazılan bir programla YSA için uygun sayısal datalar oluşturulmuştur.

5.4.2 Dataların Özellikleri

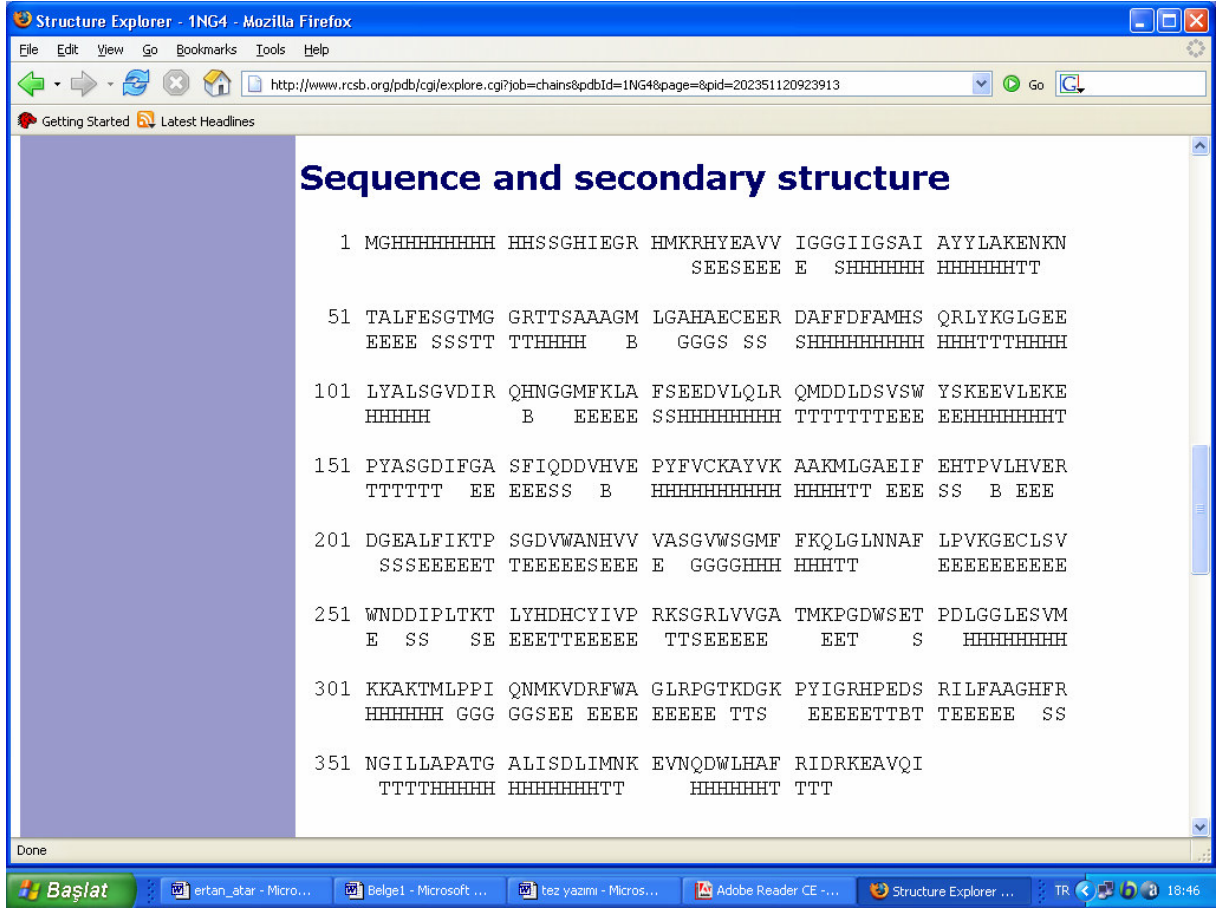
2003 yılındaki Protein Yapı Tahmini Konferansı’nda sunulan bu 15 yeni proteinin özellikleri;

Çizelge 5.1 Tezde kullanılan proteinlerin özellikleri

2003 yılındaki Protein Yapı Tahmini Konferansı’nda sunulan 15 yeni protein				
PDB için Protein ismi	Rezidü	Sarım (Coil)	Tabaka (Sheet)	Sarmal (Helix)
1ON9	523	202	87	234
1O14	511	317	78	116
1O0W	252	77	22	153
1O0U	429	199	68	162
1NYN	131	82	16	33
1NG4	390	181	93	116
1MG7	417	203	89	125
1H7M	102	40	18	44
154L	185	80	4	101
1UZC	71	29	0	42
135L	129	68	8	53
1AAB	83	44	0	39
1AAF	55	49	0	6
1KOY	62	22	0	40
1IYR	111	66	0	45
Rezidü _{Toplam} (Sheet dahil) : 3451		C _T :1659	S _T :483	H _T :1309
Rezidü _{Toplam} (Sheet hariç) : 2968				

Hemoglobin en önemli proteinlerden biridir. Birçok hastalığın belirtileri hemoglobinde belli olur. Oksijen, hemoglobin vasıtasıyla taşınır. Rezidü sayılarına göre iki tip hemoglobin vardır. Hemoglobin-141 proteini 2562 rezidü olarak 1890 sarmal ve 672 sarım yapısından oluşmuştur. Hemoglobin-146 proteini ise 2540 rezidü olarak 1995 sarmal ve 545 sarım sınıfından oluşmuştur. Hemoglobin proteininde tabaka yapısı yoktur (Pauling, 1952), (Rinsho, 1998).

Şekil 5.4 de PDB sayfasından alınmış 1NG4 proteinin birincil ve ikincil yapısı mevcuttur. Birinci satırdakiler proteinin birincil yapılarını, ikinci satırdakiler ise proteinin ikincil yapılarını göstermektedir. İkincil yapıdaki boş olanlarda sarım(coil) sınıfına ait olmaktadır. Bu bilgiler çizelge 5.2 ye göre üç sınıflı yapıya dönüştürülmüştür. Matlab'ta yazılan başka bir program ile çizelge 5.1 deki bilgiler elde edilmiştir. Doğal olarak her protein için datalar değişeceğinden tekrardan program çalıştırılmıştır. PDB sayfasından alınan veriler Matlab'ta yazılan bir program ile BIN-20 giriş kodlamasına göre istenilen pencere boyunda datalar oluşturulmuştur. Datalar Matlab 7.0'daki dizi editörü sayfasından kopyalanıp, bir .txt dosyasına kayıt edilmiştir. Her protein için ve her pencere boyutu için program tekrar çalıştırılmıştır. Böylece 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan bu 15 yeni protein için tabaka(sheet) yapısı çıkarılmış ve tabaka(sheet) yapısı çıkarılmamış olarak 11-15-19 pencere boyutuyla toplam 6 farklı data oluşturulmuştur. Bu 15 yeni proteinde görüldüğü gibi tabaka sınıfı diğer sınıflara göre çok az vardır. Bu yüzden de data çeşitliliği olması bakımından bu 15 protein için tabaka sınıfı çıkartılmış datalar da oluşturulmuştur. Ayrıca Hemoglobin-141 ve Hemoglobin-146 dataları için de yine BIN-20 yöntemine göre 11-13-15-17-19-21-23 pencere boyutlu $7+7=14$ farklı data, toplamda ise $14+6=20$ farklı data(veri) oluşturulmuştur. Yapay sinir ağları olarak ta 5 ayrı tip YSA kullanılmıştır. Bunlar: SOM, MLP, PNN, GRNN ve 2004 yılının sonunda geliştirilen yepyeni bir algoritma olan SOGR'dur.



