

T.C.
YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANABİLİM DALI

**PEDİGRİ VE MOLEKÜLER BELİRTEÇLER İLE FENOTİPİK ÖZELLİK
TAHMİN PERFORMANSININ YAPAY SİNİR AĞLARI İLE
KARŞILAŞTIRILMASI**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Mustafa Can YILMAZ
DANIŞMAN : Prof. Dr. Hayrettin OKUT

VAN-2015

T.C.
YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANABİLİM DALI

**PEDİGRİ VE MOLEKÜLER BELİRTEÇLER İLE FENOTİPİK ÖZELLİK
TAHMİN PERFORMANSININ YAPAY SİNİR AĞLARI İLE
KARŞILAŞTIRILMASI**

YÜKSEK LİSANS TEZİ

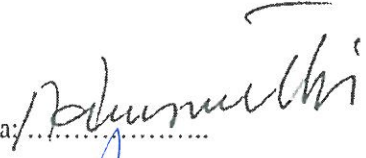
HAZIRLAYAN: Mustafa Can YILMAZ

VAN-2015

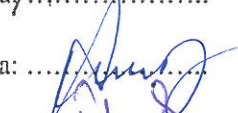
KABUL VE ONAY SAYFASI

Zootekni Anabilim Dalı'nda Hayrettin OKUT danışmanlığında, Mustafa Can YILMAZ tarafından sunulan'' Pedigri ve moleküler belirteçler ile fenotipik özellik tahmin performansının yapay sinir ağları ile karşılaştırılması'' isimli bu çalışma Lisansüstü Eğitim-Öğretim Yönetmeliği'nin ilgili hükümleri gereğince 09 / 11 / 2015 tarihinde aşağıdaki jüri tarafından oy birliği/ oy çokluğu ile başarılı bulunmuş ve Yüksek Lisans olarak kabul edilmiştir.

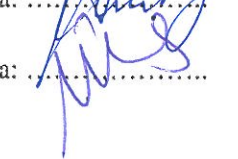
Başkan :Prof Dr. Hayrettin OKUT

İmza: 

Üye :Doç. Dr. Abdullah YEŞİLOVA

İmza: 

Üye :Yrd. Doç. Dr. Yılmaz KAYA

İmza: 

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun/...../2015 tarih ve sayılı kararı ile onaylanmıştır.

...../...../2015

Prof. Dr. Suat ŞENSOY

Enstitü Müdürü

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

Mustafa Can YILMAZ

ÖZET

PEDİGRİ VE MOLEKÜLER BELİRTEÇLER İLE FENOTİPİK ÖZELLİK TAHMİN PERFORMANSININ YAPAY SİNİR AĞLARI İLE KARŞILAŞTIRILMASI

YILMAZ, Mustafa Can
Yüksek Lisans Tezi, Zootečni Anabilim Dalı
Tez Danışmanı : Prof. Dr. Hayrettin OKUT
Kasım 2015, 61sayfa

Bu çalışmada, pedigrî ve yoğun DArT belirteç haritalarının kompleks bir fenotipik özelliği tahminleme yeteneklerinin karşılaştırılması amaçlanmıştır. Bu amaçla, üç farklı çapraz doğrulama oranı (%90, %70, %50 öğrenim kalanı test veri seti olmak üzere on tekrar) ile gizli katmanındaki aktivasyon fonksiyonu ve nöron sayısına göre farklılaşan dört yapay sinir ağı mimarisi incelenmiştir. Çalışmanın uygulama materyalini, 599 buğday hattına ait bir kompleks fenotipik özellik (tane verimi), pedigrî ve 1,279 DArT belirteç bilgisini içeren veri seti oluşturmuştur. Pedigrî ile DArT belirteç bilgisi karşılaştırıldığında, belirteç genomik bilgisinden, doğrulama oranları ve ağ mimarilerinden bağımsız olarak, daha yüksek korelasyon katsayıları ve daha küçük efektif parametre sayıları ve hata kareler ortalamaları elde edilmiştir. Bu sonuçlara göre yoğun moleküler belirteç bilgisinin pedigrîye göre tane veriminin tahmininde daha iyi bir performans gösterdiği söylenebilir. Ayrıca doğrusal olmayan sinir ağı mimarilerden doğrusal mimariler göre daha iyi sonuçlar elde edilmiştir. Bu sonuçlar, ıslah çalışmalarında, genomik belirteç haritaları ile pedigrîye göre daha yüksek verim alınabileceğini göstermiştir. Bunun yanı sıra, çok sayıda parametre ve doğrusal olmayan kompleks ilişkiler içeren yoğun belirteç haritalarının analizinde yapay sinir ağlarının etkili bir yöntem olduğu sonucuna varılabilir.

Anahtar kelimeler: Yapay sinir ağları, Bayesyen regülerizasyonu, Çeşitlilik dizileri teknolojisi, moleküler belirteç haritaları, pedigrî

ABSTRACT

COMPARISON OF PREDICTIVE PERFORMANCE OF PHENOTYPIC TRAITS WITH PEDIGREE AND MOLECULAR MARKERS THROUGH ARTIFICIAL NEURAL NETWORKS

YILMAZ, Mustafa Can
M. Sc., Animal Science
Supervisor: Prof. Dr. Hayrettin OKUT
Kasım 2015, 61 pages

In this thesis, predictive ability of pedigree and dense molecular markers on phenotypic traits was compared. For this comparison, artificial neural networks were used as statistical method. In the analysis, three different cross-validation ratio (90%, 70%, 50% learning data, left behind test data- ten times) and four network architectures differ from each other with their number of neurons and activation functions in hidden layer were tested. Data from 599 wheat lines with both phenotypic (grain yield) and genetic (pedigree and 1,270 DArT marker) information were used as study materials. Comparing genomic information with pedigree, in the results, higher correlation coefficients and less effective number of parameters and mean squared errors were obtained with marker information. From these results, it can be said that dense molecular marker information has showed better predictive performance than pedigree. In addition, better results were obtained from non-linear networks compare to linear networks. These results indicate that improving predictive efficiency in breeding studies could be accomplished through genomic marker maps. On the other hand, it can be concluded that artificial neural networks are efficient method at analyzing dense marker maps include great number parameters and complex relationships.

Key words: Artificial neural networks, Bayesian regularization, Diversity arrays technology, molecular marker maps, pedigree

ÖN SÖZ

Bu tez çalışmasında, her türlü ilgi ve yardımlarını esirgemeyen danışmanım Sayın Prof. Dr. Hayrettin OKUT'a teşekkür ederim. Ayrıca desteğini her zaman yanımda hissettiğim ve önerileri ile bana yol gösterici olan Sayın Doç. Dr. Abdullah YEŞİLOVA'ya ve Yrd. Doç. Dr. Gazel SER'e, her konuda yanımda olduğunu hissettiren ve katkıları ile bana yardımcı olan sevgili arkadaşlarım, Arş. Gör. Mehmet Reşit KARAGEÇİLİ ve Arş. Gör. Cihan ÇAKMAKÇI'ya, çalışmam süresince desteklerini ve yardımlarını esirgemeyen Arş. Gör. Talip Çakmakcı, Arş. Gör. Abdurrahim YILMAZ, Arş. Gör. Hilmi Kara ve Arş. Gör. Özlem GÜLDİĞEN'e teşekkürlerimi sunarım. Bana sonsuz manevi destek veren Arş. Gör. Dilek ÖZCAN, Arş. Gör. Merve Dilek GEBELOĞLU ve Arş. Gör. Sibel TURAN'a ve beni her konuda destekleyen sevgili aileme sonsuz teşekkürlerimi sunuyorum.

2015

Mustafa Can YILMAZ

İÇİNDEKİLER

	Sayfa
ÖZET	i
ABSTRACT	iii
ÖN SÖZ	v
İÇİNDEKİLER	vii
ÇİZELGELER LİSTESİ	ix
ŞEKİLLER LİSTESİ	xi
KISALTMALAR DİZİNİ	xiii
1. GİRİŞ	1
2. GENEL BİLGİLER	3
2.1. Diversity Arrays Technology Belirteçleri	3
2.1.1. DArT belirteçleri eldesi ve kullanımı	4
2.1.2. DArT uygulama alanları	6
2.2. Yapay Sinir Ağları	8
2.2.1. Yapay sinir ağları tarihi ve gelişimi	8
2.2.2. Biyolojik ve yapay sinir hücreleri	10
2.2.3. Yapay sinir hücresi elemanları	10
2.2.4. Yapay sinir ağları özellikleri	13
2.2.5. Yapay sinir ağları kullanım alanları	14
2.2.6. YSA öğrenme kurallarının sınıflandırılması	16
2.2.7. Basit algılayıcı modeli (Perseptron)	17
2.2.8. Çok katmanlı algılayıcı	18
3. KAYNAK BİLDİRİŞLERİ	21
4. MATERYAL VE YÖNTEM	27
4.1. Materyal	27
4.2. Yöntem	27
4.2.1. Genomik ilişkiler	27
4.2.2. Çapraz doğrulama (Cross-validation)	28
4.2.3. Yapay sinir ağları	29

4.2.4. Bayesyen regülerizasyonu	31
5. BULGULAR	41
5.1. Komplekslik Derecesine İlişkin Sonuçlar	41
5.2. Tahmin Performansına İlişkin Sonuçlar	43
6. TARTIŞMA VE SONUÇ.....	47
KAYNAKLAR.....	51
ÖZGEÇMİŞ.....	61

ÇİZELGELER LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Yapay sinir ağıları toplama fonksiyonları formülleri ve tanımları.....	11
Çizelge 2.2. Yapay sinir ağıları aktivasyon fonksiyonları formülleri ve tanımları.....	12
Çizelge 5.1. Efektif parametre sayıları tablosu.....	41
Çizelge 5.2. Shrinkage oranları tablosu.....	42
Çizelge 5.3. Test veri setlerinden elde edilen korelasyon katsayıları tablosu	44
Çizelge 5.4. Test veri setlerinden elde edilen hata kareler ortalamaları tablosu ...	45

ŞEKİLLER LİSTESİ

Şekil	Sayfa
Şekil 2.1. DArT belirteçlerinin DNA polimorfizmlerinin saptaması	5
Şekil 2.2. Rutin DArT mikroçipi şematik örneği	6
Şekil 2.3. Tek katmanlı algılayıcının şematik gösterimi.	9
Şekil 2.4. Biyolojik sinir ağı ve yapay sinir ağı.....	10
Şekil 2.5. Basit algılayıcı modeli şematik gösterimi	17
Şekil 2.6. Çok katmanlı algılayıcının şematik gösterimi.....	19
Şekil 4.1. Yapay nöronun şematik gösterimi.....	29
Şekil 4.2. İleri beslemeli yapay sinir ağının şematik gösterimi.....	30
Şekil 4.3. Yapay sinir ağları Bayesyen regülarizasyonu akış şeması.....	40
Şekil 5.1. Efektif parametre sayıları grafiği	42
Şekil 5.2. Shrinkage oranları grafiği.....	43
Şekil 5.3. Test veri setlerinden elde edilen korelasyon katsayıları grafiği	45
Şekil 5.4. Test veri setlerinden elde edilen hata kareler ortalamaları grafiği	46

KISALTMALAR DİZİNİ

Bu çalışmada kullanılmış bazı kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Kısaltmalar

ART	Adaptif rezonans teoremi
AFLP	Çoğaltılmış parça uzunluk polimorfizmi
AutoCM	Auto Contractive Map
CV	Çapraz doğrulama
CIMMYT	International Maize and Wheat Improvement Center
DArT	Diversity Arrays Technology
DNA	Deoksiribo nükleik asit
PZR	Polimeraz zincir reaksiyonu
RAPD	Rastgele arttırılmış polimorfik DNA
RFLP	Sınırlayıcı enzim parça uzunluğu çeşitliliği
RKHS	Reproducing kernel Hilbert spaces
SNP	Tek nükleotit polimorfizmi
SOM	Self-organizing map
SSR	Basit dizi tekrarları
SVR	Destek vektörü regresyonu
YSA	Yapay sinir ağları

1. GİRİŞ

Hayvan ve bitkilerin ıslah çalışmalarında, fenotipik özellik ve akrabalık ilişkileri kayıtlarının kullanıldığı pedigrî temelli geleneksel seleksiyon uzun yıllardır kullanılmaktadır. Seleksiyon çalışmalarında, bu kayıtları analiz etmek için seçilecek model ve yöntemler önemlidir. Hayvan ıslah çalışmalarında damızlık değer tahmininde en sık kullanılan klasik yöntem en iyi doğrusal sapmasız kestiricidir (BLUP; Henderson, 1984). Bitki ıslahçıları ise daha yakın zamanda, genetik değer tahmin etmede pedigrî bilgisini içeren karışık modelleri kullanmaya başlamışlardır (Oakey ve ark., 2006; Crossa ve ark., 2006; 2007).

Son yıllarda gelişen moleküler belirteç teknolojileri ile genom boyunca dağılan binlerce belirteçten oluşan genomik haritaları elde edilebilmektedir. Bu moleküler belirteç haritaları, genomik seleksiyon ile genetik değer tahmini yapılabilmesine olanak sağlamıştır. Hayvan ve bitki ıslah çalışmalarında, genetik değer tahmini veya hastalıkların kestirimi yaparken bu genomik bilgilerin hangi istatistiksel modellerle değerlendirileceği önemli bir problemidir. Genomik veri setlerindeki çok boyutluluk sorunu, belirteçler arasındaki interaksiyon veya çok kompleks fenotipik özellikler klasik yöntemlerle analize uygun olmayabilmektedir (Meuwissen ve ark., 2001).

Binlerce belirteçten oluşan genomik haritalar gibi büyük veri setlerinde önemli sorunlardan biri modelin veri ile aşırı uyum göstermesidir. Bu genelde, model çok kompleks ve modelin parametre sayısı çok fazla olduğunda meydana gelmektedir. Parametre sayısının birey sayısından fazla olması durumunda (teknik olarak bu durum $n < p$, küçük n büyük p olarak tanımlanır) model aşırı uyum gösterebilmektedir. Aşırı uyum gösteren modeller genellikle düşük tahmin performansına sahiptir. Bu durumdan bazı tekniklerle (çapraz onaylama, regülerizasyon, erken durdurma, vs.) kaçınılabilmektedir. Büyük veri setlerinde yaşanan diğer bir sorun, parametre sayısı arttıkça hata serbestlik derecesinin azalmasıdır ve bu hatanın olması gerekenden daha küçük çıkmasına, hatta klasik yöntemlerle hesaplanamamasına neden olabilmektedir. Ortak bilgi taşıyan parametreler bir araya getirilerek tek bir parametre olarak değerlendirilerek bu problem çözülebilmektedir (shrinkage). Bu LASSO, Ridge regresyon, LAR1, reproducing kernel Hilbert spaces, BayesA, BayesB, BayesC ve

BayesCpi gibi yöntemlerin kullanılması yapılabilmesine karşılık bu yöntemler doğrusal olmayan ilişkiler ile interaksiyon terimleri için uygun değildir. Fakat yapay sinir ağları, YSA, aktivasyon fonksiyonu diye adlandırılan bir dizi dönüşümler aracılığıyla doğrusal olmayan ilişkiler ile interaksiyonu terimlerini de modele dahil ederek parametre tahmini yapabilmektedir (MacKay, 1991; Haykin, 1994). Ayrıca problemlere göre farklı öğrenim algoritmalarını kullanabilmesi ve örneklerden öğrenebilen yapısı yapay sinir ağlarının diğer istatistiksel yöntemlere göre avantajları arasında sayılabilir (Tu, 1996).

Bu çalışmanın temel amacı, kompleks fenotipik özellikleri tahmin etmede pedigri ile moleküler belirteç bilgilerinin performanslarını karşılaştırmaktır. Çalışmanın uygulama materyalini, buğday hatlarına ait bir kompleks fenotipik özelliklerle beraber pedigri ve moleküler belirteç bilgilerini içeren bir veri seti oluşturmuştur. Çalışmanın ilk aşamasında, çapraz doğrulama yapmak için veri seti öğrenim ve test veri seti olmak üzere rastgele ayrılmıştır. Bununla beraber üç farklı doğrulama oranı kullanılmış ve bu oranlara ait sonuçlar ayrıca kendi aralarında karşılaştırılmıştır. Veri setlerinin analizinde, gizli katmandaki nöron sayısına ve aktivasyon fonksiyonuna göre değişen dört farklı Bayesyen YSA ağları mimarisi kullanılmıştır. Analiz sonucunda tahmin performanslarına ait parametreler elde edilmiştir.

2. GENEL BİLGİLER

2.1. Diversity Arrays Technology Belirteçleri

Son otuz yılda çok sayıda DNA tabanlı moleküler belirteç analizi yöntemi geliştirilmiştir. Bireyler ve ırklar arasındaki polimorfizmleri DNA düzeyinde tanımlayabilen moleküler belirteçler canlıların genotipinin belirlenmesinde anahtar rol oynamaktadır. Bu moleküler belirteç tekniklerinin geliştirilmesi, çalışmalar amacı doğrultusunda kullanılacak yöntemlerin seçiminde önem kazanmaktadır. Bu geliştirilen moleküler belirteç tekniklerinden ilki olan sınırlayıcı enzim parça uzunluk polimorfizmi (RFLP), Botstein ve ark. (1980) tarafından geliştirilmiştir. Teknik olarak uygulaması zahmetli olması ve yüksek miktarda DNA'ya ihtiyaç duymasına rağmen RFLP, çeşitli organizmalarda ilk moleküler haritaların yaratılmasına olanak sağlamıştır. Polimeraz Zincir Reaksiyonu tekniğinin (PZR) geliştirilmesi, yeni moleküler belirteç tekniklerinin geliştirilmesine olanak sağlamıştır. Rastgele arttırılmış polimorfik DNA (RAPD; Williams ve ark., 1990), çoğaltılmış parça uzunluk polimorfizmi (AFLP; Vos ve ark., 1995) ve genellikle bitki genetik çeşitliliğinin analizinde kullanılan basit dizi tekrarları (SSR; Weber ve May, 1989) bu moleküler belirteç teknikleri arasında gösterilebilir.

RFLP, RAPD ve AFLP gibi yöntemler bugünkü genom organizasyonu ve genetik varyasyon anlayışına büyük katkı sağlamaktadır. Ancak bu yöntemler jel elektroforezi temelli olmasından dolayı az sayıda çıktı alınabilmekte, bu da kullanım alanlarının sınırlandırılmasına neden olmaktadır (Williams ve ark., 1990; Young ve ark., 1992). Çoklu DNA parçalarının boyutlarını ayırt etmeye dayalı tüm yöntemler, jeller üzerindeki bantları tam ve doğru olarak ilişkilendirmekte zorluklar teşkil etmektedirler.

Bu kısıtlamaların üstesinden gelebilmek için önce tek nükleotit polimorfizmi (SNP) mikroçipleri ve yakın zamanda Diversity Array Technology (DArT) gibi hibridizasyon temelli yöntemler geliştirilmiştir. DNA mikro çipleri sayesinde lam gibi katı bir destek üzerinde tek seferde binlerce farklı genom bölgesi dizilenebilir. SNP mikroçip analizi ile genotipteki bu küçük fakat önemli farklılıkları tanımlamak nispeten pahalı ve zahmetli bir DNA dizileme sürecidir. Ayrıca poliploid bitkilerde genlerin çok sayıda kopyasının varlığı SNP analizinde zorluk teşkil etmektedir. DArT, SNP yöntemi

gibi mikroçip teknolojisi kullanarak hızlı ve ekonomik bir şekilde genotip analizi yapmayı sağlayan bir yöntemdir (Jacoud ve ark., 2001). Oligonükleotit çiplerinin aksine DArT, ön genom dizi bilgisine ihtiyaç duymaz, onun yerine birey veya türde genetik çeşitliliği temsil eden bir gen havuzundan elde edilen genetik bilgi alt kümesini kullanır. Bireylere ait DNA dizileri çipteki bölgelere hibridizasyon yapar ve özgül DNA parçalarının varlığı ve miktarı tespit edilerek genotiplenir. DArT, buğday (Akbari ve ark., 2006), muz (Amorim ve ark., 2009) ve şeker kamışı (Heller-Uszynska ve ark., 2011) gibi poliploid türlerde SNP teknolojisinin tersine iyi performans göstermiştir.