Şekil 5.4 Proteinlerin yapılarının elde edilmesi [14]

Çizelge 5.2 Proteinlerin ikincil yapılarının sınıflaması

İkinci yapı kısaltması	İkincil yapı genel sınıfı
H	H
G	H
I	H
E	E
B	C
T	C
S	C
	C

5.4.3 Giriş ve Çıkış Kodlaması

YSA'lara nümerik (sayısal) format uygulanması gereklidir. Giriş katmanı proteinin birincil yapısı olduğundan; 20 amino asittir. Çıkış katmanında ikincil yapı ise sarmal (heliks), tabaka (sheet) ve sarım (coil) şeklinde üç sınıfa ayrılır. Veya bazı durumlarda tabaka yapısı proteinlerde az olduğu için tabaka yapısı çıkarılıp iki sınıfa ayrılır; sarmal ve sarım şeklinde olabilir.

Çizelge 5.3 Giriş katmanı için 20 amino asitin BIN-20 yöntemi ile sayısallaştırılması

Alanine (A):	1000000000000000000000
Cysteine(C) :	0100000000000000000000
Aspartic Acid(D):	0010000000000000000000
Glutamic Acid(E):	0001000000000000000000
Phenylalanine(F):	0000100000000000000000
Glycine(G):	0000010000000000000000
Histidine(H):	0000001000000000000000
Isoleucine(I):	0000000100000000000000
Lysine(K):	0000000010000000000000
Leucine(L):	0000000001000000000000
Methionine(M):	0000000000100000000000
Asparagine(N):	0000000000001000000000
Proline(P):	0000000000000100000000
Glutamine(Q):	0000000000000010000000
Arginine(R):	0000000000000001000000
Serine(S):	0000000000000000100000
Threonine(T):	0000000000000000010000
Valine(V):	0000000000000000000100
Tryptophan(W):	00000000000000000000010
Tyrosine(Y):	00000000000000000000001

Çizelge 5.4 Üç durumlu çıkış katmanının sayısallaştırılması

Helix (Sarmal)	(H):	1
Coil (Sarım)	(C):	2
Sheet (Tabaka)	(E):	3

Çizelge 5.5 İki durumlu çıkış katmanının sayısallaştırılması

Helix (Sarmal)	(H):	1
Coil (Sarım)	(C):	2

5.4.4 YSA ile Proteinlerin İkincil Yapılarının Kestirimi

İlk önce bir YSA programı oluşturulmuştur. Eğitim bilinen birincil ve ikincil yapı dataları ile yapılmıştır. Test ise bilinen başka bir birincil ve ikincil yapı dataları ile yapılmıştır. Giriş şifrelemesi için BIN-20 kullanılmıştır. BIN-20 'de sadece bir tane bir ve diğerlerinde sıfırlar (19 tane sıfır) bulunur. Böylece 20 tane amino asit şifrelenmiş olur. Pencere boyutu ve şifreleme ile girişler karışmaz, ama giriş sayısı artar. Ağırlıklar ilk önce rastgeledir, daha sonra ağırlıklar doğru cevaba göre uyarlanmıştır. Eğitim bittiğinde ağırlıklar kesin değerini almışlardır. Pencerede merkez rezidü fokuslanır. Merkez rezidünün karşısındaki ikincil yapı o pencere için hedeflenen çıkış sayılmıştır. YSA öğrenir, (öğrenirken ağırlıklar değişir) sonra pencere sağa kayar. Tahminler yapılırken, genellikle kazanan hepsini alır prensibine göre, çıkış en güçlü çıkışın cevabını verir.

Örnek bir hesaplamayla;

5 gizli düğümlü, 17 boyutlu pencere , 3 çıkışlı, tek gizli katmanlı bir YSA için;

17×20 (BIN-20 kodlama) = 340 giriş düğümü $((17 \times 20) \times 5) + (5 \times 3) = 1715$ ağırlık gerekir.

Proteinlerin ikincil yapılarını bulmak için belli sayıda amino asiti içeren kayan pencereler kullanılmıştır. Bu pencerelerin tam ortalarında bulunan *i.* amino asitin karşında bulunan ikincil yapısı R_i , tahmin edilmeye çalışılmıştır. Oluşturulan böyle bir yapıda, merkezde bulunan amino asidin $\pm n$ sayısı kadar sağında ve solunda da amino asit bulunması gerekmektedir. Pencerenin genel şekli de; *Pencere genişliği* (P) = $2 \times n + 1$ şeklinde olmaktadır.

Örneğin;.....ACDWGV.....şeklinde giden bir birincil yapı ve aynı proteininHHCSEE.....şeklinde olan ikincil yapısı olsun. Böyle bir veri 5 amino asit içeren bir pencere ile tarandığı düşünülürse; bu veri şöyle hazırlanır; birinci pencere ACDWG ve ikincil yapısı ; C (merkez rezidüye göre tahmin yapıldığı için); ikinci pencere (birinci pencere bir birim sağa kayar) CDWGV ve ikincil yapısı S olarak tanımlanır. Böylece harflerin yerine önceki sayfada olduğu gibi 20-kodlama ile 0 ve 1 den oluşan bilgiler ve çıkış için de C,E ve H harflerini tanımlayacak sayısal değerler konularak çıkış katmanı da tanımlanmış olur. O zaman giriş katmanında 20x5 tane giriş, çıkış katmanında ise üç tane veri(1,2,3) için bir bitlik çıkış bulunur.

Kayan pencere, protein üzerinde belli amino asit sayısında birer birer hareket ederek bütün protein üzerinde hareket eder. Yani 5 amino asidi pencere boyutu olarak alırsak, bu proteinin sonuna kadar birer amino asit kaydırılarak sürülür(örnekteki gibi). Böylece proteinin bütünü taranmış olunur. Ayrıca ikincil yapısı belli olmayanlar sarım (coil) sınıfına girerler.

YSA ile protein ikincil yapısının kestirimi için ; SOGR, SOM, GRNN, PNN ve MLP algoritmaları kullanılmıştır. Bu programlardan SOGR ve SOM programları için Matlab NNet komutları kullanılmadan tez için oluşturulan yeni programlardır.

SOGR için oluşturulan program 7 bölümden oluşmuştur. 1. bölüm programı datalara uyarlama bölümü, bu bölümde kullanıcı harita boyutunu ve iterasyon sayısını girer, eğitim ve test dataları yüklenir, ağırlıklara eğitim datasından rastgele değerler atanır ve sabitler atanır. 2.bölüm ise eğitim bölümüdür, bu bölümde eğitim 4 aşamadan oluşturulmuştur, yani iterasyon sayısı dörde bölünüp, bu farklı 4 aşamada önce genel 8 komşuluk, sonra genel 4 komşuluk, sonra genel 2 komşuluk ve son adımda ise sadece genel 1 komşuluk güncellenmektedir ve her adımda öğrenme oranı lineer olarak azaltılmaktadır. Eğitim aşamasında rastgele eğitim giriş vektörü ile bütün düğümlerin ağırlık vektörleri arasındaki uzaklık hesaplanır. Sonra en küçük mesafeye sahip düğüm değeri bulunur ve o düğüm değerinin ağırlığı formüle göre güncellenir, daha sonra globale bakılarak eğitim aşamasına göre en küçük uzaklık değerine sahip düğümlerin ağırlıkları güncellenir. İterasyon sayısına göre eğitim yukarıda bahsedilen 4 adımda tamamlanır. Daha sonra sınıflama bölümü vardır, bu bölümde eğitimden sonra güncellenen ağırlıklarla yine ama bu sefer sırasıyla eğitim dataları ile ağırlıklar arasındaki uzaklıklar ölçülür. En yakın mesafeye sahip düğümün ağırlıkları o pozisyonadaki çıkışlarla sınıflanır ve o sınıfın sayıcı bir artırılır. Bir diğer bölüm ise eğitim etiketleme bölümüdür. Bu bölümde ise herbir eğitim giriş vektörü en yakın ağırlık vektörünün etiketini alır. Etiketlemede seçim; ağırlık vektörünün sınıfına kaydedilen

sayıcılarından hangisi büyükse o sayıcının sınıfını alarak seçilir. Bir diğer bölüm de eğitim başarımının ölçümüdür. Bu bölümde de eğitim için etiketlenen sınıflar ile eğitimin gerçek sınıfları karşılaştırılarak eğitim başarım oranları bulunur. Bir diğer bölüm de test etiketleme bölümüdür. Bu bölüm de eğitim etiketleme bölümü gibidir, yalnız eğitim giriş vektörleri yerine test giriş vektörleri kullanılır. Son bölümde test başarımının ölçümü, bu bölümde eğitim başarımının ölçüm bölümü gibidir, ama bu bölümde test etiketleme sınıfları ile test datasının gerçek sınıfları karşılaştırılarak test başarım sonuçları elde edilmiş olur. Öğrenme oranı 0,3 alınmıştır. İterasyon 5000, harita boyutu da 20x20 alınmıştır.

SOM için de oluşturulan program 7 bölümden oluşmuştur. SOM programının SOGR programından farkı eğitim bölümündedir, diğer bölümler aynıdır. Eğitim bölümde de eğitim 4 aşamadan oluşturulmuştur, yani iterasyon sayısını dörde bölünüp, bu farklı 4 aşamada önce lokal 8 komşuluk, sonra lokal 4 komşuluk, sonra lokal 2 komşuluk ve son adımda ise sadece kazanan düğümün ağırlığı güncellenmektedir ve her adımda öğrenme oranı lineer olarak azaltılmaktadır. SOM'un eğitim bölümünde rastgele eğitim giriş vektörü ile bütün düğümlerin ağırlık vektörleri arasındaki uzaklık hesaplanır. Sonra en küçük mesafeye sahip düğüm değeri bulunur ve o düğümün ağırlık değeri formüle göre güncellenir, daha sonra ızgara yapı komşuluğuna göre ve eğitim aşamasına göre önce lokal ızgara yapı 8 komşuluğu, sonra lokal ızgara yapı 4 komşuluğu, daha sonra lokal ızgara yapı 2 komşuluğu ve son olarak da sadece kazanan 1 düğümün ağırlığı güncellenir. İterasyon sayısına göre eğitim bu açıklanan 4 adımda tamamlanır. Öğrenme oranı 0,3 alınmıştır. İterasyon 5000, harita boyutu da 20x20 alınmıştır.

MLP' nin özellikleri ise;

```
net = newff([minmax(data)],[5 4 1],{'logsig','logsig','purelin'},'trainlm');
```

```
net.trainParam.lr= 0.3; net.trainParam.mc= 0.2; net.trainParam.goal= 0.001;
```

GRNN ve PNN ise Matlab NNet'teki standart (default değerleri ile) şeklinde kullanılmıştır. GRNN, PNN ve MLP için 1000 iterasyon kullanılmıştır.

6. SONUÇLAR, TARTIŞMA ve ÖNERİLER

Aşağıda tez için elde edilen 20 farklı tip data ve 5 farklı tip YSA ile uygulanmış ayrıntılı eğitim ve test sonuçları bulunmaktadır.

Çizelge 6.1 Hemoglobin-141 proteini ve 11- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 2047						
Eğitim Toplam Sarmal Sayısı : 1540						
Eğitim Toplam Sarım Sayısı : 507						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOG	93.3073	1910	1512	98.1818	398	78.5010
SOM	92.6722	1897	1501	97.4675	396	78.1065
GRNN	97.4597	1995	1521	98.7662	474	93.4911
MLP	97.8505	2003	1525	99.0260	478	94.2801
PNN	97.8994	2004	1526	99.0909	478	94.2801

Çizelge 6.2 Hemoglobin-141 proteini ve 11- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 683						
Test Toplam Sarmal Sayısı : 513						
Test Toplam Sarım Sayısı : 170						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOG	87.2621	596	482	93.9571	114	67.0588
SOM	86.3836	590	477	92.9825	113	66.4706
GRNN	86.2372	589	481	93.7622	108	63.5294
MLP	86.3836	590	476	92.7875	114	67.0588
PNN	86.2372	589	481	93.7622	108	63.5294

Çizelge 6.3 Hemoglobin-141 proteini ve 13- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 2015						
Eğitim Toplam Sarmal Sayısı : 1510						
Eğitim Toplam Sarım Sayısı : 505						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOG	93.8462	1891	1481	98.0795	410	81.1881
SOM	93.0025	1874	1468	97.2185	406	80.3960
GRNN	97.5186	1965	1492	98.8079	473	93.6634
MLP	96.8734	1952	1492	98.8079	460	91.0891
PNN	98.0645	1976	1496	99.0728	480	95.0495

Çizelge 6.4 Hemoglobin-141 proteini ve 13- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 673						
Test Toplam Sarmal Sayısı : 504						
Test Toplam Sarım Sayısı : 169						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOG	87.6672	590	473	93.8492	117	69.2308
SOM	86.6270	583	469	93.0556	114	67.4556
GRNN	86.0327	579	471	93.4524	108	63.9053
MLP	86.0327	579	471	93.4524	108	63.9053
PNN	86.1813	580	472	93.6508	108	63.9053

Çizelge 6.5 Hemoglobin-141 proteini ve 15- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 1985						
Eğitim Toplam Sarmal Sayısı : 1480						
Eğitim Toplam Sarım Sayısı : 505						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOG	93.5013	1856	1446	97.7027	410	81.1881
SOM	93.1989	1850	1444	97.5675	406	80.3960
GRNN	97.9345	1944	1464	98.9189	480	95.0495
MLP	86.9018	1725	1383	93.4459	342	67.7228
PNN	98.1864	1949	1466	99.0541	483	95.6436