2.1.1. DArT belirteçleri eldesi ve kullanımı

Genomun kompleks yapısını basite indirgemek için çok sayıda yöntem geliştirilmiştir. DArT yöntemi özellikle aktif genlere ait genom parçalarını seçerek önemli bir avantaj sağlar. Bu seçim, genomdaki az tekrarlı dizileri ayırt edebilen restriksiyon enzimleri aracılığı ile sağlanır.

Mikroçip hazırlama aşamaları

Öncelikle ilgili türe ait bireylerden DNA ekstrasyonu yapılmaktadır. Sonrasında her bireye ait ekstrasyondan bir miktar alınarak karıştırılır ve karışımındaki uzun DNA iplikçikleri hedef bölgelere özgü enzimlerle kesilerek küçük parçalara ayrılır. Bu küçük parçalar bir vektör aracılığı ile *E. Coli* gibi bir bakteriye aktararak klonlanır. Böylece DNA örneğindeki karmaşıklık indirgenmiş olur ve bu klonlara DNA örneğinin temsili (representation) denmektedir.

DArT genomik temsilde iki tür parçacık olduğunu varsayarak çalışır. Bunlar:

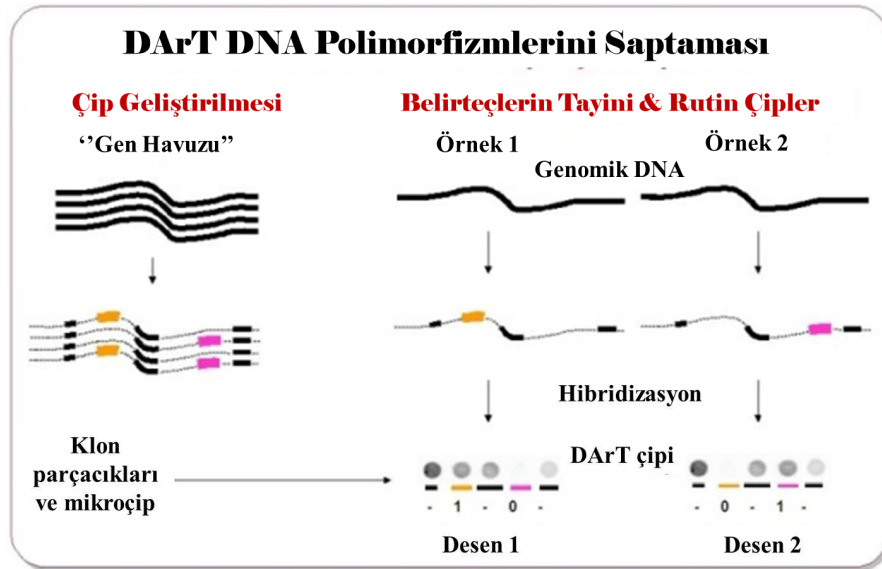
- Sabit parçacıklar; bireyler veya türlere ait DNA örneğinden hazırlanmış her temsilde bulunurlar.
- Değişken (polimorfik) parçacıklar; sadece bazı temsillerde bulunurlar.

Değişken parçacıklar, genomik temsilde hibridizasyon aracılığı ile o türe ait DArT belirteçleri mikroçiplerde var veya yok olma durumuna göre değerlendirilebilirler.

Kütüphanenin oluşturulmasında aşağıda verilen yollar izlenmektedir:

- Genetik çeşitliliği kapsayan bireylere ait genomik parçacıklardan oluşan bir gen havuzu oluşturulur.
- Bu parçacıklar bir vektör aracılığı ile *E. Coli*'ye aktarılır.
- Her koloni genomik temsile ait bir parçacığı taşır.

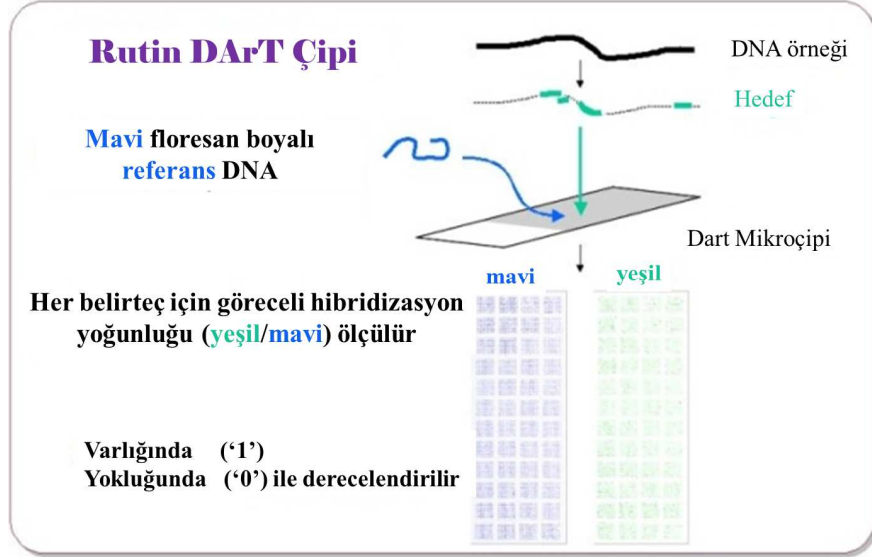
Kütüphane oluşturulduktan sonra seçili parçacıklar çoğaltılır ve cam plakalar üzerine yerleştirilerek mikroçipler oluşturulur. Bu mikroçipler her tür için bir defa oluşturulur ve türe ait gen havuzunu kapsayan bir genotip grubundan üretilen temsil parçacıklarını içerir.



Şekil 2.1. DArT belirteçlerinin DNA polimorfizmleri saptaması (Wenzl, 2008).

Genotipleme cam plaka yüzeyine kovalent bağlarla sabitlenmiş DNA parçacıkları ile genomik temsile ait parçacıklar ve vektöre ait parçacıklar arasında hibridizasyon yapma esasına göre yapılır. Bireye ait genomik materyaller bir floresan boya ile boyanırken, vektöre ait parçacıklar referans olarak başka bir floresan boya ile boyanır. Bireye ait temsildeki parçacıklar sadece mikroçip üzerindeki kendisi ile eşleşen parçacıklarla hibridizasyon yaparlar. Böylece her birey için kendine özgü bir çip deseni ortaya çıkar (Bkz. Şekil 2.1). Çiplerle genotipleme yaklaşık her 500 belirteçte 1 hata olmak üzere %99.8'lik bir hassasiyet gösterir (Akbari ve ark., 2006; Sessitsch ve ark., 2006). DArT belirteçlerini değerlendirmek için kullanılan algoritmalar belirteçlerin

kalitesini tahmin eder ve bu yeteneği farklı uygulamalarda kullanılacak belirteç setlerini belirlemekte kullanılabilir. Şekil 2.2’de rutin DArT mikroçipinin çalışma prensibi gösterilmiştir.



Şekil 2.2 Rutin DArT mikroçipi örneği (Wenzl ve ark., 2008).

2.1.2. DArT uygulama alanları

Genetik profileme ve genetik çeşitlilik haritaları

Mikroçip platformunda yeni türler için DArT belirteçleri farklı türlere ait genetik bilgiyi içeren DNA havuzundan hazırlanan binlerce DNA parçacığını içeren bir kütüphane taranarak elde edilebilir. Aynı anda analiz edilen çok sayıda DArT belirteci ile genetik çeşitlilik çalışmalarında yüksek çözünürlükte genotip incelemesi yapılabilmektedir (Jacoud ve ark., 2001; Wenzel ve ark., 2004; Akbari ve ark., 2006; Heller-Uszynska ve ark., 2006; Badea ve ark., 2008; Mace ve ark., 2008).

Genetik ve fiziksel haritalama çalışmaları

DArT belirteçleri çoğunlukla bitkilerde olmak üzere birçok organizmada genetik haritalama çalışmalarında kullanılmıştır. Şu ana kadar sadece buğday ve arpa 250’den

fazla popülasyon, haritalamada kullanılmıştır (Wittenberg ve ark, 2005; Akbari ve ark., 2006; Wenzl ve ark, 2006; Hearnden ve ark., 2007; Peleg ve ark., 2008).

Kantitatif özellik lokuslarını tanımlama

Uygun bir popülasyonda kantitatif özellik lokusları (QTL) analizi olarak bilinen yöntemin uygulanmasıyla, belirli kromozom bölgesindeki ilgili genlerin yerleri saptanabilir, etkilerinin büyüklüğü tahmin edilebilir, gen etkisinin eklemeli veya baskın olup olmadığı belirlenebilir. Uygulama kolaylığı, hızı ve göreceli düşük maliyeti nedeniyle DArT, kantitatif özellik lokusu (QTL) analizi ve bağlantı haritaları çalışmalarında kullanılmaktadır (Bedo ve ark., 2008; Crossa ve ark., 2007; Grewal ve ark., 2008; Huynh ve ark., 2008).

Genomik seleksiyon

Belirteç ve özellikler arası ilişkiler saptandığında, ilgili lokuslarla ilişkili belirteç seti seçilerek ayrı bir platformda daha az kompleks bir çip oluşturulabilir. Bu yüksek sayıda örneğin analizinde maliyet oranlarını önemli ölçüde düşürebilir. DArT belirteçleri düşük maliyet-çıkı oranına sahip olması ve poliploid organizmalarda yüksek oranda başarı göstermesi nedeniyle birçok bitki ıslah çalışmasında kullanılmıştır (Yang ve ark., 2006; Zhong ve ark., 2009; Poland ve ark., 2012).

Ekinlerde tür tanımlama ve genetik saflık testleri

DNA analizine dayanan genetik parmak izlemesi ve saflık testleri, tarım ve gıda sektöründe kullanılan modern kalite güvence testleri arasında önemli bir yere sahiptir. Çeşitli varyetelerine ait referans profilleri bulunan herhangi bir bitkiye ait tohum örneğinin genetik kimliği DArT ile tespit edilebilir. Özellikle buğday ve arpada geniş referans setleri oluşturulmuştur ve bu bitkilerde rutin olarak varyete belirleme ve hassas kontaminasyon türü ve seviyesi analizi yapılmaktadır (Liu ve ark., 2007; Bohra ve ark., 2011).

Kompleks DNA örneklerinin içeriğinin saptanması

Moleküler belirteçler, doğrudan toprak veya su gibi çevresel örneklerden elde edilen mikroorganizmaların içeriğini belirlemede kullanılan yöntemlerden biridir. İzole laboratuvar ortamında saf kültürlerde yapılan çalışmalar bakteriyal popülasyonların kendileri, çevreleri ve diğer popülasyonlarla olan etkileşimlerini analiz edememektedir. Bu nedenle, mikroorganizma koloni yapıları ve fonksiyonları doğal ortamında incelemek, mikrobiyal etkileşim belirleme ve biyoçeşitlilik alanlarında daha iyi bir bakış açısı sağlar. DNA mikroçip teknolojileri, belirli popülasyonları ve mikroorganizmaların fonksiyon gruplarını doğal çevrelerinde etkili bir şekilde izlemeye olanak sağlar (Ficetola ve ark., 2008).

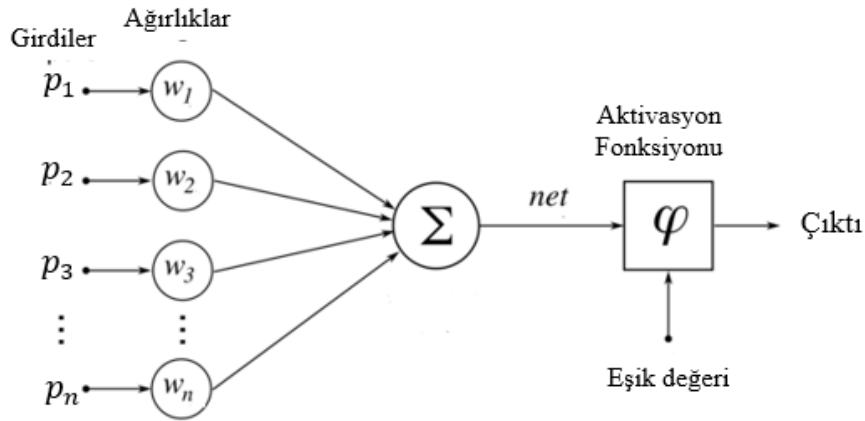
2.2. Yapay Sinir Ağları

Günümüzde, sinir sistemi ve özellikle beynin işleyişi ve yapısından yola çıkarak pek çok araştırma yapılmıştır. Bu araştırmaların sonucu olarak kompleks problemlerin çözümünde kullanılabilen yapay sinir ağları (YSA) yöntemi geliştirilmiştir. YSA, beyindeki sinir hücreleri gibi katmanlı ve paralel, birbiriyle bağlantılı, girdilerden sonuçları hesaplayabilen nöronlardan oluşmaktadır. Öğrenmeye dayalı ve adapte olabilen yapısıyla YSA, makine öğrenimi ve desen (pattern) tanıma yeteneğine sahiptir (Öztemel, 2003).

2.2.1. Yapay sinir ağları tarihi ve gelişimi

YSA'nın geçmişi modern bilgisayarlardan daha eskidir. Bu alanda ilk çalışmalar sinir hücresinin işleyiş yapısının modellenmesi ile başlamıştır. McCulloch ve Pitts (1943) tarafından yapılan çalışmada ilk yapay sinir ağı hücresini modelini (threshold logic) geliştirilmesiyle yapay sinir ağlarının temellerini oluşturulmaya başlanmıştır. Ardından sinirbilimci Hebb (1949) tarafından yapılan çalışmada sinir ağlarının parametrelerini değiştiren öğrenme kuralını (Hebbian Kuralı) geliştirmiştir. Hebbian kuralı tipik bir danışmansız öğrenme yöntemidir ve günümüzde kullanılan birçok öğrenme kuralının temelini oluşturmaktadır. 1950'li yıllarda hızlı bir gelişim gösteren

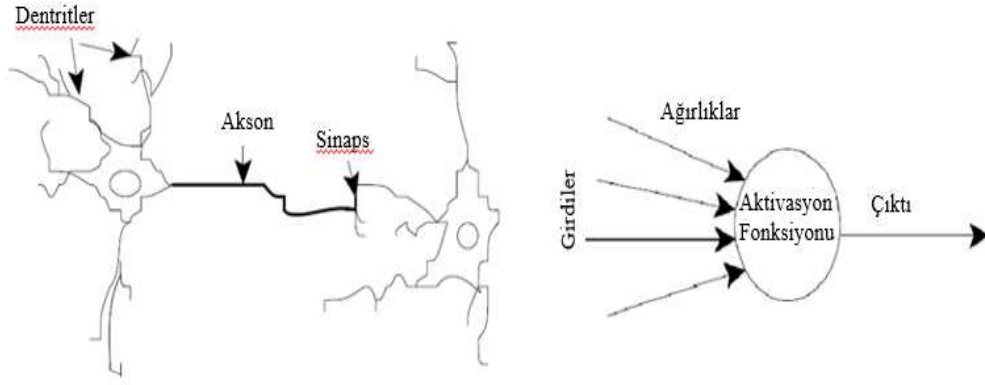
bilişim teknolojisi üzerine çalışan uzmanlar, sinirbilimcilerle temasa geçerek yapay zeka ve sinir ağları üzerine çalışmalar yapmışlardır. Bu çalışmalar sonucunda Rosenblatt, (1958) tarafından yapılan basit ve öğrenebilen bilgisayar ağı temelli desen tanıma algoritmasına dayalı çalışmada, tek katmanlı algılayıcıyı (perceptron) geliştirilmiştir (Şekil 2.3). Bu model, aynı zamanda daha sonra geliştirilecek olan ve YSA'da bir mihenk taşı sayılabilecek çok katmanlı algılayıcıların (multi layer perceptron) temelini oluşturmaktadır. Widrow ve Hoff (1960) Rosenblatt'ın algılayıcısına benzer niteliklere sahip ve öğrenim algoritması daha gelişmiş olan adaptif doğrusal nöron (ADALINE) modelini oluşturmuşlardır. Minsky ve Papert (1969) tarafından tek katmanlı algılayıcının iki temel sorunu keşfedilmiştir. Bu sorunlar tek katmanlı algılayıcıların bazı basit problemleri dahi çözmemesi ve dönemin bilgisayarlarının çok katmanlı büyük sinir ağlarını çalıştıracak kadar güçlü bir donanıma sahip olmamasıdır. Bununla beraber bu dönemde YSA'ında gelişmeler çok yavaşlamıştır. Ancak 1986 yılında geliştirilen geriye doğru yayılım (backpropagation) algoritmasıyla çok katmanlı ileri doğru besleme (multilayer perceptron) modellerinde parametrelerin nasıl güncellenebileceği bulunmuştur (Parker, 1985; LeCun, 1986). 1990'da bilgisayar teknolojisinin gelişmesi ve geliştirilen yeni yöntemler ile popülerliğini geri kazanmış ve bir bilim dalı olarak kabul görmüştür.



Şekil 2.3. Tek katmanlı algılayıcının şematik gösterimi.

2.2.2. Biyolojik ve yapay sinir hücreleri

Biyolojik sinir sistemi milyarlarca sinir hücresinin bir araya gelmesinden oluşur ve beyin çalışmasını sağlayan en temel yapı taşıdır. Sinir hücreleri bilgilerin elektriksel ve kimyasal yollarla işlenmesi ve iletilmesinden sorumludur. Yapay nöronlar modellenirken biyolojik nöronların temel özellikleri ele alınmıştır. Yapay nöron, ana hatlarıyla, ağırlıklarla (sinyallerin gücü) çarpılan girdiler (sinapslar) ve her nöronun aktivasyonunu sağlayan bir sayısal fonksiyonda meydana gelir (Şekil 2.4). YSA, yapay sinir hücrelerinin birbirleri ile çeşitli şekillerde bağlanmasından oluşmakta ve katmanlar halinde düzenlenmektedir.



Şekil. 2.4. Biyolojik sinir ağı ve yapay sinir ağı.

2.2.3. Yapay sinir hücresi elemanları

Girdiler

Girdiler sinir hücresine dış dünyadan gelen verilerdir. Girdiler nöron dış dünyadan olduğu gibi başka hücrelerden de gelebilir.

Ağırlık parametreleri

Sinir hücresine gelen girdiler, kendilerine ait bağlantılardaki ağırlık parametreleri ile ağırlandırılarak sinir hücresine gönderilir. Ağırlıklar eksi veya artı

değer alabilirler. Sinir modeline göre ağırlık parametreleri sabit olabilir veya yenilenebilir.

Toplama fonksiyonu

Sinir hücresine ağırlıklarla çarpılarak gelen net girdileri hesaplayan bir fonksiyondur. Toplama fonksiyonunda en yaygın kullanılan yöntem, kendi ağırlığı ile çarpılmış girdilerin toplandığı toplama yöntemidir. Çalışmalarda toplam fonksiyonunda çeşitli formüller kullanılmaktadır. Bir problem için en uygun toplama fonksiyonunu bulmayı sağlayan belirli bir formül yoktur. Bir yapay sinir ağı mimarisinde nöronlar birbirlerinden farklı toplama fonksiyonları kullanabilirler. Çizelge 2.1’de YSA’da kullanılan en çok kullanılan bazı toplama fonksiyonları verilmiştir.