Çizelge 6.6 Hemoglobin-141 proteini ve 15- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 661						
Test Toplam Sarmal Sayısı : 493						
Test Toplam Sarım Sayısı : 168						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOG	87.7458	580	462	93.7120	118	70.2381
SOM	86.9894	575	459	93.1034	116	69.0476
GRNN	85.9304	568	460	93.3063	108	64.2857
MLP	80.9380	535	442	89.6552	93	55.3571
PNN	86.0817	569	461	93.5091	108	64.2857

Çizelge 6.7 Hemoglobin-141 proteini ve 17- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 1953						
Eğitim Toplam Sarmal Sayısı : 1448						
Eğitim Toplam Sarım Sayısı : 505						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOGR	93.7020	1830	1417	97.8591	413	81.7822
SOM	93.5484	1827	1417	97.8591	410	81.1881
GRNN	97.9519	1913	1432	98.8950	481	95.2475
MLP	98.1567	1917	1434	99.0331	483	95.6436
PNN	98.1567	1917	1434	99.0331	483	95.6436

Çizelge 6.8 Hemoglobin-141 proteini ve 17- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 651						
Test Toplam Sarmal Sayısı : 483						
Test Toplam Sarım Sayısı : 168						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOGR	87.0968	567	452	93.5818	115	68.4524
SOM	87.2504	568	452	93.5818	116	69.0476
GRNN	85.8679	559	451	93.3747	108	64.2857
MLP	86.4823	563	446	92.3395	117	69.6429
PNN	85.8679	559	451	93.3747	108	64.2857

Çizelge 6.9 Hemoglobin-141 proteini ve 19- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 1922						
Eğitim Toplam Sarmal Sayısı : 1417						
Eğitim Toplam Sarım Sayısı : 505						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOGR	93.9646	1806	1383	97.6006	423	83.7624
SOM	93.1842	1791	1385	97.7417	406	80.3960
GRNN	97.9709	1883	1402	98.9414	481	95.2475
MLP	98.1790	1887	1404	99.0826	483	95.6436
PNN	98.1790	1887	1404	99.0826	483	95.6436

Çizelge 6.10 Hemoglobin-141 proteini ve 19- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 640						
Test Toplam Sarmal Sayısı : 473						
Test Toplam Sarım Sayısı : 167						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOGR	87.6563	561	441	93.2347	120	71.8563
SOM	87.3438	559	441	93.2347	118	70.6587
GRNN	85.6250	548	440	93.0233	108	64.6707
MLP	87.0313	557	436	92.1776	121	72.4551
PNN	85.7813	549	441	93.2347	108	64.6707

Çizelge 6.11 Hemoglobin-141 proteini ve 21- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 1891						
Eğitim Toplam Sarmal Sayısı : 1386						
Eğitim Toplam Sarım Sayısı : 505						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOG	93.1253	1761	1351	97.4747	410	81.1881
SOM	93.6013	1770	1349	97.3304	421	83.3663
GRNN	97.9376	1852	1371	98.9177	481	95.2475
MLP	97.7790	1849	1371	98.9177	478	94.6535
PNN	98.1491	1856	1373	99.0620	483	95.6436

Çizelge 6.12 Hemoglobin-141 proteini ve 21- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 629						
Test Toplam Sarmal Sayısı : 462						
Test Toplam Sarım Sayısı : 167						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOG	87.4404	550	431	93.2900	119	71.2575
SOM	87.5994	551	431	93.2900	120	71.8563
GRNN	85.3736	537	429	92.8571	108	64.6707
MLP	85.8506	540	428	92.6407	112	67.0659
PNN	85.5326	538	430	93.0736	108	64.6707

Çizelge 6.13 Hemoglobin-141 proteini ve 23- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 1859						
Eğitim Toplam Sarmal Sayısı : 1354						
Eğitim Toplam Sarım Sayısı : 505						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOGR	93.3297	1735	1322	97.6366	413	81.7822
SOM	93.9215	1746	1322	97.6366	424	83.9604
GRNN	97.9559	1821	1339	98.8922	482	95.4455
MLP	98.1173	1824	1341	99.0399	483	95.6436
PNN	98.1173	1824	1341	99.0399	483	95.6436

Çizelge 6.14 Hemoglobin-141 proteini ve 23- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 619						
Test Toplam Sarmal Sayısı : 452						
Test Toplam Sarım Sayısı : 167						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOGR	86.5913	536	420	92.9204	116	69.4611
SOM	86.9144	538	420	92.9204	118	70.6587
GRNN	84.9758	526	418	92.4779	108	64.6707
MLP	85.9451	532	418	92.4779	114	68.2635
PNN	85.1373	527	419	92.6991	108	64.6707

Çizelge 6.15 Hemoglobin-146 proteini ve 11- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 2024						
Eğitim Toplam Sarmal Sayısı : 1610						
Eğitim Toplam Sarım Sayısı : 414						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOGR	92.0455	1863	1556	96.6460	307	74.1546
SOM	91.9960	1862	1554	96.5217	308	74.3961
GRNN	97.7767	1979	1590	98.7578	389	93.9614
MLP	95.4051	1931	1592	98.8820	339	81.8841
PNN	98.5178	1994	1598	99.2547	396	95.6522

Çizelge 6.16 Hemoglobin-146 proteini ve 11- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 679						
Test Toplam Sarmal Sayısı : 536						
Test Toplam Sarım Sayısı : 140						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOGR	87.4260	591	511	95.3358	80	57.1429
SOM	86.9822	588	510	95.1493	78	55.7143
GRNN	91.8639	621	521	97.2015	100	71.4286
MLP	84.4675	571	475	88.6194	96	68.5714
PNN	91.7160	620	520	97.0149	100	71.4286

Çizelge 6.17 Hemoglobin-146 proteini ve 13- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 1995						
Eğitim Toplam Sarmal Sayısı : 1581						
Eğitim Toplam Sarım Sayısı : 414						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOGR	92.4812	1845	1534	97.0272	311	75.1208
SOM	91.8797	1833	1527	96.5844	306	73.9130
GRNN	97.7945	1951	1561	98.7350	390	94.2029
MLP	94.1353	1878	1536	97.1537	342	82.6087
PNN	98.4962	1965	1569	99.2410	396	95.6522

Çizelge 6.18 Hemoglobin-146 proteini ve 13- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 665						
Test Toplam Sarmal Sayısı : 527						
Test Toplam Sarım Sayısı : 138						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOGR	87.6692	583	501	95.0664	82	59.4203
SOM	86.7669	577	500	94.8767	77	55.7971
GRNN	91.7293	610	511	96.9639	99	71.7391
MLP	85.5639	569	472	89.5636	97	70.2899
PNN	91.7293	610	512	97.1537	98	71.0145

Çizelge 6.19 Hemoglobin-146 proteini ve 15- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 1967						
Eğitim Toplam Sarmal Sayısı : 1552						
Eğitim Toplam Sarım Sayısı : 415						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOGR	92.6792	1823	1502	96.7784	321	77.3494
SOM	92.5775	1821	1496	96.3918	325	78.3133
GRNN	98.0173	1928	1532	98.7113	396	95.4217
MLP	89.8322	1767	1479	95.2964	288	69.3976
PNN	98.6782	1941	1544	99.4845	397	95.6627

Çizelge 6.20 Hemoglobin-146 proteini ve 15- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 653						
Test Toplam Sarmal Sayısı : 517						
Test Toplam Sarım Sayısı : 136						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOGR	88.8208	580	497	96.1315	83	61.0294
SOM	86.5237	565	484	93.6170	81	59.5588
GRNN	91.5773	598	501	96.9052	97	71.3235
MLP	85.4518	558	480	92.8433	78	57.3529
PNN	91.4242	597	503	97.2921	94	69.1176

Çizelge 6.21 Hemoglobin-146 proteini ve 17- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 1935						
Eğitim Toplam Sarmal Sayısı : 1524						
Eğitim Toplam Sarım Sayısı : 411						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOGR	92.7132	1794	1476	96.8504	318	77.3723
SOM	92.5065	1790	1467	96.2598	323	78.5888
GRNN	97.9845	1896	1504	98.6877	392	95.3771
MLP	90.9044	1759	1459	95.7349	300	72.9927
PNN	98.8114	1912	1517	99.5407	395	96.1071

Çizelge 6.22 Hemoglobin-146 proteini ve 17- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 645						
Test Toplam Sarmal Sayısı : 508						
Test Toplam Sarım Sayısı : 137						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOGR	88.2171	569	486	95.6693	83	60.5839
SOM	86.8217	560	475	93.5039	85	62.0438
GRNN	91.4729	590	493	97.0472	97	70.8029
MLP	83.7209	540	463	91.1417	77	56.2044
PNN	91.4729	590	495	97.4409	95	69.3431

Çizelge 6.23 Hemoglobin-146 proteini ve 19- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 1935						
Eğitim Toplam Sarmal Sayısı : 1496						
Eğitim Toplam Sarım Sayısı : 409						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOGR	92.2835	1758	1450	96.9251	308	75.3056
SOM	92.5459	1763	1454	97.1925	309	75.5501
GRNN	98.1627	1870	1478	98.7968	392	95.8435
MLP	98.1627	1870	1479	98.8636	391	95.5990
PNN	98.9501	1885	1490	99.5989	395	96.5770

Çizelge 6.24 Hemoglobin-146 proteini ve 19- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 635						
Test Toplam Sarmal Sayısı : 599						
Test Toplam Sarım Sayısı : 136						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOGR	88.0315	559	474	94.9900	85	62.5000
SOM	87.7165	557	477	95.5912	80	58.8235
GRNN	91.3386	580	483	96.7936	97	71.3235
MLP	89.7637	570	470	94.1883	100	73.5294
PNN	91.3386	580	485	97.1944	95	69.8529

Çizelge 6.25 Hemoglobin-146 proteini ve 21- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 1875						
Eğitim Toplam Sarmal Sayısı : 1468						
Eğitim Toplam Sarım Sayısı : 407						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOG	92.6400	1737	1422	96.8665	315	77.3956
SOM	92.0000	1725	1408	95.9128	317	77.8870
GRNN	98.5600	1848	1454	99.0463	394	96.8059
MLP	93.8133	1759	1413	96.2534	346	85.0123
PNN	99.0400	1857	1463	99.6594	394	96.8059

Çizelge 6.26 Hemoglobin-146 proteini ve 21- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 625						
Test Toplam Sarmal Sayısı : 499						
Test Toplam Sarım Sayısı : 126						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOG	86.4000	540	465	93.1864	75	59.5238
SOM	87.5200	547	461	94.0816	86	63.7037
GRNN	91.3600	571	475	96.9388	96	71.1111
MLP	87.5200	547	455	92.8571	92	68.1481
PNN	91.3600	571	477	97.3469	94	69.6296

Çizelge 6.27 Hemoglobin-146 proteini ve 23- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 1845						
Eğitim Toplam Sarmal Sayısı : 1439						
Eğitim Toplam Sarım Sayısı : 406						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOG	92.7371	1711	1385	96.2474	326	80.2956
SOM	92.8997	1714	1389	96.5254	325	80.0493
GRNN	98.6450	1820	1427	99.1661	393	96.7980
MLP	98.6450	1820	1427	99.1661	393	96.7980
PNN	99.0786	1828	1434	99.6525	394	97.0443

Çizelge 6.28 Hemoglobin-146 proteini ve 23- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 615						
Test Toplam Sarmal Sayısı : 480						
Test Toplam Sarım Sayısı : 135						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOG	87.8049	540	451	93.9583	89	65.9259
SOM	87.4797	538	454	94.5833	84	62.2222
GRNN	91.5447	563	466	97.0833	97	71.8519
MLP	88.9431	547	449	93.5417	98	72.5926
PNN	91.5447	563	469	97.7083	94	69.6296

Çizelge 6.29 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 11- pencere boyutu için eğitim sonuçları (tabaka sınıfı çıkarılmış)

Eğitim Toplam Rezidü Sayısı : 1878						
Eğitim Toplam Sarmal Sayısı : 914						
Eğitim Toplam Sarım Sayısı : 964						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOGR	72.8967	1369	722	78.9934	647	67.1162
SOM	72.8435	1368	726	79.4311	642	66.5975
GRNN	99.8935	1876	913	99.8906	963	99.8963
MLP	87.1672	1637	792	86.6521	845	87.6556
PNN	99.9468	1877	914	100.00	963	99.8963

Çizelge 6.30 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 11- pencere boyutu için test sonuçları (tabaka sınıfı çıkarılmış)

Test Toplam Rezidü Sayısı : 760						
Test Toplam Sarmal Sayısı : 383						
Test Toplam Sarım Sayısı : 377						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOGR	56.9554	434	258	67.3629	176	46.6844
SOM	56.6929	432	258	67.3629	174	46.1538
GRNN	23.2283	177	114	29.7650	63	16.7109
MLP	62.3360	475	258	67.3629	217	57.5597
PNN	50.5249	385	380	99.2167	5	1.3263

Çizelge 6.31 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 15- pencere boyutu için eğitim sonuçları (tabaka sınıfı çıkarılmış)

Eğitim Toplam Rezidü Sayısı : 1836						
Eğitim Toplam Sarmal Sayısı : 880						
Eğitim Toplam Sarım Sayısı : 956						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOGR	74.0741	1360	704	80.0000	656	68.6192
SOM	71.0784	1305	653	74.2045	652	68.2008
GRNN	99.8911	1834	879	99.8864	955	99.8954
MLP	77.2331	1418	662	75.2273	756	79.0795
PNN	99.9455	1835	880	100.00	955	99.8954

Çizelge 6.32 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 15- pencere boyutu için test sonuçları (tabaka sınıfı çıkarılmış)

Test Toplam Rezidü Sayısı : 685						
Test Toplam Sarmal Sayısı : 361						
Test Toplam Sarım Sayısı : 324						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOGR	59.2431	407	241	66.7590	166	51.2346
SOM	54.1485	372	224	62.0499	148	45.6790
GRNN	23.2897	160	83	22.9917	77	23.7654
MLP	64.3377	442	241	66.7590	201	62.0370
PNN	52.5473	361	361	100.00	0	0