Çizelge 2.1. Yapay sinir ağları toplama fonksiyonları formülleri ve tanımları

Yöntem	Formül	Açıklama
Toplam	$Net = \sum_i G_i A_i$	Girdiler ağırlık parametreleri ile çarpıldıktan sonra toplamaları alınarak net girdi hesaplanır.
Çarpım	$Net = \prod_{i=1} G_i A_i$	Ağırlıklar ile çarpılan girdiler ardından birbirileri ile çarpılarak hesaplanır.
Maksimum	$Net = \text{Max}(G_i A_i)$	Girdiler kendi ağırlık parametreleri ile çarpıldıktan sonra en büyük değer net girdi olarak kabul edilir.
Minimum	$Net = \text{Min}(G_i A_i)$	Girdiler kendi ağırlık parametreleri ile çarpıldıktan sonra en küçük değer net girdi olarak kabul edilir.
Kümülatif Toplam	$Net = Net(\text{eski}) + \sum_i G_i A_i$	Ağırlıklarla çarpılan girdiler önce gelen bilgilerle toplanarak net girdi elde edilir.

Aktivasyon fonksiyonu

Bu fonksiyon, nörona gelen girdiyi işleyerek nöronun bu girdiden üreteceği çıktıyı belirler. Nöronlarda, hücrenin gerçekleştireceği işleve ve girdi-çıkı yapısına göre çeşitli aktivasyon fonksiyonları kullanılabilir. Seçilen belli bir aktivasyon fonksiyonu çözülmeye çalışılan problem özelliğine göre diğer aktivasyon fonksiyonlarına göre daha tatmin edici sonuç verebilmektedir. Bir problem için en uygun fonksiyonu araştırmacı denemeleri ve tecrübe ile belirlenebilir. En uygun aktivasyon fonksiyonunu gösteren genel bir formül henüz bulunmamıştır. Çizelge 2.2’de YSA’da kullanılan bazı aktivasyon fonksiyonları verilmiştir.

Çizelge 2.2. Yapay sinir ağları aktivasyon fonksiyonları formülleri ve tanımları

Fonksiyon İsmi	Formül	Açıklama
Doğrusal Fonksiyon	$F(\text{Net})=A * \text{Net}$	Doğrusal problemleri çözmek için kullanılır. Girdiler sabit bir sayıyla çarpılarak çıktı hesaplanır.
Adım Fonksiyonu	$F(\text{Net}) \begin{cases} 0 \leq \text{Net} \text{ ise } +1 \\ 0 > \text{Net} \text{ ise } 0 \end{cases}$	Gelen Net girdi sıfırın altındaysa 0, sıfır veya üstünde ise 1 değerini alır.
Simetrik Adım Fonksiyonu	$F(\text{Net}) \begin{cases} 0 \leq \text{Net} \text{ ise } +1 \\ 0 > \text{Net} \text{ ise } -1 \end{cases}$	Gelen Net girdi sıfırın altındaysa -1, sıfır veya üstündeyse +1 değerini alır.
Log-Sigmoid Aktivasyon Fonksiyonu	$F(\text{NET}) = \frac{1}{1 + e^{-\text{Net}}}$	Sürekli ve türevi alınabilen ve doğrusal olmayan bir fonksiyondur. Çıktısı 0 ile 1 arasında bir değer alır
Hiperbolik Tanjant Sigmoid	$F(\text{Net}) = \frac{e^{\text{Net}} + e^{-\text{Net}}}{e^{\text{Net}} - e^{-\text{Net}}}$	Log-Sigmoid ile benzer bir fonksiyondur. -1 ile + 1 arasında değerler alabilir.

Sinir hücresinin çıktısı

Aktivasyon fonksiyonu tarafından hesaplanarak dış dünyaya veya başka bir nörona gönderilen çıktı değeridir. Bir hücrenin çıktısı, kendine veya başka bir nörona girdi olabilir.

2.2.4. Yapay sinir ağları özellikleri

YSA'nın en önemli özelliği ağın öğrenme yeteneğidir. Öğrenim sürecinde, problemin genel özelliklerini elde eder ve böylelikle öğrenim sırasında kullanılmayan girdiler için de anlamlı yanıtlar üretebilir. Sinir ağına çözülmek üzere bir problem verildiğinde, ağ gözlemleri kullanarak o problemi çözebilecek en uygun yöntemi arar. Kullanılan bu gözlemlere öğrenim (training) gözlemleri denilmektedir. Problemi çözecek herhangi bir yöntem tanımlamadan, bir program veya özel bir algoritma geliştirmeden sadece benzer problemlerin çözümünü değerlendirerek problemi çözebilir.

YSA, değişiklere kendini adapte edebilir. Yani, belirli bir problemi çözmek amacıyla öğretilen ağ, problemdeki değişikler veya yeni bilgiler karşısında tekrar öğrenebilmekte, kendini yenileyebilmektedir. YSA uygulamaları genellikle yazılım tabanlı olmasına rağmen, donanımsal bir çok YSA uygulaması vardır. Yüksek paralel işlem yeteneğine sahip donanım sistemleri ile uygulanabilir. Bu özellik, YSA'nın hızlı bilgi işleme yeteneğini artırır ve gerçek zamanlı uygulamalarda arzu edilmektedir. Paralel yapısı nedeniyle ağın sahip olduğu bilgi, ağdaki bütün bağlantılar üzerine dağılmış durumdadır. Bu nedenle, ağdaki bazı bağlantıların veya hücrelerinin etkisiz hale gelmesi, ağın doğru bilgi üretmesini önemli ölçüde etkilemez (Haykin, 1994).

Yapay sinir ağlarının başlıca avantajları:

- YSA makine öğrenmesi gerçekleştirebilirler.
- Hata toleransına sahiptirler. Eksik bağlantı veya nöron performanslarını önemli düzeyde etkilemez.
- Kendi kendine öğrenebilme ve organize etme yetenekleri vardır.
- YSA modelleri girdiler ve çıktılar arasındaki doğrusal olmayan ilişkileri tam olarak belirleyebilirler.

- Bilgiler veri tabanları veya belirli dosyalarda değil ağına tamamına yayılarak değerler ile ölçülen ağ bağlantılarında saklanmaktadır.
- Öğrenimleri sırasında kendilerine verilen örneklerden genellemeler çıkarırlar ve bu genellemeler ile yeni örnekler hakkında bilgi üretebilirler.
- Sinir ağlarında farklı öğrenme algoritmaları kullanılabilmektedir.
- YSA modellerinde verilerin dağılımı için herhangi bir varsayım öngörülmemektedir.

Yapay sinir ağlarının başlıca dezavantajları:

- Paralel işlem yapabilme yeteneği olan yüksek donanımlara ihtiyaç duyarlar.
- Probleme uygun ağ yapısının belirlenmesi için geliştirilmiş bir kural yoktur. Uygun ağ yapısı deneyim ve deneme yanılma yolu ile belirlenmektedir.
- Sinir ağları birer kara kutudur. Ağına kurduğu nedensel ilişkileri ve problem çözme yöntemini tanımlamak ve gizli katmandaki sonuçları yorumlamak çok zordur.
- YSA modelleri aşırı uyum (overfitting) sorununa yatkın olabilmektedirler (Jain, 1996; Tu, 1996; Öztemel, 2003)
-

2.2.5. Yapay sinir ağları kullanım alanları

YSA, özellikle eksik ve normal olmayan verilerle çalışabilme yeteneği sayesinde günümüzde en başarılı ve güncel desen tanıyıcı ve sınıflandırıcı yöntemler arasında sayılabilir. Özellikle çok sayıda bilgi işlemi gereken problemlerde başarılı bir şekilde çalışabilmektedir. Yapay sinir ağlarının, doğrusal olmayan, çok boyutlu, gürültülü, karmaşık, kesin olmayan, eksik, kusurlu, hata olasılığı yüksek sensör verilerinin olması ve problemin çözümü için özellikle belli bir matematik modelin ve algoritmanın bulunmaması hallerinde yaygın olarak kullanıldıkları görülmektedir.

YSA, genellikle şu temel fonksiyonları yerine getirmek için kullanılmaktadır:

Tahmin

Sisteme ait tahmin edilebilir parametreler genellikle denklemlerle değerlendirilerek ve ifade edilmektedir. Tahmin ise bu denklemlerin değerlendirmesi ve

çözümleridir. YSA'nın avantajı, sadece örnek verilerle tahmin yapabilmesi ve herhangi bir ek bilgiye ihtiyaç duymamasıdır. Sistem, geçmişe ait verilerden gizli bağımlılıkları ve ilişkileri öğrenerek gelecekteki olayları tahmin etmeye çalışmaktadır. Meteoroloji, borsa ve döviz kuru tahmini, jeolojide deprem tahmini kullanıldığı alanlara örnek gösterilebilir (Refenes ve ark., 1994; Taylor ve Buizza, 2002; Adeli ve ark., 2009).

Sınıflandırma

Sınıflandırmada kullanılan yapay sinir ağları, kendilerine sunulan verileri kategorize etmektedirler. Makine öğrenmesi terminolojisinde, bu bir danışmanlı öğrenim olarak kabul edilebilmektedir. El yazısı ve biyometrik tanımlama buna örnek olarak gösterilebilir (Gader ve ark., 1997; El-Bakry, 2001).

Veri ilişkilendirme

Ağa tanıtılan verilerin, hatalı veya eksik olup olmadığını belirlenir. Örnekten öğrendiği bilgilerle eksik olan bilgileri tamamlar veya hatalı verileri düzenleyebilir. Eksik resmin tamamlanması buna örnek olarak verilebilir (Moolman ve ark., 1994).

Veri filtreleme

Bu amaçla kullanılan ağlar, büyük ölçekli veriler arasından önemli bilgilere ulaşmayı sağlarlar. Günümüzde birçok veri arasından ihtiyaç olan bilgilere ulaşmak önemli bir hal almıştır. Veri tabanlarında ve yüksek hacimli veri elde edilen çalışmalarda yaygın olarak kullanılmaktadır (Brameier ve ark., 2001).

Tanıma ve eşleştirme

Değişik şekil ve desenleri tanıma, bilgilerin doğruluğunu, düzenini, sürekliliğini belirlemede ve eşleştirme fonksiyonlarında YSA kullanılabilir. Ses ve yüz tanıma, şekil analizleri, tıpta tanı koyma (kanser taraması vb.) gibi işlemlerde kullanılabilir (Wu ve ark., 1993; Rowley ve ark., 1998).

2.2.6. YSA öğrenme kurallarının sınıflandırılması

Yapay sinir ağları gibi öğrenmeye dayalı sistemler farklı öğrenme stratejileri kullanılmaktadır. Öğrenmede kullanılan kurallara ve metotlara göre ağın parametreleri değişmektedir. Temel olarak 3 farklı öğrenme kuralından söz edilebilir.

Danışmanlı öğrenme

Bu öğrenmede örnekler sisteme girdi/çıkı seti olarak verilir. Her örnek, bir girdi ve o girdi karşılığında oluşması gereken çıktı çiftinden oluşmaktadır. Danışmanlı öğrenme, öğrenim veri setini analiz eder ve bu bilgiyi yeni örnekleri haritalamak için kullanır. Ağ verilen girdiler ile istenen çıktıları elde edebilmek için ağın parametrelerini günceller. Ağın çıktıları ile beklenen çıktılar arasındaki fark hesaplanarak ağın yeni parametreleri bu hataya göre düzenlenmektedir.

Danışmansız öğrenme

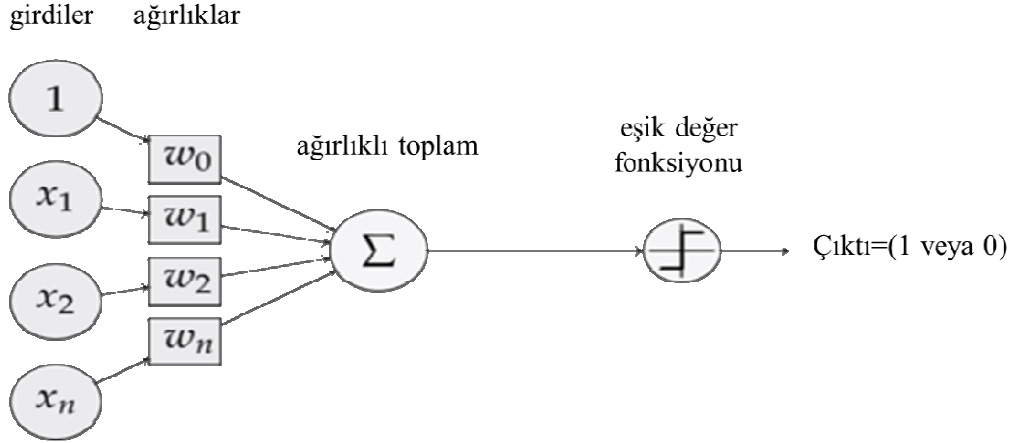
Danışmansız öğrenmede sistemin öğrenmesine yardımcı olan herhangi bir danışman yoktur. Öğrenme sırasında ağa sadece örnek girdiler verilir ve ağ her bir örneği kendi içinde sınıflandıracak şekilde ağ kurallarını oluşturur. Bu öğrenme daha sınıflandırma problemleri için kullanılmaktadır. Adaptif rezonans teoremi (ART) ve Kohonen'ın 1982 yılında geliştirdiği Self-organizing map (SOM) danışmansız öğrenmeye örnek gösterilebilir (Carpenter ve Grossberg, 1988).

Destekleyici öğrenme

Bu öğrenmede, sistem öğrenmesi için girdi setine ait hedef çıktıları sisteme göstermek yerine ağın girdilere göre kendi çıktısını üretmesini bekler ve üretilen çıktının doğruluğuna göre bir sinyal gönderir. Ağ bu sinyale göre parametrelerini güncelleyerek öğrenme sürecini devam ettirmektedir.

2.2.7. Basit algılayıcı modeli (Perseptron)

Perseptron, ikili sayı sistemine dayanan bir danışmanlı öğrenme modelidir. Ağ girdinin hangi gruba ait olduğunu belirler. Doğrusal bir sınıflandırıcıdır. Çıktı değerinin hesaplanmasında eşik değer fonksiyonu kullanılır.



Şekil 2.5. Basit algılayıcı modeli şematik gösterimi (Bargen, 2013).

Şekil 2.5'te de görüldüğü gibi, perseptron tek bir sinir hücresinden oluşur. Girdiler ve her girdi setine ait çıktılar ağırlıklı olarak tanıtılır. Sonraki adımda, öğrenme kuralına göre ağırlık çıktı değerleri hesaplanır ve hesaplanan çıktı değeri, olması gereken çıktı değerinden farklı ise ağırlıklar ve eşik değeri öğrenme kuralına göre yenilenir.

Basit algılayıcı öğrenme algoritması

Ağırlık girdi seti ve o sete ait beklenen çıktı tanıtılır. Çıktı, sadece 1 ve 0 değerini alabilir.

$$f(x) = \begin{cases} w * x + b > 0 \text{ ise } 1 \\ w * x + b \leq 0 \text{ ise } 0 \end{cases} \quad (2.1)$$

Eş. 2.1'de w ağırlık parametrelerini, x girdileri, ve b ise sapmayı temsil etmektedir. Eğer hesaplanan çıktı ile beklenen çıktı aynı olursa ağırlık parametrelerinde herhangi bir değişiklik yapılmamaktadır. Eğer farklı çıkarsa hesaplanan çıktının beklenen çıktıdan büyük veya küçük çıkmasına göre iki farklı değer almaktadır.

Ağın beklenen çıktısı 0 iken ağın gerçekleşen çıktısı 1 ise ağırlık değerleri azaltılmaktadır. Ağırlıklar parametreleri, girdi değerinin belirli bir oranı (λ) kadar azaltılır.

$$w_n = w_o - \lambda x \quad (2.2)$$

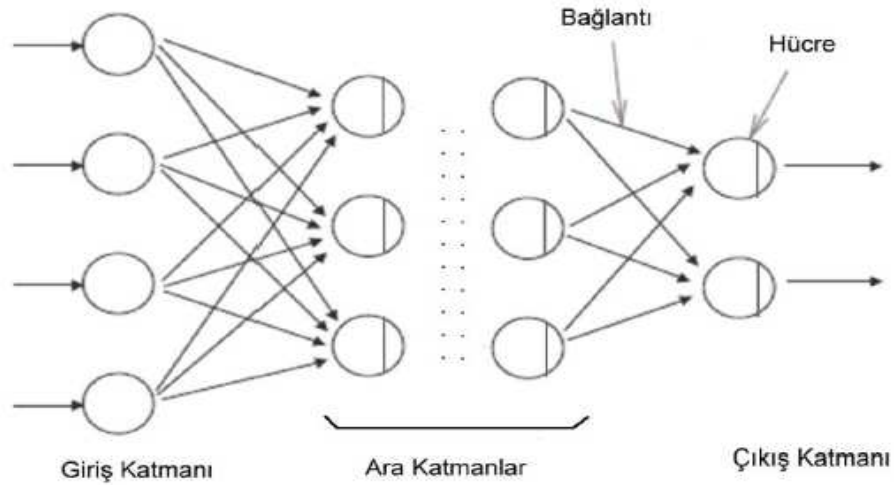
Ağın beklenen çıktısı 1 iken gerçekleşen çıktının 0 olması durumunda ise ağırlık parametrelerinin değerleri artırılır.

$$w_n = w_o + \lambda x \quad (2.3)$$

Yukarıdaki adımlar doğru sınıflandırmalar yapılmaya kadar tekrar edilmektedir (Haykin, 1994; Hagan ve ark., 1996).

2.2.8. Çok katmanlı algılayıcı

Yapay sinir ağının öğrenmesi istenen örnekteki girdiler ve çıktılar arasındaki bağlantılar doğrusal değilse tek katmanlı algılayıcı ile doğru bir sınıflandırma yapılamaz. Günlük olaylar arasındaki ilişkilerin büyük çoğunluğunun doğrusal olmaması, araştırmacıları bu alanda çalışmaya itmiştir. Çok katmanlı algılayıcı doğrusal olmayan ilişkileri belirleyebilen ilk sinir ağı sistemidir ve yapay sinir ağlarında yeni bir dönemin açılmasını sağlamıştır. Bu sinir ağı tipi günümüzde hemen hemen bütün mühendislik problemlerine çözüm üretebilecek güce sahiptir. Özellikle sınıflandırma, tanıma, sinyal işleme ve genelleme yapmayı gerektiren problemler için önemli bir çözüm aracıdır. Bu sinir ağı tipi, sayısal formüller içeren çoklu katmanlardan oluşur ve aralarındaki bağlantılar ileriye doğru kurulmuştur. Her nöron kendinden sonraki katmandaki nöronlara bağlanır. Girdi katmanı, gizli katman ve çıktı katmanı olmak üzere 3 temel katmandan oluşur (Şekil 2.6).



Şekil 2.6. Çok katmanlı algılayıcının şematik gösterimi.

Girdi katmanı

Yapay sinir ağına dış dünyadan bilgilerin geldiği katmandır. Bu katmanda bilgiler işlenmeden gizli katmana gönderilir. Şekil 2.6'da görüldüğü gibi bir girdi sonraki katmandaki tüm nöronlara bağlıdır.

Gizli katman

Girdi katmanından çıkan bilgiler, gizli katmana gönderilir. Gizli katman sayısı ağın mimarisine göre birden fazla olabilmektedir. Gizli katmanlardaki nöron sayıları girdi ve çıktı sayısından bağımsızdır. Katman ve nöron sayısının artırılarak daha karmaşık problemler çözülebilir ve fakat bu artış hesaplama süresini de artırmaktadır.