Çizelge 6.33 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 19- pencere boyutu için eğitim sonuçları (tabaka sınıfı çıkarılmış)

Eğitim Toplam Rezidü Sayısı : 1780						
Eğitim Toplam Sarmal Sayısı : 868						
Eğitim Toplam Sarım Sayısı : 912						
YSA Yöntemleri	Eğitim Başarım Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarım Sayısı	Eğitim Sarmal Başarım Oranı (%)	Eğitim Sarım Başarım Sayısı	Eğitim Sarım Başarım Oranı (%)
SOGR	72.0787	1283	685	78.9171	598	65.5702
SOM	72.0787	1283	678	78.1106	605	66.3377
GRNN	100.000	1780	868	100.000	912	100.000
MLP	95.2247	1695	822	94.7005	873	95.7237
PNN	100.000	1780	868	100.00	912	100.00

Çizelge 6.34 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 19- pencere boyutu için test sonuçları (tabaka sınıfı çıkarılmış)

Test Toplam Rezidü Sayısı : 738						
Test Toplam Sarmal Sayısı : 390						
Test Toplam Sarım Sayısı : 348						
YSA Yöntemleri	Test Başarım Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarım Sayısı	Test Sarmal Başarım Oranı (%)	Test Sarım Başarım Sayısı	Test Sarım Başarım Oranı (%)
SOGR	59.4595	440	285	73.0769	155	44.5402
SOM	55.9459	414	258	66.1538	156	44.8276
GRNN	29.0541	215	129	33.0769	86	24.7126
MLP	64.5946	478	263	67.4359	215	61.7816
PNN	52.7027	390	390	100.000	0	0

Çizelge 6.35 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 11- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 2008								
Eğitim Toplam Sarmal Sayısı : 841								
Eğitim Toplam Sarım Sayısı : 865								
Eğitim Toplam Tabaka Sayısı : 302								
YSA Yöntemleri	Eğitim Başarı Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarı Sayısı	Eğitim Sarmal Başarı Oranı (%)	Eğitim Sarım Başarı Sayısı	Eğitim Sarım Başarı Oranı (%)	Eğitim Tabaka Başarı Sayısı	Eğitim Tabaka Başarı Oranı (%)
SOGR	62.9482	1264	656	78.0024	538	62.1965	70	23.1788
SOM	64.2928	1291	662	78.7158	563	65.0867	66	21.8543
GRNN	99.9004	2006	840	99.8811	864	99.8844	302	100.00
MLP	96.7131	1942	833	99.0488	838	96.8786	271	89.7351
PNN	99.9502	2007	841	100.00	864	99.8844	302	100.00

Çizelge 6.36 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 11- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 804								
Test Toplam Sarmal Sayısı : 355								
Test Toplam Sarım Sayısı : 334								
Test Toplam Tabaka Sayısı : 115								
YSA Yöntemleri	Test Başarı Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarı Sayısı	Test Sarmal Başarı Oranı (%)	Test Sarım Başarı Sayısı	Test Sarım Başarı Oranı (%)	Test Tabaka Başarı Sayısı	Test Tabaka Başarı Oranı (%)
SOGR	47.2637	380	231	65.0704	133	39.8204	16	13.9130
SOM	47.1393	379	212	59.7183	159	47.6048	8	6.9565
GRNN	40.6716	327	16	4.5070	311	93.1138	0	0
MLP	50.3731	405	219	61.6901	129	38.6228	57	49.5652
PNN	44.7761	360	351	98.8732	9	2.6946	0	0

Çizelge 6.37 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 15- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 1986								
Eğitim Toplam Sarmal Sayısı : 840								
Eğitim Toplam Sarım Sayısı : 834								
Eğitim Toplam Tabaka Sayısı : 312								
YSA Yöntemleri	Eğitim Başarı Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarı Sayısı	Eğitim Sarmal Başarı Oranı (%)	Eğitim Sarım Başarı Sayısı	Eğitim Sarım Başarı Oranı (%)	Eğitim Tabaka Başarı Sayısı	Eğitim Tabaka Başarı Oranı (%)
SOGR	62.8902	1249	634	75.4762	547	65.5875	68	21.7949
SOM	63.1923	1255	648	77.1429	531	63.6691	76	24.3590
GRNN	99.8993	1984	839	99.8810	833	99.8801	312	100.000
MLP	94.2598	1872	811	96.5476	768	92.0863	293	93.9103
PNN	99.9496	1985	840	100.000	833	99.8801	312	100.000

Çizelge 6.38 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 15- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 766								
Test Toplam Sarmal Sayısı : 338								
Test Toplam Sarım Sayısı : 327								
Test Toplam Tabaka Sayısı : 101								
YSA Yöntemleri	Test Başarı Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarı Sayısı	Test Sarmal Başarı Oranı (%)	Test Sarım Başarı Sayısı	Test Sarım Başarı Oranı (%)	Test Tabaka Başarı Sayısı	Test Tabaka Başarı Oranı (%)
SOGR	49.6084	380	234	69.2308	134	40.9786	12	11.8812
SOM	46.7363	358	213	63.0178	135	41.2844	10	9.9010
GRNN	38.7728	297	13	3.8462	284	86.8502	0	0
MLP	43.7337	335	220	65.0888	67	20.4893	48	47.5248
PNN	44.1253	338	338	100.000	0	0	0	0

Çizelge 6.39 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 19- pencere boyutu için eğitim sonuçları

Eğitim Toplam Rezidü Sayısı : 1867								
Eğitim Toplam Sarmal Sayısı : 799								
Eğitim Toplam Sarım Sayısı : 782								
Eğitim Toplam Tabaka Sayısı : 286								
YSA Yöntemleri	Eğitim Başarı Oranı (%)	Toplam Eğitim Başarısı	Eğitim Sarmal Başarı Sayısı	Eğitim Sarmal Başarı Oranı (%)	Eğitim Sarım Başarı Sayısı	Eğitim Sarım Başarı Oranı (%)	Eğitim Tabaka Başarı Sayısı	Eğitim Tabaka Başarı Oranı (%)
SOGR	61.5961	1150	653	81.7272	452	57.8005	45	15.7343
SOM	64.0064	1195	643	80.4756	495	63.2992	57	19.9301
GRNN	100.000	1867	799	100.000	782	100.000	286	100.000
MLP	99.7322	1862	797	99.7497	782	100.000	283	98.9510
PNN	100.000	1867	799	100.000	782	100.000	286	100.000

Çizelge 6.40 2003 yılındaki Protein Yapı Tahmini Konferansı'nda sunulan 15 yeni protein ve 19- pencere boyutu için test sonuçları

Test Toplam Rezidü Sayısı : 825								
Test Toplam Sarmal Sayısı : 358								
Test Toplam Sarım Sayısı : 343								
Test Toplam Tabaka Sayısı : 124								
YSA Yöntemleri	Test Başarı Oranı (%)	Toplam Test Başarısı	Test Sarmal Başarı Sayısı	Test Sarmal Başarı Oranı (%)	Test Sarım Başarı Sayısı	Test Sarım Başarı Oranı (%)	Test Tabaka Başarı Sayısı	Test Tabaka Başarı Oranı (%)
SOGR	48.0000	396	246	68.7151	141	41.1079	9	7.2581
SOM	46.0606	380	238	66.4804	132	38.4840	10	8.0645
GRNN	35.8788	296	23	6.4246	273	79.5918	0	0
MLP	48.3636	399	193	53.9106	152	44.3149	54	43.5484
PNN	43.6364	360	358	100.000	2	0.5831	0	0

SOGR, SOM'a karşı 2004 yılının sonunda tamamlanmış yepyeni bir algoritmadır. Sonuçlardan da görüleceği gibi SOGR, SOM'dan ve diğer algoritmalarından da daha iyi bir sonuç vermiştir. SOGR ve SOM için 3 boyutlu topoloji kullanılmıştır. SOGR ve SOM'da ilk önce 8 komşuluk kullanılmıştır ve bu komşuluklar eğitim işleminde 4 kademedede azaltılmıştır. SOGR'ın SOM'dan farkı SOM'da belli bir komşuluk ilişkisi var. SOM'da kazanan düğüm ve onun komşularının ağırlıkları değişmektedir. SOGR'da ise her zaman kazanan düğümlerin ağırlıkları değişmektedir, lokal komşulukların bir anlamı yoktur, SOGR'da global sıralama vardır. Örneğin SOGR'da bir düğüm kazanırsa, daha sonra lokal komşuluğa bakılmaksızın, genele bakılarak ondan sonraki ağırlıkla giriş arasında en yakın mesafeye sahip olan düğüm kazanan düğüm olur ve sadece kazanan düğümlerin ağırlıkları değişir. Aynı zamanda SOGR, SOM'dan daha kısa sürede cevap vermektedir.

Bu çalışmada YSA algoritmalarında SOGR ve SOM için harita boyutu : 20x20 , iterasyon ise 5000 olarak alınmıştır. Diğer YSA algoritmalarında 1000 iterasyon alınmıştır. Çünkü GRNN ve PNN da iterasyon sayısı değişse bile sonuç değişmemektedir. MLP de ise iterasyon 1000 olmasına rağmen eğitim daha küçük iterasyon sayılarında tamamlanmaktadır. SOGR, SOM ve MLP programları 5 defa çalıştırılıp, ortalamaları alınmıştır.

Bu çalışmada görüldüğü gibi diğer yapılan çalışmalara göre daha yüksek doğruluk oranları, daha farklı YSA metodları ve de yepyeni bir algoritma olan SOGR'un ne kadar iyi sonuç verdiği görülmüştür. En iyi sonuç 11-13-15 pencere boyutlarında görülmüştür. Ayrıca en başarılı sınıflama sarmal yapısında görülmüştür. Tabaka yapısı proteinlerde en az görülen yapıdır ve tabaka yapısı genel doğruluk oranını düşürmektedir. Konferansta sunulan proteinlerin toplam rezidü sayıları ile hemoglobinlerin toplam rezidü sayıları birbirine yakın olmasına rağmen kestirim sonuçları birbirinden farklıdır. Bunun nedeni ise proteinlerin rezidüler ve ikincil yapılar arasındaki korelasyon farkıdır.

YSA algoritmaları incelendiği zaman, test doğruluk sonuçları için en iyi sonucu SOGR, eğitim doğruluk sonuçları için en iyi sonucu ise PNN vermiştir. Eğitim için özellikle GRNN ve PNN ile çok yüksek doğruluk oranı elde edilmiştir. MLP de veriyi çok iyi öğrenmiştir, çünkü geriye yayılım algoritması ile kendini her veriye göre uyarlayarak o yöne doğru yaklaşım yapmaktadır. Genelde ve özellikle de hemoglobinlerde GRNN ve PNN ise genel bir ifadeyle; kendilerine göre bir eğri uydurma (olasılık dağılım fonksiyonu) ile daha dengeli bir yaklaşım yapmıştır. Ayrıca GRNN ve PNN diğer algoritmalara göre daha hızlıdır.

Ayrıca GRNN,PNN,SOGR ve SOM, MLP'ye göre eğitim datasında bir değişiklik olduğunda YSA modelinin yeniden uyarlanması gerek yoktur. Genel olarak PNN ve GRNN sonuç olarak birbirine çok benzemektedir.

Sonuç olarak insan sağlığı bakımından çok önemli olan proteinlerin fonksiyonelliği (proteinlerin görevlerinin bilinmesi) için mutlaka proteinlerin ikincil yapılarının bilinmesi gereklidir. Proteinlerin ikincil yapılarının bilinebilmesi için deneysel yöntemler vardır, ama bu yöntemler çok zaman almakta, çok zor olmakta ayrıca da çok pahalı yöntemlerdir. Bu çalışma da gösteriyor ki YSA ve özellikle de SOGR diğer bir çok alanda olduğu gibi proteinlerin ikincil yapılarının kestiriminde de oldukça iyi sonuçlar vermiştir.

KAYNAKLAR

- Altun H., Eminoğlu U. ve Tezekici B.S., (2002), " Mlp Yapay Sinir Ağlarında Öğrenme Sürecinin Aktivasyon Fonksiyonu ve İstatiksel Değişim Gösteren Giriş Verilerine Bağımlılığı " Eleco'2002 Elektrik-Elektronik-Bilgisayar Mühendisliği Sempozyumu, Elektronik pp 310-314, Bursa, Aralık 2002.
- Barnes C., (1985), Invitation to Biology, Worth publishers, Inc, New York , s. 49
- Başbuğ R., (Şubat 1994), "Yapay Sinir Ağları", Byte, s.167.
- Carey J. ve Hanley V., (1997), Proteins, Princeton University, Biophysical Society On-line Textbook, Princeton, NJ 08544-1009
- Cichocki A. ve Unbehaven R., (1993), Neural Networks for Optimization and signal Processing, John Wiley and Sons, Gb. Stuttgart, pp.39-40.
- Demirsoy A., (1995), Kalıtım ve Evrim, Meteksan Yayıncılık, Ankara, Yedinci Baskı.
- De Wilde P., (1997), Neural Network Models, Second ed., Springer, London, pp.53-61.
- Ergezer H., Dikmen M. ve Özdemir E., (2003), " Yapay sinir ağları ve tanıma sistemleri ", Pivolka, 2(6), 14-17.
- Ersoy O., Erer I. ve Sağlam M.İ., (2004), " Self Organizing Global Ranking Algorithm and Its Applications ", Proceedings of the Conference on Smart Engineering Design (ANNIE 04) in St. Louis, November 2004
- Gözükara E., (1997), Biyokimya, İnönü Üniversitesi Tıp Fakültesi Biyokimya Ana Bilim Dalı Başk., Nobel Tıp Kitabevleri , Üçüncü Baskı, Cilt1. s.123-124-580
- Hagan M.T., Demuth H.B ve Beale M., (1996), Neural network design, PWS Publishing Company, Boston.
- Haykin S., (1994), Neural Networks: A Comprehensive Foundation, Macmillan College Publishing Company, Florida
- Hecht - Neilsen R., (1989), Neurocomputing, Addison-Wesley Pub. Comp., ,Massachusetts, pp.48-49.
- Hertz J., Krogh A. ve Palmer R. G., (1991) Introduction To The Theory of Neural Computation, Addison-Wesley Publishing Company ,Redwood City (CA), p.10
- Holley L.H. ve Karplus M., (1989), " Protein secondary structure prediction with a neural network " ,Department of Chemistry, Harvard University, Cambridge, MA 02138,Proc. Nati. Acad. Sci. USA, Vol. 86, pp. 152-156, Biophysics.
- Jones R.N. ve Karp A., (1990), Introducing Genetics, John Murray, Isbn 0-7195-4235-9. Chapter 14 & 17 p.186-243.
- Kangas J.A., Kohonen T.K. ve Laaksonen J.T., (1990), " Variants of Self-Organizing Maps " IEEE Transactions on Neural Networks, Vol. 1, No ,1.
- Kohonen T., (1997), Self-Organizing Maps, 2nd edition, Berlin: Springer.
- Lipton B., (2001), The Biology of Belief, Ingram, Baker & Taylor, New Leaf, Regional Wholesalers, Midpoint Trade, Chicago