Çıktı katmanı

Gizli katmandan gelen bilgileri işleyerek dış dünyaya çıktıları gönderen katmandır. Bu katmandaki her nöron bir önceki katmandaki tüm nöronlara bağlıdır ve her nöron bir tane çıktı üretmektedir.

Çok katmalı algılayıcıda öğrenme, ağın örnek veri seti işlenirken her tekrarda bağlantı parametrelerini beklenen çıktıya göre yenilemesiyle sağlanır. Bu, bir

danışmanlı öğrenme örneğidir ve geri yayılım (backpropagation) ile çalışır. Çok katmanlı algılayıcının öğrenme kuralına, basit algılayıcı modelindeki en küçük kareler yönteminin genelleştirilmiş hali denebilir.

Geri yayılım (Backpropagation)

Ağın ürettiği çıktı ile beklenen çıktı arasındaki farkı karşılaştırarak hata hesaplanır. Geri yayılım algoritmasında bu hata ağırlık parametrelerine dağıtılarak sonraki iterasyonda hatanın azaltılması sağlanır. İleri yayılım içinde kullanılan aktivasyon fonksiyonlarının türevi alınarak aynı bağlantılar aracılığıyla, geri doğru yayılmaktadır. Çıktı katmanındaki toplam hata aşağıdaki eşitlikle hesaplanır (Hagan ve ark., 1996).

$$E_D = \frac{1}{2} \sum_{j=1}^p (t_j - \hat{t}_j)^2 \quad (2.4)$$

Buna, amaç fonksiyonu da denmektedir. Geri yayılım algoritmasının hedefi bu fonksiyonu minimum yapmaktır. YSA, amaç fonksiyonunun minimum değer alması için ağırlık parametreleri gradyan alçalma (gradient descent) yöntemi ile düzenlenmektedir. Bu yöntemin amacı YSA ağırlıklarının en uygun biçimde değiştirilmesini sağlamaktır. YSA'daki her ağırlık değerinin değişme miktarı;

$$\Delta w_{ij} = -\alpha \frac{\partial E}{\partial w_{ij}} \quad (2.5)$$

eşitliğinden yola çıkarak bulunmaktadır Burada α öğrenme katsayısıdır. Ağırlık parametresindeki değişim, eski ağırlığa eklenir. -1 ile çarpım hata oranının minimuma doğru güncellenmesi için gereklidir (Bishop, 1995).

3. KAYNAK BİLDİRİŞLERİ

Jacoud ve ark. (2001) tarafından Diversirty Arrays Technology RFLP, SSR ve SNP gibi moleküler belirteç teknolojilerine bir alternatif olarak geliştirilmiştir. Çalışmada, dokuz farklı pirinç türünden (üç indica, altı japonica) elde edilen DNA parçacıkları DArT mikroçiplerinde analiz edilerek DNA polimorfizmleri tespit edilmiştir. Özel DNA parçacıklarının varlığına ve yokluğunu değerlendiren DArT'ın mikroçip temelli herhangi genomda veya gen havuzunda kullanılabilecek hızlı ve ekonomik bir uygulama olduğunu öne sürmüşlerdir. Böylece ilerde yapılacak genetik haritalama, gen kaynakları karakterizasyonu, gen etiketleme, moleküler belirteç destekli seleksiyon gibi DArT uygulamalarının önü açılmıştır.

Günümüzde birçok canlı türüne ait DArT mikroçipleri geliştirilmiştir. Şu ana kadar çoğu bitki olmak üzere 127 türe ait 667 DArT kütüphanesi oluşturulmuştur. DArT ile yapılan ilk çalışmalar genellikle bitkilerde DNA profillemesi ve genetik değer arttırmaya yönelik olmuştur. Wenzl ve ark., (2004) tarafından yapılan çalışmada arpada tüm genom profillemesi çalışmasında elde edilen DArT çipleri %98'lik çağrı oranı ve %99.8'lik bir puanlama hassasiyeti göstermiştir. Bu çalışma, DArT belirteçlerinin genetik haritalama ve genetik çeşitlilik analizinde başarılı bir şekilde kullanılabileceğini göstermiştir. Manyok'ta (*Manihot esculenta*) yapılan genotipleme çalışmasında %98.1 çağrı oranı ve %99.8'lik puanlama hassasiyeti ile benzer sonuçlar elde edilmiştir (Xia ve ark., 2005).

Model organizma *Arabidopsis thaliana* mevcut genom dizileri kullanılarak 107 DArT belirtecinin niteliği ve belirteçlerin moleküler bazda polimorfik özellikleri değerlendirilmiştir. Analizde, genetik bağlantı haritasındaki 107 DArT belirtecinin polimorfik özellikler üzerinde yadsınamaz etkisinin olduğu tespit edilmiştir. Sonuçlar DArT belirteçlerini kullanarak orta yoğunluklu genetik bağlantı haritaları oluşturulabilineceğini göstermiştir (Wittenberg ve ark., 2004).

Hekzaploid ekmek buğdayında (*Triticum aestivum L.*) yapılan DNA profillemesi çalışmasında daha önce yapılan diploid bitkilerle yapılan çalışmalarla benzer sonuçlar gözlenmiştir (%95 çağrı oranı, %99.8 puanlama hassasiyeti) . Ayrıca toplamda 339 SSR, RFLP ve AFLP belirteci bilgisi bulunan ve aynı DArT sayıda belirteci bulunan

Cranbrook × Halberd çaprazlanmış haploid hatlarda çağrı oranı ve genom kapsama derecesi bakımından karşılaştırılmıştır (Akbari ve ark., 2006). Ekmek buğdayı ile yapılan diğer bir genetik çeşitlilik çalışmasında 256 DArT, 16 AFLP primer kombinasyonu ve 633 SSR belirteci kullanılmıştır. DArT belirteçlerinin AFLP'ye göre daha yüksek, SSR ile ise benzer oranda polimorfik bilgi sağladığı tespit edilmiştir (Stodart 2008).

Poliploid buğday bitkileri üzerinde genetik çeşitlilik, genom profillemeye ve haritalama çalışmalarında DArT başarılı sonuçlar vermiştir (Badea ve ark., 2008; Francki ve ark., 2008; Harvorth ve ark., 2009; Peleg ve ark., 2009; Griffiths ve ark., 2009-2010; Zhang ve ark., 2011)

Mace ve ark. (2008) tarafından *Sorghum bicolor* üzerinde yapılan genetik çeşitlilik analizi ve harita çalışmasında 90 bitki hattından 508 DArT belirteci elde edilmiştir. Yüksek sayıda DArT belirtecinin genotipler arasındaki genetik ilişkileri hassas bir şekilde tahmin edebilmesinin yanında belirteçlerin genom üzerinde dağılımının moleküler ıslah çalışmaları ve genetik uygulamalarda avantaj sağladığı belirtilmiştir.

Genetik benzerlik ve bağlantı haritaları çalışmada 11520 çavdar klonu ve toplam 2700 DArT belirteci kullanılmıştır. Analiz sonucunda 1022 DArT belirtecini çavdar varyete ve hatlar arasında genetik ilişkilerin belirlenmesinde etkili olduğu ortaya çıkmıştır (Bolibok-Bragoszewska ve ark., 2009).

Toplamda 168 birey olmak üzere *Musa acuminata* ve *Musa balbisiana* isimli iki farklı yabani muz türünün kullanıldığı DNA analizi çalışmasında toplam 836 DArT belirteci geliştirilmiştir. DArT ile muz genotiplerinin diğer moleküler belirteçlere göre daha hızlı ve ucuz bir şekilde analiz edilebildiği belirtilmiştir (Risterucci ve ark., 2009)

Yüksek ekonomik değere sahip diğer bir poliploid bitki olan şeker kamışı üzerinde yapılan genetik çeşitlilik analizi çalışmasında uygun maliyetli ve etkili bir yöntem olduğu gösterilmiştir (Heller-Uzynska ve ark., 2010).

DArT belirteçleri ile çalışılan ilk meyve ağacı olan elma ağacına ait genetik bağlantı haritası Schouten ve ark. (2011) tarafından oluşturulmuştur. Elde edilen DArT belirteçlerinde tekrarlı yüksek benzerliğe sahip dizilere bağlı olarak %52-54 artık belirteç olduğu tespit edilmiştir. DArT belirteçlerinin elma ağacının genetik haritalamasında uygun ve düşük maliyetli bir yöntem olduğunu kanıtladığı belirtilmiştir.

Raman ve ark. (2013) tarafından DArT ve diğer moleküler belirteçlerle dört farklı kıta altı farklı kolza tohumu (*Brassica napus*) popülasyonuna ait haritayı temel alınarak *Brassica rapa* ve *Brassica oleracea*'ya ait tarımsal kalitatif ve kantitatif özelliklerle ilişkili DArT lokusları tanımlanmıştır. Fenolojik bileşenler, tohum ve yağ kalitesi özellikleri, bor verimliliği, sakaroz taşıma, erkek kısırlığı ve yanıkara hastalığına direnç gibi özelliklerle bağlantılı DArT belirteçleri tanımlanmıştır.

Yapay sinir ağları ile SNP gibi hibridizasyon temelli yüksek çıktıya sahip genetik verilerin analizi 2000'li yıllarda başlamıştır ve genellikle bu çalışmalar tıp alanında yapılmıştır. Tomida ve ark. (2002) tarafından çalışmada, SNP bilgisi ile astım bronşit, atopik dermatit, alerjik rinit ve alerjik konjonktivit gibi alerjik hastalıkların tahmininde YSA ve çoklu regresyon analizi karşılaştırılmıştır. YSA ile hastalıkların tanı tahmini %78'lik bir doğruluk oranıyla çoklu regresyon analizine göre daha iyi bir performans göstermiştir.

Tomita ve ark. (2004) tarafından çocuklarda alerjik astım tanısı ile ilgili olarak yapılan çalışmada, uygun 25 SNP'den YSA ile 10 efektif SNP belirlenmiştir. Bu 10 SNP ile YSA ile analizi sonucunda öğrenim veri setinde %97.7, değerlendirme veri seti ile %74.4 korelasyon tespit edilmiştir.

Kronik hepatit C hastalarının tedavisinde interferon tek başına veya ribavirin kombinasyonları ile verilebilmektedir. Çalışmada SNP, viral genotip, viral yük, yaş ve cinsiyet analizi yapılarak tedavi için en etkin interferon kombinasyonu tahmin edilmeye çalışılmıştır. 523 kronik hepatit C hastasına ait kan örnekleri kullanılmıştır. İnterferon sinyalizasyon ve immunomodülasyon yolları ile bağlantılı altı gene ait 7 SNP tespit edilmiştir. SNP, viral genotip, viral yük, yaş ve cinsiyet parametrelerini ele alan YSA modelinden %77.4 ortalama doğruluk oranı elde edilmiştir. Bu çalışma, kişiye özgü ilaç kullanımda genetik belirteç kullanımının gelecek vaat ettiğini göstermiştir (Lin ve ark., 2006).

Curtis (2007) tarafından yapılan çalışmada ise simülasyon ile elde edilmiş SNP veri setlerinde YSA ve diğer çoklu belirteç testlerinin performansı karşılaştırılmıştır. YSA, standart haploit temelli yöntemlere göre haploit frekanslarının tahmininde daha iyi bir performans göstermiştir. Permutasyon testi, YSA ile karşılaştırılabilir sonuçlar vermesine karşın permutasyon testinde geçerli bir p değeri elde etme gerekliliği

uygulama hızını düşürmektedir. Özellikle yüksek sayıda belirteç bulunan verilerin analizinde YSA'nın kullanılması önerilmiştir.

Chun ve ark. (2008) çalışmasında, gastrik kanserinin saptanmasında yüksek yoğunlukta SNP belirteçleri ve makine öğrenmesine benzer bir yöntem olan durum tabanlı çıkarsama kullanılmıştır. Bu çalışma, veri madenciliği uygulamalarının insan genom bilgisinden yararlanarak tıp bilişimi alanında kullanılabileceğini göstermiştir.

Buscema ve ark. (2012) yaptığı çalışmada ise hem fenotipik hem genetik faktörlerin etkilediği olduğu bir hastalık olan amiotrofik lateral sklerozun genomik bilgi ile ilişkisi nispeten yeni bir yapay sinir ağı varyasyonu olan Auto Contractive Map (AutoCM) ile analiz edilmiştir. AutoCM öğrenme görevine YSA'den farklı olarak rastgele ağırlık ataması yapmadan başlar. Ağırlıklar, tüm çıkış noktalarında değerler sıfırlandığında yakınsama kriterine ulaşır. Ağırlıkların matrisi, Öklid uzayı çarpılarak bir veri seti oluşturur. Bu vaka kontrol çalışmasında, AutoCM tabanlı veri madenciliği yaklaşımı ile 35 gene ait SNP analiz edilmiştir. Toplam kompleksliğe 0.17 katkıda bulunan koruyucu faktörü olarak gösterilen üç gen ve kompleksliğe 0.13 katkıda bulunan duyarlılık faktörü olarak gösterilen üç gen tespit edilmiştir

de los Campos ve ark. (2009) tarafından yapılan çalışmada Bayesyen Lasso regresyon yöntemi ile pedigree ve yoğun belirteç haritaları ile kantitatif özelliklerin arasındaki ilişkiler analiz edilmiştir. Çalışmada 1,884 fareye ait 10,946 polimorfik belirteç ve 599 buğday hattına ait 1,279 DArT belirteç ve pedigree bilgilerinden oluşan iki veri seti kullanılmıştır. Kantitatif özellik olarak vücut kütle endeksi ve tane verimi seçilmiştir. İki veri setine pedigree temelli, belirteç temelli ve pedigree&belirteç temelli üç farklı model uygulanmıştır. Pedigree&belirteç temelli model diğer modellere göre daha iyi bir tahmin performansı göstermiştir. Belirteç haritalarının, pedigree ile tespit edilemeyen bazı genomik bilgileri tespit edebildiği belirtilmiştir.

Crossa ve ark. (2010) yapılan çalışmada, doğrusal karışık model, reproducing kernel Hilbert spaces (RKHS) ve Bayesyen Lasso yöntemlerinin genetik değer tahmini performansını karşılaştırılmıştır. Çalışmada, de los Campos ve ark. (2009) tarafından da kullanılan buğday veri seti ve 300 mısır hattına ait 1,148 SNP bulunan bir veri seti kullanılmıştır. Analizden önce gözlemlere 10 tekrar rastgele CV uygulanmıştır. Buğday veri setinde belirteç temelli modeller, pedigree temelli modellere göre analiz yapan yöntemden bağımsız olarak daha iyi bir tahmin performansı göstermişlerdir. Bayesyen

Lasso ve RKHS yöntemleri doğrusal karışık modele göre daha iyi bir performans göstermelerine karşın kendi aralarında genel bir üstünlük tespit edilmemiştir.

Long ve ark. (2011) tarafından yapılan çalışmada destek vektörü regresyonu (SVR) ile Bayesyen Lasso yönteminin tahmin performansını Holştayn sığırı ve buğday veri setinde karşılaştırılmıştır. 4,703 holştayn sığırının 3,305'ini (%70) rastgele öğrenim grubu olarak seçmiş geri kalan 1,398 (%30) sığır ise test grubu olmuştur. Genetik belirteç olarak her sığır için 32,518 SNP bilgisi ve fenotipik özellik olarak ise geleneksel döl kontrolü ile elde edilmiş sütçülük fiziksel özellikleri aktarma yeteneği bilgileri kullanılarak analiz yapılmıştır. 599 buğday hattı, %80 öğrenim, kalan %20 test olmak üzere rastgele 50 kere rastgele ayrılmıştır. 1,270 DArT belirteci ile tane veriminin analizi yapılmıştır. Kullanılan kernel fonksiyonuna göre doğrusal veya doğrusal olmayan ϵ -SVR ve en küçük kareler SVR ile Bayesyen Lasso karşılaştırılmıştır. Tahmin performansı, hata kareler ortalaması ve test veri setinden elde edilen korelasyon katsayıları ile ölçülmüştür. Doğrusal olmayan SVR doğrusal SVR'ye göre daha iyi, Bayesyen Lasso ile ise benzer performans göstermiştir. Buğday verisinde doğrusal olmayan ϵ -SVR, doğrusal SVR'ye göre % 17,5'lik büyük bir artış göstermiştir

Okut ve ark. (2011) tarafından yapılan çalışmada 1896 fareye ait fenotip ve genotip (12,320 lokus) bilgisi kullanılarak yapılan çalışmada Bayesyen ileri beslemeli sinir ağları ile vücut kitle endekslerini analiz edilmiştir. Çalışmada gizli katmanındaki nöron sayısı ve aktivasyonuna göre sekiz farklı sinir ağı mimarisi kullanılmıştır. Gizli katmandaki nöron sayısı arttıkça efektif parametre sayısı bir noktaya kadar artmış ve sonrasında dengelenmiştir. Gizli katmanında beş nöron bulunan sinir mimarisi aralarındaki fark önemsiz çıksa da sinir ağları arasında en küçük hatayı ve en yüksek korelasyonu göstermiştir. Ayrıca 17 SNP'nin diğerlerine göre vücut kütle indeksine daha fazla etkisi olduğu tespit edilmiştir. Bu çalışma, Bayesyen sinir ağlarının kompleks özelliklerdeki doğrusal olmayan ilişkileri analiz etmede etkili bir yöntem olduğunu göstermiştir.

Gianola ve ark. (2011) tarafından yapılan çalışmada Jersey ineği ve buğday veri setinde Bayesyen sinir ağları ile pedigree ve genetik belirteç bilgisinin tahmin yeteneğini karşılaştırılmıştır. Jersey veri setinde 297 birey 35,798 SNP bilgisi ve fenotipik süt, yağ ve protein verimi bulunmaktadır. Jersey veri setinde SNP verileri kullanılarak yapılan süt, yağ ve protein verimi tahmini pedigree kullanılarak yapılan tahmine göre daha iyi bir

performans göstermiştir. Doğrusal olmayan mimariler buğdayda daha fazla olmak üzere iki veri setinde de doğrusal ağırlara göre daha iyi tahmin performansı göstermişlerdir.

4. MATERİYAL VE YÖNTEM

4.1. Materyal

Bu çalışmada, CIMMYT Global Wheat Breeding Programı'nda geliştirilen 599 buğday hattından oluşan veri seti kullanılmıştır. Veri setinde fenotipik özellik olarak tane verimi seçilmiş ve dört farklı hedef çevre grubu belirlenmiştir. 599 hat arasındaki eklemeli genetik ilişki matrisi (A) McLaren ve ark., (2005) tarafından tanımlandığı şekilde, pedigrî bilgisinden yararlanılarak oluşturulmuştur. Bu matrisin elemanları, hat çiftleri arasındaki akrabalık katsayılarının iki katına eşittir.

Veri setindeki buğday hatları Tricarte Pty. Ltd. (Canberra, Australia; <http://www.triticarte.com.au>) tarafından üretilmiş ve 1447 Diversity Array Technology (DArT) belirteci kullanılarak genotiplenmiştir. Bu belirteçler varlıklarında bir yokluklarında sıfır olmak üzere iki farklı değer alabilir. Bu veri setinde 1 olarak kodlanmış alellerin minimum ve maksimum frekansı sırasıyla 0.008 ve 0.987 ve ortalama frekansı ise 0.561'dir. Alel frekansı <0.05 veya >0.95 olan belirteçler veriden çıkarılmıştır. Eksik genotipler marker genotiplerinin marjinal dağılışımdan örnekler kullanılarak atanmıştır. Bu örnekler Bernoulli ($\hat{p}_j$) dağılımı gösterir. $\hat{p}_j$ eksik olmayan genotiplerden hesaplanarak tahmin edilen alel frekansıdır. İşlemlerden sonra 1279 belirteç kalmıştır.

Analizlerde MATLAB 2013a'nın Neural Network Toolbox[®] paketi kullanılmıştır (Beal ve ark., 2010). Ayrıca çapraz doğrulama için SAS 9.4 paket programı kullanılmıştır.

4.2. Yöntem

4.2.1. Genomik ilişkiler

Toplam genetik etkilerin doğrusal model tahminleri, karışık modellerden veya seleksiyon indeks eşitliklerinden elde edilebilir. $\mathbf{u}$, her alel veya belirteç için eklemeli genetik etkiyi içeren vektör olsun. $\mathbf{M}$ her bireyde kalıtımı yapılan alelleri gösteren ilişki

matrisini ifade etmektedir. **M** matrisi majör ve minör alelleri gösteren 1 ve 0'lardan oluşur. **MM'** matrisinin diyagonal elemanları bireylerdeki toplam majör alel sayısını, diyagonal olmayanlar elemanlar ise bireyler arasındaki ortak alel sayısını vermektedir. **M'M** matrisini diyagonal elemanları her belirtece ait majör alel sayısını, diyagonal olmayan elemanlar ise iki birey arasında kaç defa farklı alellerin kalıtımının yapıldığını göstermektedir.

p_i her belirteçteki minör alel frekansı, **P** ise her belirtece ait p_i 'nin ilgili sütunun tamamına yazılmasıyla elde edilen matrisi ifade etmektedir. **M** matrisinin **P** matrisinden çıkarılması, **u**'nun beklenen değerini tayin etmek için gerekli olan **Z** matrisini verir.

Genomik ilişki matrisi $G = ZZ' / \left[2 \sum p_i(1 - p_i) \right]$ formülü ile hesaplanmıştır.

$2 \sum p_i(1 - p_i)$ ile bölüm genomik ilişki matrisini eklemeli genetik ilişki matrisine benzer şekilde oranlamaktadır (VanRaden, 2007; 2008).

4.2.2. Çapraz doğrulama (Cross-validation)

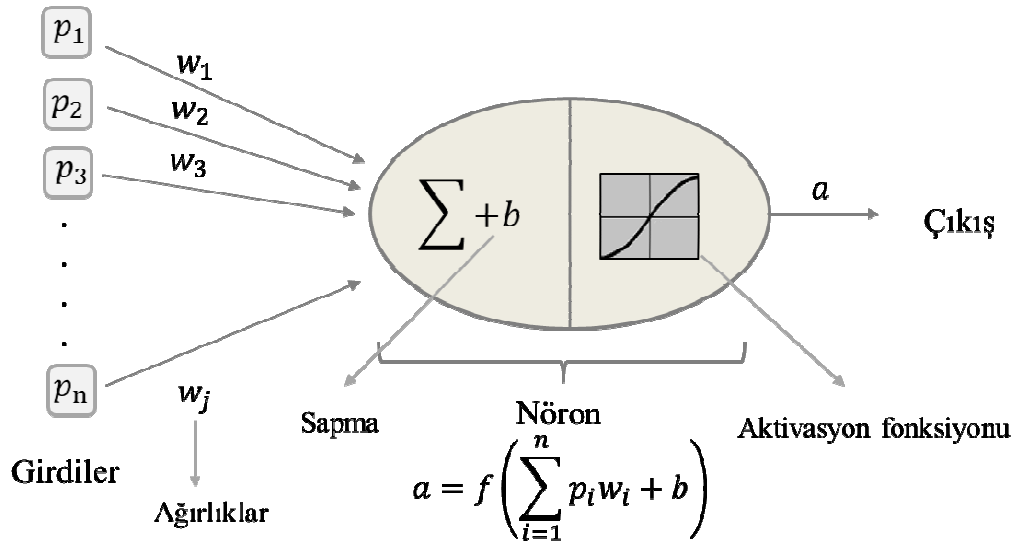
Bitki ıslahında modele dahil olmayan hatlara ait ileriye dönük tahmin performansının tespiti önemli bir problemdir (Crossa ve ark., 2010). Çapraz doğrulama istatistiksel analiz sonuçlarının bağımsız veri setinde değerlendirildiği bir yöntemdir. Genellikle kestirimci modelin performansını tahmin etmede kullanılmaktadır. Bu yöntem doğrultusunda, veri seti %90, %70 ve %50'si öğrenim veri seti kalanı test veri seti olmak üzere iki gruba ayrılmıştır. Bu işlem her oran için 10 defa tekrar edilmiştir. İlk olarak kullanılan ağ mimarilerine öğrenim veri seti tanıtılarak ve sinir ağına öğrenim yapılmış, ardından test veri setinde geleceğe yönelik tahmin parametreleri elde edilmiştir.

4.2.3. Yapay sinir ağıları

İleri beslemeli ağlar

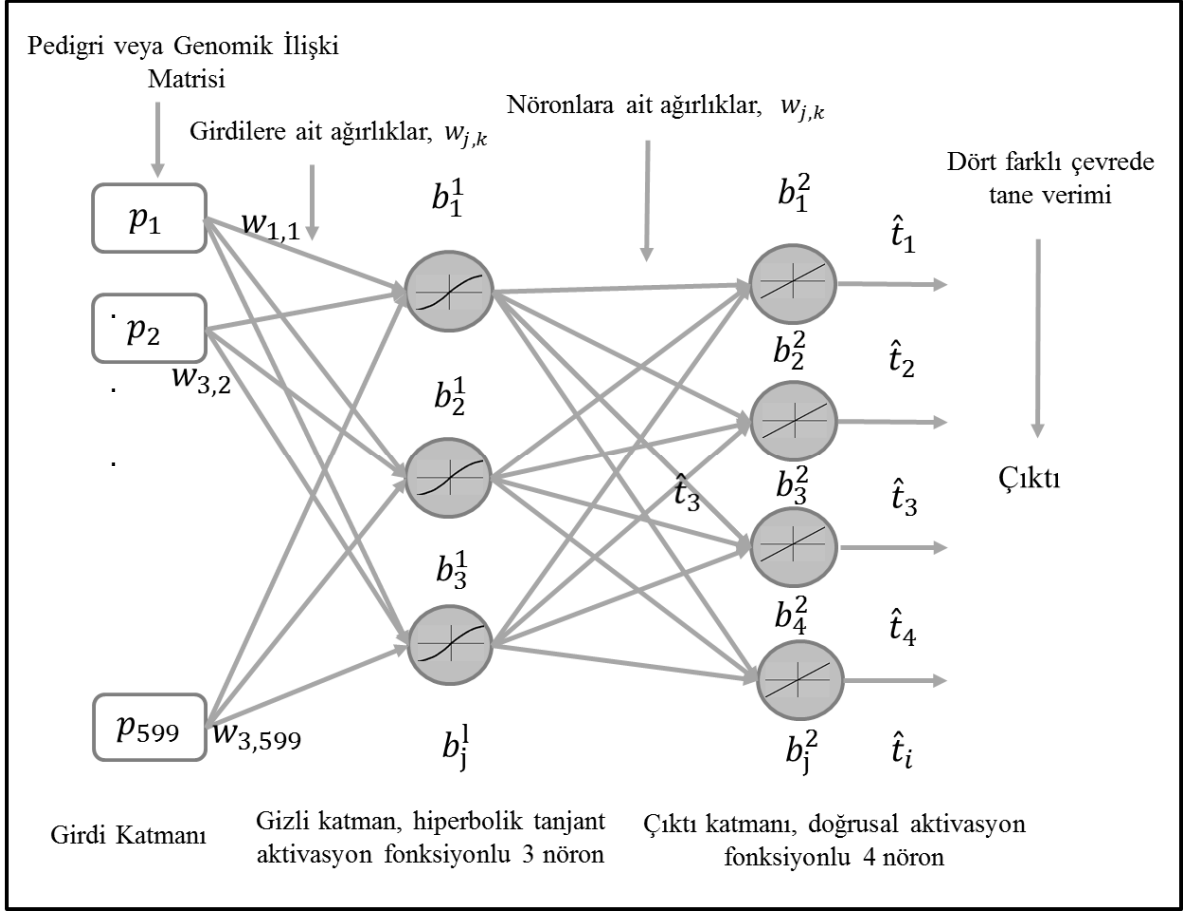
Yapay nöron

Bir YSA modelinin temel elemanı ve işlem birimidir (Şekil 4.1)



Şekil 4.1. Yapay nöronun şematik gösterimi.

Şekilde 4.2’de kullanılan 3 katmanlı ileri beslemeli sinir ağı gösterilmiştir. İlk katman 599 tane buğday hattına ait pedigree ve genomik ilişki matrisi bulunan girdi katmanıdır. Ortadaki gizli katmanda farklı sayılarda nöronlar kullanılmıştır, fakat örnek şekilde 3 nöronlu mimari gösterilmiştir. En sağdaki katmanda, bağımlı değişkenlerin bulunduğu dört çıktı nöronu bulunmaktadır. Her girdi, ağıdaki her bir nörona bilinmeyen ağırlık parametreleri, w , katsayıları ile bağlanır. Nöronda toplanan girdiler aktivasyon fonksiyonu ile dönüştürülür. Bu aktive edilmiş net girdiler, çıktı nöronuna başka bilinmeyen w katsayısıyla bağlanır ve aktivasyon fonksiyonu tekrar dönüşüme uğrarlar. Benzer süreç her bir nöron için uygulanır.



Şekil 4.2. İleri beslemeli yapay sinir ağının şematik gösterimi.

Pedigri veya genomik ilişki matrislerinde girdi (p) olarak 599 buğday hattına ait veri bulunmaktadır. Her p_k ($k=1,2,\dots,599$), $w_{j,k}$ (j nöronu, k girdiyi ifade eder.) katsayıları ile 3 gizli nörona bağlıdır. Her nöronun kendine ait bir bias (sapma) parametresi b_j^l (j nöronu, l katmanı ifade eder.) vardır. $\hat{t}_i$, her i 'ye ait tahmin edilen çıktı değeridir (Şekil 4.2).

Cebirsel olarak sürecin gösteriminde; t_i , her bireye ($i = 1,2,\dots,n$) ait gözlenen fenotipik özelliği göstermektedir. p_i , her bir bireyin genetik belirteç bilgilerini ifade etmektedir. Nöron sayısı S olan gizli katmanda k ($k=1,2,\dots, S$) nöron numarasını ifade etmektedir. $w'_k p_i$, k nöronuna özgü bilinmeyen bağlantı parametresini tanımlar. $w'_k p_i$ 'ye bias (b_j^l) eklenir ve doğrusal veya doğrusal olmayan bir aktivasyon fonksiyonu $f(.)$ ile aktive edilmektedir. Sonrasında, k nöronuna ait aktive edilmiş

sonuç çıktı katmanına gönderilir. Her bir nörona ait sonuçların toplamı Eş.4.1'deki gibi yazılabilir;

$$b + cg \left[\sum_{k=1}^S w_k f_k (w'_k p_i) \right] \quad (4.1).$$

Eş. 4.1'de b genel bias (sapmalı) parametresi ve c aktive edilmiş sonuçlara ait regresyonu tanımlamaktadır. $g(.)$ doğrusal veya doğrusal olmayan diğer bir aktivasyon fonksiyonu iken $w_1, w_2, \dots, w_S$ her bir aktive edilmiş sonuca ait regresyondur. Bağımlı değişken (fenotip) ile girdiler arasında bağlantı Eş 4.2'de olduğu gibi yazılabilir (Gianola ve ark., 2011);

$$t_i = b + cg \left[\sum_{k=1}^S w_k f_k (w'_k p_i) \right] + e_i; \quad i = 1, 2, \dots, n \quad (4.2).$$

3.2.4. Bayesyen regülarizasyonu

YSA'nın öğrenim verisine uyumlu bir ağ oluşturmasının yanında ağın oluşumuna dahil olmayan gözlemleri iyi bir şekilde tahmin etme becerisinin de olması beklenmektedir. Bu regülarizasyon adı verilen yöntemle, ağın bağlantı parametrelerine sınırlamalar getirilerek başarılmaktadır. YSA öğrenim sürecinde genellikle hata kareler toplamını, $F = E_D$, minimize etmeyi amaçlamaktadır (Eş. 4.3).

$$F = E_D (D | w, M) = \sum_{i=1}^N (\hat{t}_i - t_i)^2 \quad (4.3)$$

Hata kareler toplamı, $F = E_D$, aynı zamanda amaç fonksiyonu (objective function) olarak isimlendirilir. Bu, Bayesyen regülarizasyon algoritması dışındaki algoritmalar tarafından kullanılan performans kriteridir.

Bayesyen regülarizasyon yönteminde F performans kriteri kullanılmaktadır. F :

$$F = \beta E_D + \alpha E_W \quad (4.4)$$

şeklinde ifade edilir (Hagan ve ark., 1996). Eşitlik 4.4’de Bayesyen regülarizasyon algoritmasında E_D yanında bir de E_W performans kriteri olarak dahil edilmektedir. Burada E_W , mimarideki parametre kareler toplamını göstermektedir. F amaç fonksiyonunda yer alan α ve β hiper parametreler olarak bilinmektedir. $\alpha \ll \beta$ ise, öğrenim algoritması hatayı küçülmeye zorlamaktadır, eğer $\alpha \gg \beta$ ise öğrenim algoritması ağırlık parametrelerinin sonsal (posterior) yoğunlukları sıfır etrafında yoğunlaşmasına neden olmaktadır: Bunun sonucu olarak daha az pürüzlü bir ağ cevabı ortaya çıkmaktadır. Bayesyen regülarizasyon algoritmasında αE_W ağırlık azalma (weight decay) olarak bilinmekte ve bu terim penaltı fonksiyonu şeklinde görev yaparak ağırlıkları (w) sıfır etrafında sınırlamaya çalışmaktadır. Bu işlem klasik regresyon modelinde shrinkage (daraltıcı) olarak bilinir. Bayesyen regülarizasyon yöntemi böylece girdi (bağımsız değişken) ve çıktı (tahmin) arasındaki ilişkiyi daha az parametre ile tahmin eder ve aşırı uyumu ortadan kaldırır. Bayesyen regülarizasyonu sonucunda tahmin için kullanılan bu azaltılmış parametre sayısına teknik olarak “efektif parametre sayısı” terimi kullanılmaktadır (MacKay, 1991).

Ağın öğrenim için kullandığı veri seti ile uyumu ve tahminleme yeteneği arasındaki uyumun elde edilmesinin doğal bir yolu Bayes yöntemlerinden yararlanmaktır. Ağın parametreleri (M) ile çıktıların yoğunluk fonksiyonun birbirinden bağımsız olduğu varsayılır. D veri setini göstermek üzere, olabilirlik fonksiyonu;

$$P(D | b, w, \sigma^2, M) = \prod_{i=1}^n N(t_i | b, w, \sigma^2, M) \quad (4.5)$$

biçiminde yazılabilir. Eş. 4.5’de $N(.)$ normal olabilirlik fonksiyonunu; b sapma parametresini; w tüm bağlantı parametreleri; σ^2 hata varyansını ve M ise ağ mimarisini ifade eder. YSA’da mimari uygulanan aktivasyon fonksiyonları, katman sayısı ve uygulanan algoritmaların tamamını ifade eder. Elde edilen tahmin için Eş 4.6 elde edilir:

$$b + cg \left[\sum_{k=1}^S w_k g_k \left(b_k + \sum_{j=1}^n a_{kj} u_j^{**[k]} \right) \right], \quad i = 1, 2, \dots, n \quad (4.6).$$

b parametresindeki sapmalı sorunu, ortalamadan sapmaları çıkararak veya düzgün (flat) prior (önsel) değer atanarak ortadan kaldırılabilir. Bayesyen yaklaşımında, regülerize edilmiş sinir ağları kullanıldığında her bağlantı parametresi için aynı normal ilksel dağılım atanmaktadır. Bağlantı parametrelerinin sonsal olasılık yoğunluk fonksiyonu aşağıdaki gibi yazılabilir.

$$P(w | D, \alpha, \beta, M) = \frac{P(D | w, \beta, M) P(w | \alpha, M)}{P(D | \alpha, \beta, M)} \quad (4.7)$$

Eş. 4.7’de , $P(w | \alpha, M)$ ifadesi ağırlık parametreleri bilgisini veren ilksel yoğunluğu göstermektedir. $P(D | w, \beta, M)$ ise ağırlık parametreleri w , β ve ağ mimarisi, M , bilindiğinde veri için olabilirlik yoğunluğunu ifade etmektedir. $P(D | \beta, \alpha, M)$ ise toplam olasılığın 1 olmasını sağlayan normalizasyon faktörüdür.

Öğrenim veri setindenki hatanın bağımsız ve Gauss dağılımına sahip olduğunu varsayılırsa, ağırlık parametrelerinin ilksel dağılımı Gauss olmaktadır. Gauss ilksel olasılık yoğunluk fonksiyonu;

$$P(D | w, \beta, M) = \frac{1}{Z_D(\beta)} \exp(-\beta E_D) \quad (4.8)$$

şeklinde yazılabilirken, $\beta = 1/(2\sigma^2)$, şeklinde yazılabilmektedir. Bu durumda

$$Z_D(\beta) = (2\pi\sigma^2) = \left(\frac{\pi}{\beta}\right)^{N/2} \text{ şeklinde yazılabilmektedir.}$$

Eş. 4.7’nin $P(w | \alpha, M)$ kısmına ilksel olabilirlik denmektedir. Bu ağırlık parametreleri hakkındaki ön bilgiyi ifade eder. Bayes istatistiği ilksel yoğunluk aracılığıyla ilksel bilginin ilişkilendirilmesine olanak tanımaktadır. Ağırlıkların sıfır etrafında yoğunlaştığı varsayılırsa, ağırlık parametrelerine ait sıfır ortalamaya sahip Gauss ilksel olasılık yoğunluk fonksiyonu;

$$P(w | \alpha, M) = \frac{1}{Z_w(\alpha)} \exp(-\alpha E_w) \quad (4.9)$$

şeklinde yazılabilir. E_w ağırlık kareler toplamını simgeler ve $\alpha = 1/(2\sigma_w^2)$ şeklinde gösterilebilir. Bu durumda $Z_w(\alpha) = (2\pi\sigma_w^2)^{n/2} = \left(\pi/\alpha\right)^{n/2}$ şeklinde yazılabilir. n ağıdaki ağırlık ve sapmaların toplam sayısını ifade etmektedir.

Eş. 4.7’de $P(D | \beta, \alpha, M)$ kısmına teknik olarak kanıt (evidence) denmektedir. Bu ağırlık parametreleriyle w , bağlantılı bir fonksiyonu değil, normalizasyon terimidir. Amaç sonsal olabilirliği $P(w | D, \alpha, \beta, M)$, maksimize edecek ağırlık parametrelerini bulmak ise $P(D | \beta, \alpha, M)$ üzerinde durmaya gerek yoktur. Fakat, $P(D | \beta, \alpha, M)$ hiper parametrelerin α ve β , tahmininde önemlidir.

Gauss varsayımları Eş. 4.7’de yerine yazılırsa ağırlık parametreleri w , sonsal olasılık yoğunluk fonksiyonu aşağıdaki gibi yazılabilir.

$$P(w | D, \alpha, \beta, M) = \frac{\frac{1}{Z_w(\alpha)} \frac{1}{Z_D(\beta)} \exp(-(\beta E_D + \alpha E_w))}{\text{Normalizasyon Faktörü}} = \frac{\frac{1}{Z_F(\alpha, \beta)} \exp(-F(w))}{\text{Normalizasyon Faktörü}} \quad (4.10)$$

Bayesyen yaklaşımında optimum ağırlıkların sonsal olasılığı maksimize etmesi beklenir. Sonsal olasılığı maksimize etmek aynı zamanda regülerize amaç fonksiyonunu $F = \beta E_D + \alpha E_w$ minimize etmektedir (Haykin ve ark., 1994; Hagan ve ark., 1996).