- Maggiora G.M., Zhang C.T., Chou K.C. ve Elrod D.W., (1996), *Neural Networks in Qsar and Drug Design*, Academic Press, London, pp 255-279, ISBN 0122138155
- Masters T., (1993), *Practical Neural Network Recipes in C++*, Academic Press, New York, ISBN: 0-12-479040-2.
- Mount D.W., (2001), *Bioinformatics: Sequence and Genome Analysis*, University Of Arizona, Tucson, Cold spring Harbor Laboratory Press, ISBN 0879696087, New York.
- Pauling L., (1952), ‘ ‘ The Hemoglobin Molecule in Health and Disease. *Proceedings of the American Philosophical Society*, Vol. 96, No: 5, October, 555-565.
- Qian N., ve Sejnowski T.J., (1988), " Predicting the secondary structure of globular proteins using neural network models ", *J. Mol. Biol.*, 202, 865-884.
- Rinsho N., (1998), ‘ ‘ Human Hemoglobin Structure and Respiratory Transport. *Department of Human Genetics*’, *National Institute of Genetics*, 9, 54, 2320- 2325.
- Rojas R., (1996), *Neural Networks: A Systematic Introduction* , Springer-Verlag, Berlin, pp.150.
- Rost B. ve Sander C., (1994), " Combining Evolutionary Information and Neural networks to Predict Protein Secondary Structure ", *Vol. 19*, 55-72
- Sağıroğlu Ş., Beşdok E. ve Erler M., (2003), *Mühendislikte Yapay Zeka Uygulamaları Yapay Sinir Ağları*, *Ufuk Yayınları*.
- Schalkoff R.J., (1997), *Artificial Neural Networks*, McGraw-Hill Inc., Singapore.
- Specht D.F., (1999), “Probabilistic Neural Networks”, *Neural Networks*, vol.3, pp.100-118.
- Specht, D.F., (1991), “A General Regression Neural Network”, *IEEE Trans. Neural Networks*, Vol.2, No 6, pp. 568-576.
- Specht D.F., (1990), “Probabilistic neural networks”, *Neural networks*, vol.3, pp. 109-118.
- Tang J., (1995), *Design and Implementation of a Neural Network For the Evaluation of Protein Secondary Structure Prediction Using the DSSP and Chou-Fasman Algorithms*, *Biomolecular Engineering Research Center Boston University, Msc., Research Science Institute*.
- Temussi P.A., (1976), "Structural Characterization of Prebiotic Polypeptides", *Journal of Molecular Evolution*, 7:105
- Trippi R. R. ve Turban E., (1996), *Neural Network in Finance and Investing*, *Irwin Professional Pub.*, Chicago, p.4.
- Vemuri V. R., (1992), " Artificial Neural Networks: Concepts and Control Applications ", *IEEE Computer Society Press*, Los Alamitos, California, p.42.
- Wu C.H. ve McLarty J.W., (2000), *Neural Networks and Genome Informatics*, *University of Texas Health Center at Tyler Department of Epidemiology and Biomathematics*, Elsevier Science Ltd., ISBN 0080428002, Kidlington, 2000
- Yıldırım T. ve Kayaer K., (2003), " Medical Diagnosis on Pima Indian Diabetes Using General Regression Neural Networks ", *Int. Conf. on Artificial Neural Networks- Int. Conf. on Neural Information Processing*, 26-29 Haziran 2003, İstanbul.

Yıldırım T., (2003), " Uzaktan Algılama ve Yapay Sinir Ağları ", Türkiye I. Uluslararası Uydu-Anten-Elektronik İletişim Sistemleri Sempozyumu, 1 - 2 Mayıs 2003 Altınpark Fuar Merkezi, Ankara.

Yüçetürk A. C., (2000), Yapay Sinir Ağları Kullanılarak Örüntü Sınıflandırma ve Tanıma, Doktora Tezi, Ege Üniversitesi.

Zengin M. ve Aygün E., (1992), Atom ve Molekül Fiziği, Ankara Üniversitesi Yayınları.

İNTERNET KAYNAKLARI

- [1] www.genetikbilimi.com
- [2] www.blc.arizona.edu
- [3] www.fst.rdg.ac.uk
- [4] <http://dl.clackamas.cc.or.us>
- [5] www.biostat.wisc.edu/~craven/
- [6] <http://biyoloji.ksu.edu.tr/biyoinfo>
- [7] www.accessexcellence.org
- [8] www.mayo.caltech.edu/projects
- [9] <http://muhendislik.erciyes.edu.tr/tr/bilgisayar/bilakad/ss>
- [10] <http://www-biology.ucsd.edu/undergrad>
- [11] <http://www.uwyo.edu/>
- [12] <http://www.jcu.edu.au>
- [13] www.cs.wfu.edu/
- [14] www.rcsb.org/pdb/
- [15] <http://batitrakya.dostweb.com/yapaysiniraglari>
- [16] <http://www.anu.edu.au/>
- [17] <http://www.mathworks.com/access/helpdesk/help/toolbox/nnet/>

ÖZGEÇMİŞ

Doğum Tarihi 28.08.1980

Doğum Yeri İstanbul

Lise 1994-1998 Şişli Teknik Lisesi
(Elektrik-Elektronik Bölümü) (Okul 1.liği)

Lisans 1998-2002 Sakarya Üniversitesi Mühendislik Fakültesi
Elektrik-Elektronik Mühendisliği Bölümü
(Fakülte 1.liği ve Bölüm 1.liği)

Yüksek Lisans 2003-2005 Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü
Elektronik ve Haberleşme Müh. Anabilim Dalı,
Elektronik Programı

Çalıştığı Kurumlar

2003-2004 Baykal Akümülatör Tic. Ltd. Şti. (Yarı Zamanlı)

2004-2005 M.E.B. Cumhuriyet İlköğretim Okulu ve M.E.B.
Fevzi Çakmak İlköğretim Okulu, Vekil Sınıf Öğrt.