α ve β parametrelerinin optimizasyonu

Amaç fonksiyonu parametreleri, α ve β ’yı Bayes kuralıyla optimize edilirken kullanılacak $P(\alpha, \beta | D, M)$ ’ye ait olasılık yoğunluk fonksiyonu;

$$P(\alpha, \beta | D, M) = \frac{P(D | \alpha, \beta, M) P(\alpha, \beta | M)}{P(D | M)} \quad (4.11)$$

şeklinde yazılabilir. Eğer $P(\alpha, \beta | M)$ ilksel yoğunluğunun değişmez olduğu varsayılırsa olabilirlik fonksiyonunu $P(D | \alpha, \beta, M)$, maksimize etmek aynı zamanda sonsal olasılık fonksiyonunu maksimize etmektedir. Bu eşitlikteki olabilirlik fonksiyonu aynı zamanda Eş. 4.7’de normalizasyon faktörüdür. Bu bilgilere dayanarak Eş. 4.7 normalizasyon faktörü için Eş. 4.12’deki gibi çözülebilmektedir (Bishop 1995):

$$\begin{aligned}
 P(D | \alpha, \beta, M) &= \frac{P(D | w, \beta, M) P(w | \alpha, M)}{P(w | D, \alpha, \beta, M)} \\
 &= \frac{\left[\frac{1}{Z_D(\beta)} \exp(-\beta E_D) \right] \left[\frac{1}{Z_w(\alpha)} \exp(-\alpha E_w) \right]}{\frac{1}{Z_F(\alpha, \beta)} \exp(-F(w))} \\
 &= \frac{Z_F(\alpha, \beta)}{Z_D(\beta) Z_w(\alpha)} \times \frac{\exp(-\beta E_D - \alpha E_w)}{\exp(-F(w))} = \frac{Z_F(\alpha, \beta)}{Z_D(\beta) Z_w(\alpha)} \\
 &= \frac{Z_F(\alpha, \beta)}{\left(\frac{2\pi}{\beta} \right)^{N/2} \left(\frac{2\pi}{\alpha} \right)^{m/2}} \tag{4.12}
 \end{aligned}$$

$Z_D(\beta) Z_w(\alpha) \cdot Z_F(\alpha, \beta)$ Taylor serisi açılımı kullanılarak tahmin edilebilmektedir. Amaç fonksiyonu minimum noktasında, küçük bir alan etrafında kuadratik bir yapıya sahip olduğu için ikinci dereceden Taylor serisi uygulanarak $F(w)$ sonsal yoğunluğunun minimum noktasının, w^{MAP} , civarında açılımı yapılabilir:

$$F(w) \approx F(w^{\text{MAP}}) + \frac{1}{2} (w - w^{\text{MAP}})^T \mathbf{H}^{\text{MAP}} (w - w^{\text{MAP}}) \tag{4.13}.$$

$\mathbf{H}^{\text{MAP}}$ ise w^{MAP} dikkate alınarak amaç fonksiyonu için elde edilmiş olan Hessian matrisidir. Bu yaklaşım ile sonsal yoğunluk ifadesi Eş. 4.14’te olduğu gibi yazılabilir.

$$P(w|D, \alpha, \beta, M) \approx \frac{1}{Z_F} \exp \left[-F(w^{\text{MAP}}) - \frac{1}{2} (w - w^{\text{MAP}})^T \mathbf{H}^{\text{MAP}} (w - w^{\text{MAP}}) \right] \quad (4.14)$$

Gauss yoğunluğu standart formu aşağıdaki gibi gösterilebilmektedir.

$$P(w) = \frac{1}{\sqrt{(2\pi)^n |(\mathbf{H}^{\text{MAP}})^{-1}|}} \exp \left(-\frac{1}{2} (w - w^{\text{MAP}})^T \mathbf{H}^{\text{MAP}} (w - w^{\text{MAP}}) \right) \quad (4.15)$$

Eş. 4.15, Eş. 4.14 ile eşitlenirse $Z_F(\alpha, \beta)$ 'nin çözümü yapılabilir:

$$Z_F(\alpha, \beta) \left| \mathbf{H}^{\text{MAP}} \right|^{\frac{1}{2}} \exp \left(-F(w^{\text{MAP}}) \right) \quad (4.16).$$

Bu sonuç, Eş. 4.12'de yerine yazılırsa α ve β 'nin optimum değerleri minimum noktasında hesaplanabilir. Bu, Eş. 4.12'de her tarafın log'unun türevi alınıp sıfıra eşitlenerek gerçekleştirilebilir (Eş. 4.17).

$$\begin{aligned} \log P(D|\alpha, \beta, M) &= \log(Z_F) - \log(Z_D(\beta)) - \log(Z_w(\alpha)) \\ &= \frac{n}{2} \log(2\pi) - \frac{1}{2} \log \det(\mathbf{H}^{\text{MAP}}) - F(w^{\text{MAP}}) - \frac{N}{2} \log\left(\frac{\pi}{\beta}\right) - \frac{n}{2} \log\left(\frac{\pi}{\alpha}\right) \\ &= -F(w^{\text{MAP}}) - \frac{1}{2} \log \det(\mathbf{H}^{\text{MAP}}) + \frac{N}{2} \log(\beta) + \frac{n}{2} \log(\alpha) + \frac{n}{2} \log 2 - \frac{N}{2} \log \pi \end{aligned} \quad (4.17)$$

Amaç fonksiyonunun, F , Hessian matrisi, $\mathbf{H} = \nabla^2 F = \nabla^2(\beta E_D) + \nabla^2(\beta E_w) = \beta \mathbf{B} + 2\alpha \mathbf{I}$ şeklinde yazılabilir ve burada $\mathbf{B} = \nabla^2 E_D$ 'dir. Eğer $\mathbf{H}$ 'nin özdeğeri λ^h ve $\beta \mathbf{B}$ 'nin özdeğeri λ^b ise, $\lambda^h = \lambda^b + 2\alpha$ şeklinde yazılabilir. Ardından, denklemin ikinci teriminin α 'ya göre türevi alınabilir. Bir matrisin determinanı o matrisin özdeğerlerinin bir ürünü olduğundan, Hessian matrisinin tersinin izi, $\text{tr}(\mathbf{H}^{-1})$, Eş. 4.18'deki gibi hesaplanabilmektedir (Haykin, 1994):

$$\begin{aligned}
\frac{\partial}{\partial \alpha} \log \det \mathbf{H} &= \frac{1}{2 \det \mathbf{H}} \frac{\partial}{\partial \alpha} \left(\prod_{k=1}^n \lambda_k^h \right) = \frac{1}{2 \det \mathbf{H}} \left[\sum_{i=1}^n \left(\prod_{j \neq i} (\lambda_j^b + 2\alpha) \right) \frac{\partial}{\partial \alpha} (\lambda_i^b + 2\alpha) \right] \\
&= \frac{\sum_{i=1}^n \left(\prod_{j \neq i} (\lambda_j^b + 2\alpha) \right)}{\prod_{i=1}^n (\lambda_i^b + 2\alpha)} = \sum_{i=1}^n \frac{1}{\lambda_i^b + 2\alpha} = \text{tr}(\mathbf{H}^{-1}) \quad (4.18).
\end{aligned}$$

Aynı terimin türevi bu sefer β 'ya göre alınıp ve efektif parametre sayısını gösteren γ parametresi aşağıdaki gibi tanımlanarak sonraki adım için açılımı yapılabilir.

$$\begin{aligned}
\gamma &\equiv n - 2\alpha \text{tr}(\mathbf{H}^{-1}) \\
&= n - 2\alpha \sum_{i=1}^n \frac{1}{\lambda_i^b + 2\alpha} = \sum_{i=1}^n \left(1 - \frac{2\alpha}{\lambda_i^b + 2\alpha} \right) = \sum_{i=1}^n \left(\frac{\lambda_i^b}{\lambda_i^b + 2\alpha} \right) = \sum_{i=1}^n \frac{\lambda_i^b}{\lambda_i^h} \quad (4.19)
\end{aligned}$$

Bu işlemin ardından, $\frac{1}{2} \log \det(\mathbf{H}^{\text{MAP}})$ 'ın β 'ya göre türevi alınırsa;

$$\begin{aligned}
\frac{\partial}{\partial \beta} \frac{1}{2} \log \det \mathbf{H} &= \frac{1}{2 \det \mathbf{H}} \frac{\partial}{\partial \beta} \left(\prod_{k=1}^n \lambda_k^h \right) \\
&= \frac{1}{2 \det \mathbf{H}} \frac{\partial}{\partial \beta} \left(\prod_{i=1}^n (\lambda_i^b + 2\alpha) \right) \\
&= \frac{1}{2 \det \mathbf{H}} \left[\sum_{i=1}^n \left(\prod_{j \neq i} (\lambda_j^b + 2\alpha) \right) \frac{\partial}{\partial \beta} (\lambda_i^b + 2\alpha) \right] \\
&= \frac{1}{2} \frac{\sum_{i=1}^n \left(\prod_{j \neq i} (\lambda_j^b + 2\alpha) \left(\frac{\lambda_i^b}{\beta} \right) \right)}{\prod_{i=1}^n (\lambda_i^b + 2\alpha)} = \frac{1}{2\beta} \sum_{i=1}^n \frac{\lambda_i^b}{\lambda_i^b + 2\alpha} = \frac{\gamma}{2\beta} \quad (4.20)
\end{aligned}$$

sonucu elde edilmektedir.

Böylece $\log P(D | \alpha, \beta, M)$ 'nin her terim için türevini alarak sıfıra eşitlenebilir. Öncelikle, α için türevi alınarak Eş 4.21 elde edilebilmektedir.

$$\begin{aligned}
\frac{\partial}{\partial \alpha} \log P(D | \alpha, \beta, M) &= \frac{\partial}{\partial \alpha} F(\mathbf{w}^{\text{MAP}}) - \frac{\partial}{\partial \alpha} \frac{1}{2} \log \det(\mathbf{H}^{\text{MAP}}) + \frac{\partial}{\partial \alpha} \frac{n}{2} \log \alpha \\
&= \frac{\partial}{\partial \alpha} (\alpha E_w(\mathbf{w}^{\text{MAP}})) - \text{tr}(\mathbf{H}^{\text{MAP}})^{-1} + \frac{n}{2\alpha^{\text{MAP}}} \\
&= -E_w(\mathbf{w}^{\text{MAP}}) - \text{tr}(\mathbf{H}^{\text{MAP}})^{-1} + \frac{n}{2\alpha^{\text{MAP}}} = 0
\end{aligned} \tag{4.21}$$

Terimler tekrar düzenlenir ve efektif parametre sayısını, γ , kullanılırsa;

$$\begin{aligned}
E_w(\mathbf{w}^{\text{MAP}}) &= \frac{n}{2\alpha^{\text{MAP}}} - \text{tr}(\mathbf{H}^{\text{MAP}})^{-1} \\
2\alpha^{\text{MAP}} E_w(\mathbf{w}^{\text{MAP}}) &= n - 2\alpha^{\text{MAP}} \text{tr}(\mathbf{H}^{\text{MAP}})^{-1} = \gamma \\
\alpha^{\text{MAP}} &= \frac{\gamma}{2E_w(\mathbf{w}^{\text{MAP}})}
\end{aligned} \tag{4.22}$$

α^{MAP} parametresini elde edilmiş olur.

Aynı yol β için tekrar edilirse Eş. 4.23 elde edilir.

$$\begin{aligned}
\frac{\partial}{\partial \beta} \log P(D | \alpha, \beta, M) &= \frac{\partial}{\partial \beta} F(\mathbf{w}^{\text{MAP}}) - \frac{\partial}{\partial \beta} \frac{1}{2} \log \det(\mathbf{H}^{\text{MAP}}) + \frac{\partial}{\partial \beta} \frac{N}{2} \log \beta \\
&= \frac{\partial}{\partial \beta} (\beta E_D(\mathbf{w}^{\text{MAP}})) - \frac{\gamma}{2\beta^{\text{MAP}}} + \frac{N}{2\beta^{\text{MAP}}} \\
&= -E_D(\mathbf{w}^{\text{MAP}}) - \frac{\gamma}{2\beta^{\text{MAP}}} + \frac{N}{2\beta^{\text{MAP}}} = 0
\end{aligned} \tag{4.23}$$

Terimler tekrar düzenlenirse;

$$E_D(\mathbf{w}^{\text{MAP}}) = \frac{N}{2\beta^{\text{MAP}}} - \frac{\gamma}{2\beta^{\text{MAP}}}$$

$$\beta^{\text{MAP}} = \frac{N - \gamma}{2E_D(\mathbf{w}^{\text{MAP}})}$$

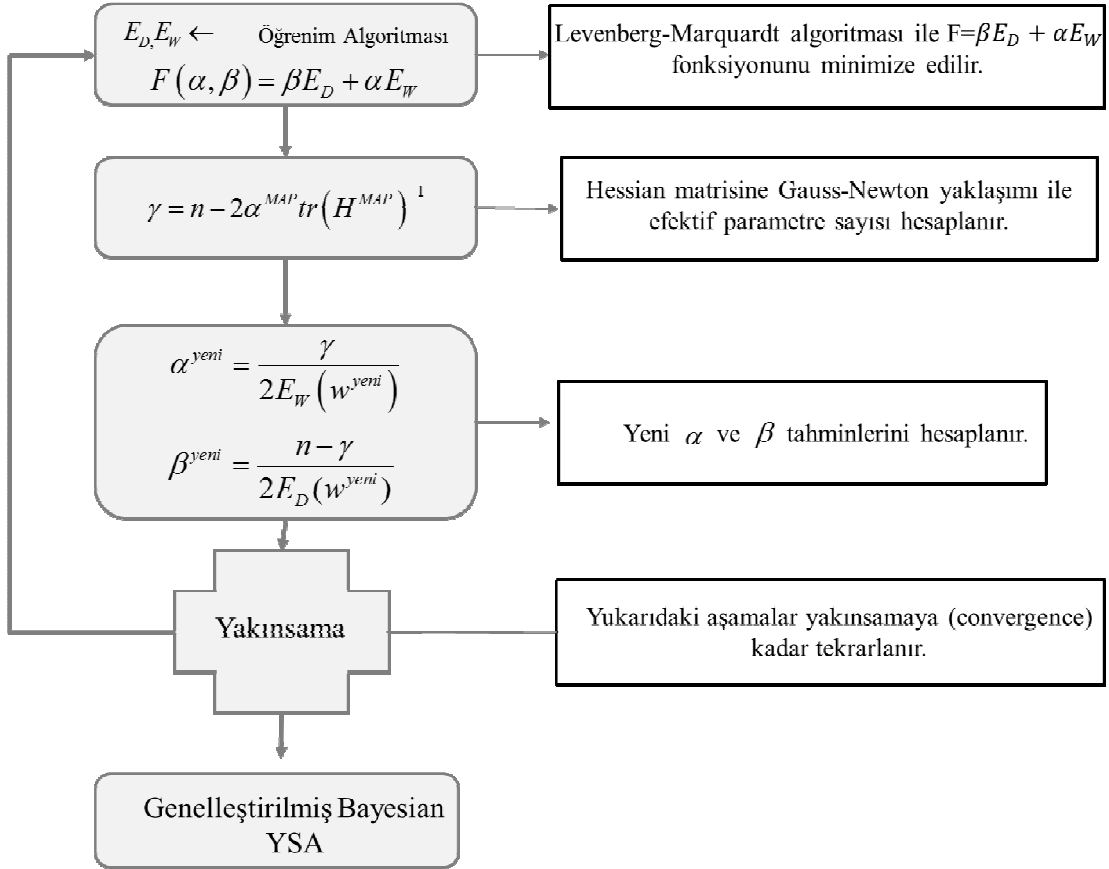
β^{MAP} parametresi elde edilebilmektedir.

Sonuç olarak edilen iki parametrenin eşitlikleri aşağıdaki gibidir (Haykin, 1994; Bishop, 1995; Hagan ve ark., 1996).

$$\alpha^{\text{MAP}} = \frac{\gamma}{2E_w(\mathbf{w}^{\text{MAP}})} \text{ ve } \beta^{\text{MAP}} = \frac{N - \gamma}{2E_D(\mathbf{w}^{\text{MAP}})} \quad (4.25; 4.26)$$

Efektif parametre sayısının eşitliğinde, $\gamma = n - 2\alpha^{\text{MAP}} \text{tr}(\mathbf{H}^{\text{MAP}})^{-1}$, γ , yapay sinir ağı mimarisinde kaç tane parametrenin hatayı düşürmede efektif rol oynadığı göstermektedir. Burada n toplam parametre sayısını ifade eder ve γ sıfır ile n arasında herhangi bir değer alabilmektedir

Bayesyen optimizasyonu ile parametrelerin regülerizasyonu başlıca şu adımlardan oluşur: 1) Başlangıçta α , β ve ağırlık parametreleri atanır. 2) Amaç fonksiyonunu minimize etmek için Levenberg-Marquardt algoritmasının bir adımı kullanılmaktadır. 3) Levenberg-Marquardt algoritmasında Hessian matrisine Gauss-Newton yaklaşımı uygulanarak efektif parametre sayısı hesaplanabilir. 4) Yeni α ve β parametreleri hesaplanır. 5) Yakınsama sağlanana kadar 2-4 adımları tekrarlanmaktadır (Şekil 4.3).



Şekil 4.3. Yapay sinir ağları Bayesyen regülarizasyonu akış şeması (Okut ve ark., 2011).

5. BULGULAR

5.1. Komplekslik Derecesine İlişkin Sonuçlar

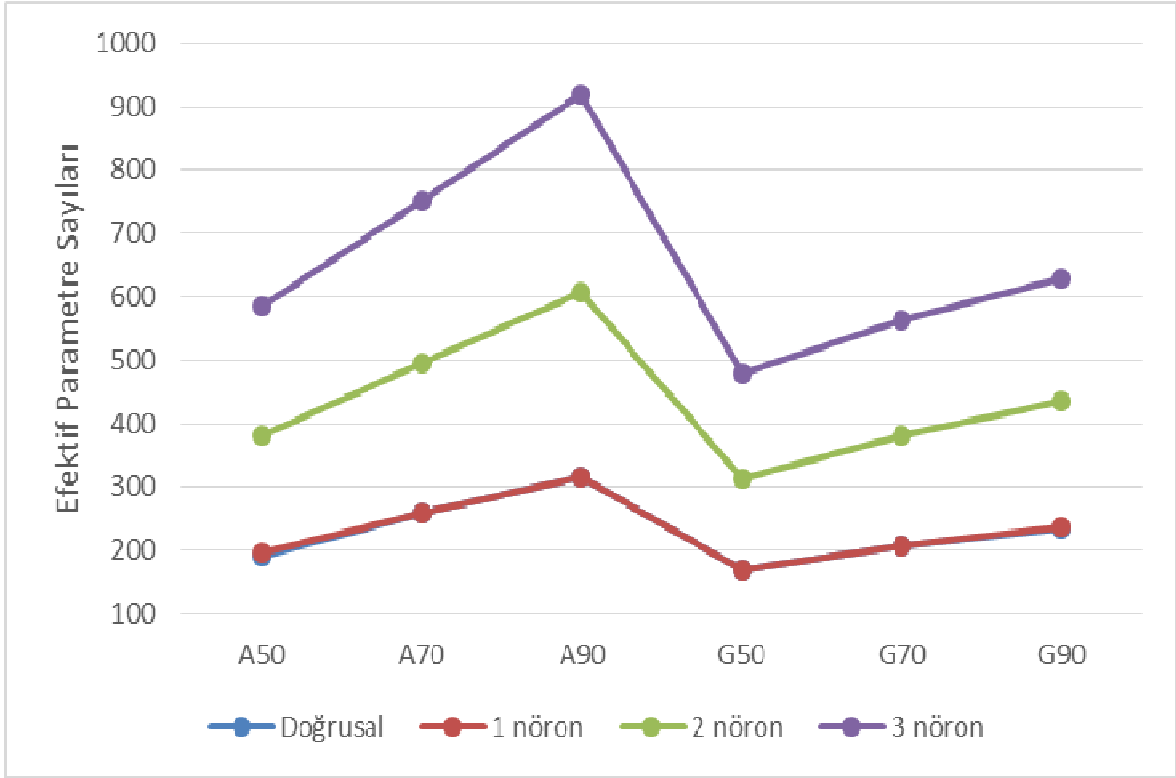
Bayesyen yapay sinir ağları, öğrenim sırasında ağırlık parametrelerini cezalandırarak (penalize) ağın komplekslik derecesinin azalmasını sağlar. Bu ağın, öğrenim veri seti aşırı uyumunu engellemektedir. Çizelge 5.1’de ortalama efektif parametre sayıları verilmiştir.

Çizelge 5.1. Efektif parametre sayıları tablosu ($\pm$ standart hata)

Ağ Mimarisi	A50	A70	A90	G50	G70	G90
Doğrusal	191.4 $\pm$ 2.7	258.8 $\pm$ 2.4	315.9 $\pm$ 3.4	169.1 $\pm$ 3.5	208.2 $\pm$ 4.4	234.3 $\pm$ 3.5
1 nöron	196.0 $\pm$ 2.5	259.4 $\pm$ 2.0	316.3 $\pm$ 2.1	170.0 $\pm$ 3.2	208.2 $\pm$ 3.8	237.4 $\pm$ 1.9
2 nöron	380.4 $\pm$ 7.5	494.4 $\pm$ 4.9	606.4 $\pm$ 3.7	312.4 $\pm$ 3.8	382.1 $\pm$ 7.3	436.2 $\pm$ 3.0
3 nöron	584.1 $\pm$ 11.8	750.2 $\pm$ 6.6	920.5 $\pm$ 2.8	477.7 $\pm$ 6.5	563.6 $\pm$ 8.9	627.3 $\pm$ 5.3

Efektif parametre sayısı (γ), pedigrinin kullanıldığı tüm ağ mimarilerinde genomik bilginin kullanıldığı ağlara göre bariz bir şekilde daha yüksek çıkmıştır. Çizelge 5.1’de görüldüğü gibi pedigrini kullanılan mimarilerde ortalama γ 191.4 ile 920.5 arasında değişirken, moleküler belirteçlerin kullanıldığı ağlarda 169.1 ile 627 arasında değişmektedir. Moleküler bilgilerin kullanıldığı ağlarda efektif parametre sayısının daha az olması shrinkage ile modelin daha fazla genelleştirilebildiği bunun da model ile öğrenim veri setinin aşırı uyumundan kaçınmayı sağladığı söylenebilir.

Çizelge 5.1’de görüldüğü üzere hem pedigrini hem de genomik bilginin kullanıldığı YSA mimarilerinde öğrenim setindeki birey sayısı arttıkça efektif parametre sayısı da artmaktadır. En düşük efektif parametre sayısı 169.1 ile G50 doğrusal ağ yapısından elde edilmiştir. G90 ve doğrusal ağ kesişimde ise bu sayı 234.3’tür. G50’e göre G90 kullanılan mimarilerden ortalama %36, G70 kullanılanlardan ise ortalama %21 daha fazla efektif parametre sayısı elde edilmiştir. Pedigrini de ise, A90 kullanılan mimarilerden A50’den ortalama %60, A70 kullanılanlardan ise ortalama A50’den ortalama %30 daha fazla efektif parametre sayısına elde edilmiştir. Çizelge 5.1’deki efektif parametre sayılarına ait sonuçlar Şekil 5.1’de grafik olarak verilmiştir.



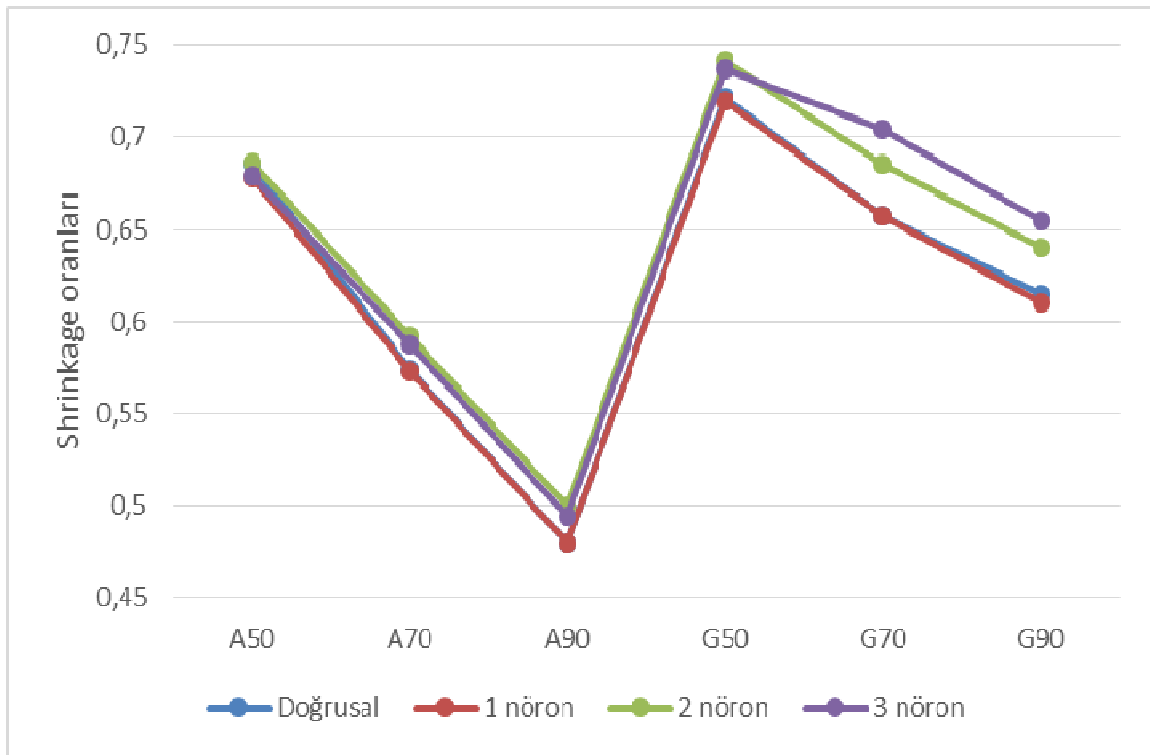
Şekil 5.1. Efektif parametre sayıları grafiği.

En düşük efektif parametre sayısına sahip sinir ağı mimarisi doğrusal yapıda olandır ve gizli katmanında bir nöron bulunan doğrusal olmayan mimari ile aralarında büyük bir fark gözlenmemiştir (Çizelge 5.1). Fakat nöron sayısı arttıkça efektif parametre sayısı net bir şekilde artmaktadır. Bu toplam parametre sayısının artışı ile açıklanabilir. Örneğin gizli katmanında bir nöron bulunan ağ mimarilerinde toplam 608 parametre mevcut iken bu sayı gizli katmanında üç nöron bulunan ağ mimarilerinde 1816'ya çıkmaktadır. Efektif parametre sayısını toplam parametre sayısına bölünür ve sonuç birden çıkarılırsa elde edilen shrinkage oranlarının birbirlerine yakın olduğu görülebilir (Çizelge 5.2).

Çizelge 5.2. Shrinkage oranları tablosu

Ağ Mimarisi	A50	A70	A90	G50	G70	G90
Doğrusal	0.685	0.574	0.480	0.722	0.657	0.615
1 nöron	0.678	0.573	0.480	0.720	0.657	0.610
2 nöron	0.686	0.592	0.500	0.742	0.685	0.640
3 nöron	0.679	0.588	0.494	0.737	0.704	0.655

Çizelge 5.2’de öğrenim veri setindeki birey sayısı artıkça shrinkage oranının azaldığı görülmektedir. Efektif parametre sayısı ile shrinkage oranı arasında ters orantı vardır. Pedigri kullanılan modellerde gizli katmanında iki ve üç nörona sahip olan mimarilerde doğrusal ve doğrusal olmayan mimarilere göre shrinkage oranında bir artış gözlenmektedir. Bu fark genomik bilgi kullanılan mimarilerde daha fazladır.



Şekil 5.2. Shrinkage oranları grafiği.

5.2. Tahmin Performansına İlişkin Sonuçlar

Yapay sinir ağına tanıtılan öğrenim veri setine dayanarak oluşturulan modellerin tahmin performansı kriterleri olarak, test veri setlerinden elde edilen korelasyon katsayıları ve hata kareler ortalamaları kullanılmıştır. Test verisinden elde edilen bu sonuçlar, ağların daha önce karşılaşmadığı verilerde ileriye dönük tahmin performansını değerlendirilebilmesini sağlar. Çizelge 5.3’te öğrenim setlerinde oluşturulan modellere ait test veri setinden elde edilen ortalama korelasyon katsayıları verilmiştir.

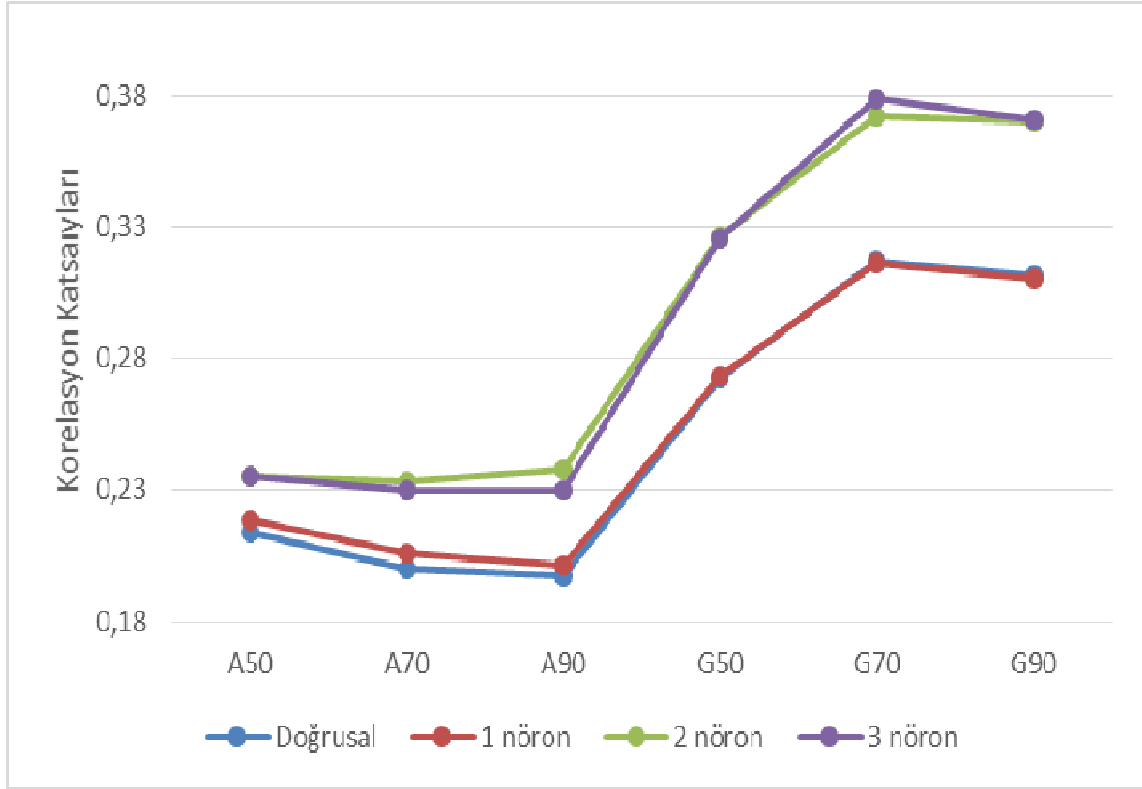
Çizelge 5.3. Test veri setlerinden elde edilen korelasyon katsayıları tablosu ($\pm$ standart hata)

Ağ Mimarisi	A50	A70	A90	G50	G70	G90
Doğrusal	0.22 $\pm$ 0.01	0.21 $\pm$ 0.01	0.20 $\pm$ 0.02	0.27 $\pm$ 0.01	0.32 $\pm$ 0.01	0.31 $\pm$ 0.03
1 nöron	0.21 $\pm$ 0.01	0.20 $\pm$ 0.01	0.19 $\pm$ 0.02	0.27 $\pm$ 0.01	0.32 $\pm$ 0.01	0.31 $\pm$ 0.03
2 nöron	0.24 $\pm$ 0.01	0.23 $\pm$ 0.01	0.24 $\pm$ 0.02	0.33 $\pm$ 0.01	0.37 $\pm$ 0.01	0.37 $\pm$ 0.03
3 nöron	0.24 $\pm$ 0.01	0.23 $\pm$ 0.01	0.23 $\pm$ 0.02	0.33 $\pm$ 0.01	0.38 $\pm$ 0.01	0.37 $\pm$ 0.03

Çizelge 5.3'te sonuçlardan görüldüğü gibi ortalama korelasyon katsayıları 0.19 ile 0.38 arasında değişmektedir. DArT belirteçlerinin kullanıldığı modellerde ortalama korelasyon katsayılarının pedigrinin kullanıldığı modellere göre net bir şekilde daha yüksek olduğu görülmektedir. En düşük ortalama korelasyon katsayısı sonucu 0.19, A90'ın kullanıldığı gizli katmanında doğrusal olmayan bir nöron bulunan ağ mimarisinden, en büyük korelasyon katsayısı ise 0.38 ile G70 ve gizli üç nöron bulunan mimariden elde edilmiştir.

Çizelge 5.3'te ağ mimarileri arasındaki ortalama korelasyon katsayılarına bakıldığında gizli katmanında bir nöron bulunan mimarilerdeki sonuçların iki ve üç nöron bulunanlara göre daha düşük olduğu görülebilmektedir.

Pedigri kullanılan mimarilerde farklı çapraz doğrulama oranlarında ortalama korelasyon katsayılarının daha homojen dağıldığı görülebilir. DArT belirteçlerinin kullanıldığı ağ mimarilerde G50, diğer iki çapraz doğrulama oranının kullanıldığı mimarilere göre daha düşük ortalama korelasyon katsayılarına sahiptir (Çizelge 5.3). Çizelge 5.3'teki korelasyon katsayılarına ait sonuçlar Şekil 5.3'te grafik olarak verilmiştir.



Şekil 5.3. Test veri setlerinden elde edilen korelasyon katsayıları grafiği.

Tahmin performansını değerlendirme kullanılan diğer bir kriter test verisinden elde edilen hata kareler ortalamasıdır. Elde edilen ortalama hata kareler ortalamaları Çizelge 5.4'te verilmiştir.

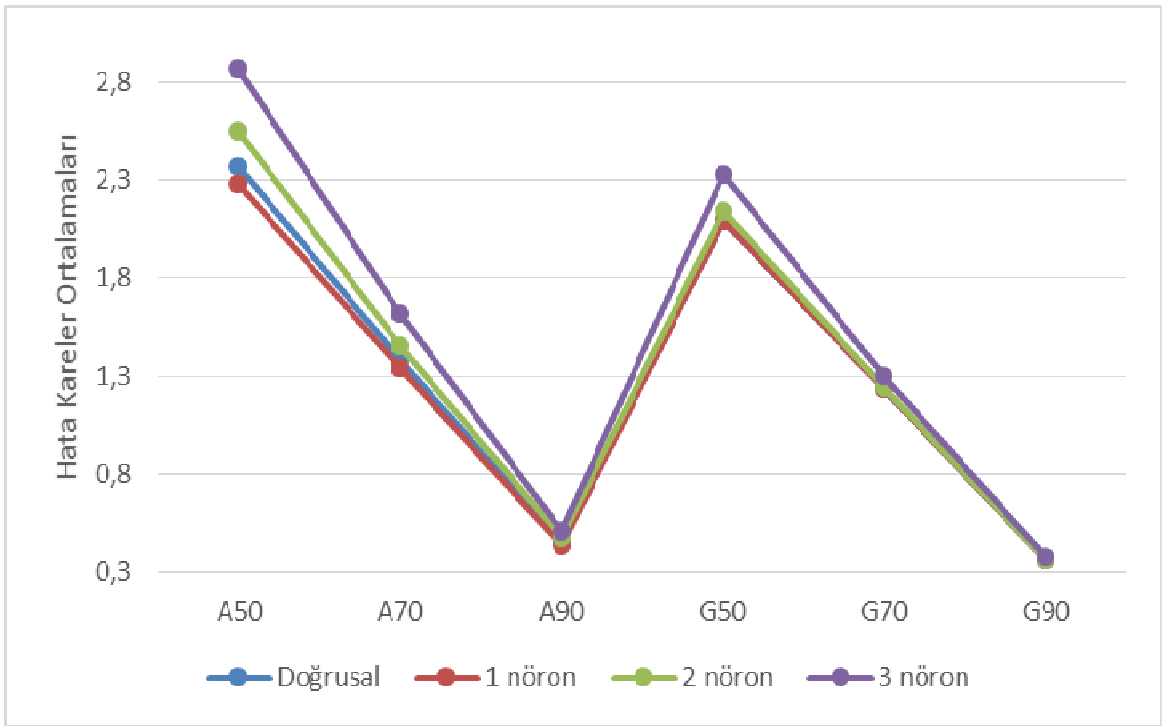
Çizelge 5.4. Test veri setinden elde edilen hata kareler ortalamaları tablosu ($\pm$ standart hata)

Ağ Mimarisi	A50	A70	A90	G50	G70	G90
Doğrusal	2.37 $\pm$ 0.08	1.38 $\pm$ 0.02	0.45 $\pm$ 0.03	2.09 $\pm$ 0.04	1.23 $\pm$ 0.03	0.36 $\pm$ 0.02
1 nöron	2.30 $\pm$ 0.06	1.34 $\pm$ 0.02	0.44 $\pm$ 0.03	2.09 $\pm$ 0.04	1.24 $\pm$ 0.03	0.37 $\pm$ 0.02
2 nöron	2.54 $\pm$ 0.08	1.45 $\pm$ 0.03	0.48 $\pm$ 0.03	2.14 $\pm$ 0.04	1.25 $\pm$ 0.03	0.36 $\pm$ 0.02
3 nöron	2.96 $\pm$ 0.11	1.62 $\pm$ 0.03	0.51 $\pm$ 0.03	2.33 $\pm$ 0.04	1.30 $\pm$ 0.03	0.38 $\pm$ 0.02

Test veri setinden elde hata kareler ortalamaları, Çizelge 5.4'te görüldüğü gibi 0.36 (G90-doğrusal) ile 2.96 (A50-3 nöron) arasında değişmektedir. Pedigri kullanılan ağlarda hata kareler ortalamaları, genomik bilginin kullanıldığı ağlara göre net bir şekilde daha yüksek olduğu görülmektedir. Genomik bilginin kullanıldığı ağlarda korelasyon katsayılarının yüksekliğine ters orantılı bir şekilde hata kareler

ortalamalarının pedigri bilgisinin kullanıldığı ağlara göre daha düşük çıkması beklenen bir sonuçtur.

Çizelge 5.4'te çapraz doğrulama oranlarına bakıldığında test veri seti küçüldükçe hata kareler ortalaması da düşmektedir. Ağ mimarilerine bakıldığında ise nöron sayısı arttıkça hata kareler ortalaması da artmaktadır. Çizelge 5.4'teki hata kareler ortalamalarına ait sonuçlar Şekil 5.4'te grafik olarak verilmiştir.



Şekil 5.4. Test veri setlerinde elde edilen hata kareler ortalamaları grafiği.

5. TARTIŞMA VE SONUÇ

Bu çalışmada öncelikli olarak, ileri beslemeli YSA ile pedigree ve yoğun DArT belirteçlerinin, buğdayda kompleks bir fenotipik özellik olan tane verimini tahmin performansları analiz edilmiştir. Ayrıca, farklı çapraz doğrulama oranları kullanılarak bu oranlar arasındaki farklar değerlendirilmiştir. Yüksek sayıda moleküler bilgi ve kompleks bir fenotipik özellik içeren veri setinde, efektif parametre sayısını azaltarak modeli genelleştirmek ve aşırı uyumdan kaçınmak için Bayesyen YSA kullanılmıştır (Foresee ve Hagan 1997; Demuth ve ark., 2009; Okut ve ark., 2011). Gizli katmanda farklı nöron sayıları ve bu nöronlarda farklı aktivasyon fonksiyonlarının (doğrusal ve hiperbolik tanjant aktivasyon fonksiyonu) kullanıldığı çeşitli YSA mimarileri denenmiştir (Gianola ve ark., 2011).

Çalışmada, YSA'ların tahmin performanslarının değerlendirilmesinde efektif parametre sayısı, korelasyon katsayısı ve hata kareler ortalaması olmak üzere temel üç parametre kullanılmıştır. Bu parametrelerden efektif parametre sayıları öğrenim veri setinden, korelasyon katsayıları ve hata kareler ortalamaları ise test veri setinden elde edilmiştir. Test veri setinden elde edilen korelasyon katsayıları ve hata kareler ortalamaları ağa daha önce tanıtılmamış verilerdeki tahmin performansını gösteren parametrelerdir. Bu nedenle, bu parametrelere kestirimci korelasyon sayısı ve kestirimci hata kareler ortalaması da denmektedir.

Efektif parametre sayısı, shrinkage etkisini ve YSA'nın öğrenim veri setindeki bilgilerin oluşturulan modelde ne kadar genelleştirebildiğinin bir göstergesidir (Fernandez ve Cabellero, 2006; Fernandez ve ark., 2011). YSA, Bayesyen yaklaşımı yüksek sayıda parametre veri setlerine bağlı aşırı uyum ve aşırı öğrenim problemlerini azaltabilmektedir. (Sorich 20003). Bunu, ağırlık parametrelerinin penaltı fonksiyonu olan E_w ile sağlamaktadır. Bu penaltı fonksiyonu ile ortak bilgi taşıyan parametreler bir araya getirilerek ve etkisi sifıra yakın parametreler çıkarılarak parametre sayısı azaltılmaktadır. Bu azaltmış parametre sayısına efektif parametre sayısı denmektedir. (MacKay, 1991; Foresee ve Hagan, 1997). Efektif parametre sayısı, pedigree bilgisinin kullanıldığı ağlarda 191 ile 920, belirteç bilgisinin kullanıldığı ağlarda ise 169 ile 627 arasında değişirken toplam parametre sayısı 608 ile 1816 arasında değişmektedir (Çizelge 5.1). Bu, shrinkage etkisi ve yapay sinir ağlarının çok boyutluluk sorunu ile

nasıl başa çıktığı hakkında fikir vermektedir (Gianola ve ark., 2011). Ayrıca, hangi genetik bilginin kullanıldığı fark etmeksizin gizli katmandaki nöron sayıları arttıkça efektif parametre sayısının da arttığı görülmektedir (Çizelge 5.1; Şekil 5.1). Bu, toplam parametre artışının bir sonucu olarak değerlendirilebilir (Winkler ve Burden, 2004). Çizelge 5.1’de farklı çapraz doğrulama oranlarına bakıldığında öğrenim veri setindeki birey sayısı arttıkça efektif parametre sayısı artmaktadır. Bundan yola çıkarak, sinir ağının öğreniminde ağa tanıtılan veri sayısının artmasının ağın komplekslik derecesini arttırdığı söylenebilir.

YSA’nın tahmin performansını değerlendirme önemli bir parametre kestirici korelasyon da denebilen test veri setinden edilen korelasyon katsayılarıdır. Çizelge 5.3’te genomik bilgilerin kullanıldığı mimarilerde, pedigree kullanılanlara göre korelasyon katsayılarının net bir şekilde daha yüksek olduğu görülmektedir. Bu sonuç, aynı veri setinde pedigree ve moleküler belirteçlerin tahmin yeteneklerinin karşılaştırıldığı de los Campos ve ark. (2009) ve Crossa ve ark. (2010) bulgularıyla paralellik göstermektedir. Crossa ve ark. (2010) tarafından yapılan çalışmada RKHS, Bayesyen Lasso ve en iyi doğrusal yansız kestirici (BLUP) yöntemleri karşılaştırılmış genomik veri seti ile pedigree modeline göre RKHS’de ortalama %21.55, Bayesyen Lasso’da %8.5, BLUP’ta %3.1 daha yüksek korelasyon katsayısı elde edilmiştir. Lasso yöntemi, regresyon katsayılarını regresyon katsayılarının mutlak değerleri toplamı (L_1) ile cezalandırır. Lasso, bu cezalandırma fonksiyonu ile değişken seçimi yapmakta ve regresyon katsayılarına shrinkage uygulamaktadır (Tibshirani, 1996). Ayrıca moleküler belirteç kullanılan yöntemler arasında eklemeli bir model olan Bayesyen Lasso’ya göre RKHS’de test veri seti ile daha yüksek korelasyon gözlenmesi, RKHS’nin kompleks ilişkileri açığa çıkarmada daha başarılı olmasına dayandırılabilir. Bu kompleks genetik ilişkileri tanımlamada, RKHS’nin temel bileşeni, kompleks genetik ilişkileri tanımlayabilecek uygun kernel matrisini seçmesidir. Bu sayede, genotip ile fenotip arasında spesifik bir fonksiyonel yapıya gerek duymadan tahmin yapılabilir (Gianola ve Van Kaam, 2008). Benzer şekilde, bu çalışmada ise Çizelge 5.3’te, doğrusal ve doğrusal olmayan tek nörona sahip ağ mimarileri arasında önemli bir fark olmamasından yola çıkarak, doğrusal olmayan genetik ilişkilerin tek nöronun açıklayabileceğinden daha kompleks bir yapıya sahip olduğu sonucu çıkarılabilir. YSA ile doğrusal olmayan ilişkilerin tespitinde en az iki doğrusal olmayan nöron ihtiyacı

benzer çalışmalardaki sonuçları da desteklemektedir (Winkler ve Burden, 2004; Okut ve ark., 2011; Gianola ve ark., 2011). Long ve ark. (2011) çalışmasında ise aynı veri setinde (%80 öğrenim % 20 test veri seti olmak üzere 50 tekrar çapraz doğrulama) DArT belirteçleri ile fenotipik özellik tahmininde, radyal temelli kernel fonksiyonu kullanılan SVR modellerinden doğrusal kernel fonksiyonuna sahip modellere göre daha yüksek korelasyon katsayıları elde edilmiştir. Çapraz doğrulama oranlarına göre korelasyon katsayılarını karşılaştırdığımızda genomik bilgilerin kullanıldığı ağlarda G70'te, G50 ve G90'a göre daha iyi sonuçlar elde edilmesi seçimde G70'i çapraz doğrulama oranı seçiminde öne plana çıkarmıştır.

Çapraz doğrulama ile test veri setinden elde edilen diğer önemli bir parametre olan hata kareler ortalamasıdır. Çizelge 5.4'te Pedigri ve moleküler belirteç genotiplerinin kullanıldığı mimariler karşılaştırıldığında belirteç genotiplerinden, ağ mimarisine ve çapraz doğrulama oranına bakılmaksızın daha küçük hata değerleri elde edilmiştir. Belirteç bilgisi kullanan mimarilerinden daha yüksek korelasyon katsayıları ve daha düşük parametre sayıları elde edildiği göz önüne alınırsa küçük hata kareler ortalaması elde edilmesi beklenen bir sonuçtur. Farklı yapay sinir ağı mimarilerine bakıldığında nöron sayısı arttıkça hata kareler ortalamalarının da arttığı görülmektedir. Nöron sayısı artışına paralel olarak kestirimci korelasyon katsayılarının artmasına rağmen hatanın artmasının nedeni olarak nöron sayısı artışı ile artan efektif parametre sayısı gösterilebilir.

Bu çalışmada, buğdayda kompleks bir özellik olan tane veriminin YSA ile analizinde pedigriye göre daha yüksek test korelasyon katsayıları ve daha küçük efektif parametre sayıları ve hata kareler ortalamaları elde edilen yoğun DArT moleküler belirteçlerinin daha üstün bir tahmin yeteneği gösterdiği söylenebilir. Farklı çapraz doğrulama oranlarının kullanıldığı çalışmada, öğrenim/test verisi oranı arttıkça efektif parametre sayısının arttığı ve hata kareler ortalamasının düştüğü belirlenmiştir. Pedigri verisinin kullanıldığı ağlarda çapraz doğrulama oranının korelasyon katsayısına önemli bir etkisi gözlenmezken, genomik verinin kullanıldığı ağlarda G50'de korelasyon katsayıları diğer iki orana göre net bir şekilde daha düşüktür. Ayrıca sonuçlara bakıldığında, sinir ağı mimarisi seçiminde gizli katmanında tek nöron bulunan ağların fenotipik ve genetik özellikler arasındaki ilişkileri belirlemede yetersiz kaldığı gözlenmiştir.

Sonuç olarak, pedigrî ve fenotipik özelliklerin kaydı tutularak genetik değer ve fenotipik özellik tahmini çok uzun süredir başarılı bir şekilde yapılmaktadır. Fakat bu çalışmada olduğu gibi tüm kromozomu kaplayan yoğun genetik belirteç haritaları ile pedigrîye oranla daha iyi genetik tahmin yapmak mümkün olabilmektedir. Pedigrî ile belirlenemeyen genetik ilişkileri tespit edebilmesi ve uzun yıllar süren kayıtlara ihtiyaç duymaması moleküler belirteç haritalarını ön plana çıkarmaktadır. Teknolojinin gelişmesiyle, erişimi kolaylaşan DArT ve SNP gibi yüksek çıktı veren moleküler tekniklerin genetik değer belirleme ve ıslah çalışmalarında daha yaygın olarak kullanılması kaçınılmazdır. Yapay sinir ağları, bu gibi çok sayıda parametre içeren yoğun belirteç haritaları ile kompleks fenotiplerin analizinde başarılı bir yöntem olarak değerlendirilebilir.

KAYNAKLAR

- Adeli H., Panakkat, A., 2009. Probabilistic neural network for earthquake magnitude prediction. *Neural Networks*, **22**: 1018-1024.
- Akbari, M., Wenzl, P., Caig, V., Carlig, J., Xia, L., Yang, S., Uszynski, G., Mohler, V., Lehmensiek, A., Howes, N., Sharp, P., Huttner, E., Kilian, A., 2006. Diversity arrays technology (DArT) for highthroughput profiling of the hexaploid wheat genome. *Theor. Appl. Genet.*, **113**: 1409-1420.
- Amorim, E. P., Vilarinhos, A. D., Cohen, K. O., Amorim, V. B. O., Santos-Serejo, J. A., Silva, S. O., Pestana, K. N., Santos, V. J., Paes, N. S., Monte, D. C., Rei, R. V., 2009. Genetic diversity of carotenoid-rich bananas evaluated by Diversity Arrays Technology (DArT). *Genet Mol Biol*, **32**: 96-103.
- Badea, A., Eudes, F., Graf, R. J., Laroche, A., Gaudet, D. A., Sadasivaiah, R. S., 2008. Phenotypic and marker-assisted evaluation of spring and winter wheat germplasm for resistance to Fusarium head blight. *Euphytica*, **164**: 803-819.
- Bargen, D., 2013. Programming a perceptron in python. <https://blog.dbrgn.ch/2013/3/26/perceptrons-in-python/> Eriřim tarihi: 04.11.2015.
- Beal, M.H., Hagan, M.T., Demuth, H.B., 2010. *Neural Network Toolbox' 6 User's Guide*. The MathWorks, Inc.
- Bedo, J., Wenzl, P., Kowalczyk, A., Kilian, A., 2008. Precision mapping and statistical validation of quantitative trait loci by machine learning. *BMC Genetics*, **9**: 35-52.
- Bishop C. M., 1995. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford. 482.
- Bohra, A., Dubey, A., Saxena, R. K., Penmetsa, R. V., Poornima, K. N., Kumar, N., Framer, A. D., Srivani, G., Upadhyaya, H. D., Gothwal, R., Ramesh, S., Singh, D., Saxena, K., Kavikishor, P. B., Singh, N. K., Town, C. D., May, G. D., Cook, D. R., Varshney, R. K., 2011. Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea (*Cajanus spp.*). *BMC Plant Biol*, **11**: 56-71.

- Bolibok-Bragoszewska, H., Heller-Uszyńska, K., Wenzl, P., Uszyński, G., Kilian, A., Rakoczy-Trojanowska, M., 2009. DArT markers for the rye genome-genetic diversity and mapping. **BMC Genomics**, **10**: 578-588.
- Botstein, D., White, R., Skolnick, M., Davis R., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **Am. J. Hum. Genet.**, **32**: 314-331.
- Brameier, M., Banzhaf, W., 2001. A comparison of linear genetic programming and neural networks in medical data mining. **IEEE Trans. Evol. Comput.**, **5**: 17-26.
- Buscema, M., Penco, S., Grossi, E., 2012. A novel mathematical approach to define the genes/SNPs conferring risk or protection in sporadic amyotrophic lateral sclerosis based on auto contractive map neural networks and graph theory. **Neurol Res Int.**, **2012**: 13.
- Carpenter, G. A., Grossberg, S., 1988. The ART of adaptive pattern recognition by a self-organizing neural network. **Computer**, **21**: 77-88.
- Chun, S., Kim, J., Hahm, K., Park, Y., Chun S., 2008. Data mining technique for medical informatics: Detecting gastric cancer using case-based reasoning and single nucleotide polymorphisms. **Expert Systems**, **25**: 163-172.
- Crossa, J., Burgueno, J., Cornelius, P. L., McLaren, G., Trethowan R., Krishnamachari A., 2006. Modeling genotype 3 environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. **Crop Sci.**, **46**: 1722-1733.
- Crossa, J., Burgueño, J., Dreisigacker, S., Vargas, M., Herrera-Foessel, S. A., Lillemo, M., Singh, R. P., Trethowan, R., Warburton, M., Franco., J., Reynolds, M., Crouch, J. H., Ortiz, R., 2007. Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. **Genetics**, **177**: 1889-1913.
- Crossa, J., de los Campos, G., Perez, P., Gianola, D., Dreisigacker, S., Burgueño, J., Araus, J. L., Makumb, D., Yan, J., Singh, R., Arief, V., Banzinger, M., Braun, H. J., 2010. Prediction of genetic values for quantitative traits in plant breeding using pedigree and molecular markers. **Genetics**, **186**: 713-724.
- Curtis, D., 2007. Comparison of artificial neural network analysis with other multimarker methods for detecting genetic association. **BMC Genet**, **8**: 49-56.

- de los Campos, G., Gianola, D., Rosa, G. J. M., 2009. Reproducing kernel Hilbert spaces regression: A general framework for genetic evaluation. *J. of Anim. Sci.*, **87**: 1883-1887.
- Demuth, H., Beale, M., Hagan, M., 2009. *Neural Network Toolbox™ 6 User's Guide*. The MathWorks Inc. Natick, Massachusetts, Birleşik Devletler. 410.
- El-Bakry, H. M., 2001. Automatic human face recognition using modular neural networks. *Machine Graphics & Vision Journal*, **10**: 47-73.
- Fernandez, M., Caballero, J., 2006. Ensembles of bayesian-regularized genetic neural networks for modeling of acetylcholinesterase inhibition by huprines. *Chem Biol Drug Des*, **68**: 201-212.
- Fernandez, M., Caballero, J., Fernandez, L., Sarai, A., 2011. Genetic algorithm optimization in drug design QSAR: Bayesian-regularized genetic neural networks (BRGNN) and genetic algorithm-optimized support vectors machines (GA-SVM). *Mol Divers*, **15**: 269-289.
- Francki, M. G., Walker, E., Crawford, A. C., Broughton, S., Ohm, H. W., Barclay, I., Wilson R. W., Mclean, R., 2009. Comparison of genetic and cytogenetic maps of hexaploid wheat (*Triticum aestivum* L.) using SSR and DArT markers. *Mol Genet Genomics*, **281**: 181-191.
- Ficetola, G. F., Miaud, C., Pompanon, F., Taberlet, P., 2008. Species detection using environmental DNA from water samples. *Biology Letters*, **4**: 423-425.
- Foresee, F. D., Hagan, M. T., 1997. Gauss-Newton approximation to Bayesian regularization. *Proceedings of the 1997 International Joint Conference on Neural Networks*. 1977, Birleşik Devletler. 1930-1935.
- Gader, P., Mohamed, M., Chiang, J. H., 1997. Handwritten word recognition with character and inter-character neural networks. *IEEE Trans. Syst. Man. Cybern.*, **27**: 158-164.
- Gianola, D., Fernando R. L., 1986. Bayesian methods in animal breeding theory. *J. Anim. Sci.*, **63**: 217-244.
- Gianola, D., Okut, H., Wiegel, K. A., Rosa, G. J. M., 2011. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. *BMC Genet.*, **12**: 87-100.

- Gianola, D., van Kaam, J. B., 2008. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. ***Genetics*, 178:** 2289-2303.
- Grewal, T. S., Rossnagel, B. G., Pozniak, C. J., Scoles, G. J., 2008. Mapping quantitative trait loci associated with barley net blotch resistance. ***Theor Appl Genet*, 116:** 529-539.
- Griffiths, S., Simmonds, J., Leverington, M., Wang, Y., Fish, L., Sayers, L., Alibert, L., Orford, S., Wingen, L., Herry, L., Faure, S., Laurie, D., Bilham, L., Snape, S., 2009. Meta-QTL analysis of the genetic control of ear emergence in elite European winter wheat germplasm. ***Theoretical and Applied Genetics*, 119:** 383-395.
- Hagan, M., Beale, M., H., Demuth, H., 1996. ***Neural Network Design***. PWS Publishing Company, Boston, MA. 1012.
- Haykin, S., 1994. ***Neural Networks: A Comprehensive Foundation***. Macmillan, New York, NY. 842.
- Hearnden, P. R., Eckermann, P. J., McMichael, G. L., Hayden M. J., Eglinton J. K., Chalmers, K. J., 2007. A genetic map of 1,000 SSR and DArT markers in a wide barley cross. ***Theor Appl Genet*, 115:** 383-39.
- Hebb, D. O., 1949. ***The Organization of Behaviour***. Wiley. New York, Birleşik Devletler. 378.
- Heller-Uszynska, K., Uszynski, G., Huttner, E., Evers, M., Carling, J., Caig, V., Aitken, K., Jackson, P., Piperidis, G., Cox, M., Gilmour, R., D'Hont, A., Butterfield, M., Glaszmann J., Kilian, A., 2010. Diversity Arrays Technology effectively reveals DNA polymorphism in a large and complex genome of sugarcane. ***Mol Breed*, 28:** 37-55.
- Horvath, A., Didier, A., Koenig, J., Exbrayat, F., Charmet, G., Balfourier, F., 2009. Analysis of diversity and linkage disequilibrium along chromosome 3B of bread wheat (*Triticum aestivum* L.). ***Theoretical and Applied Genetics*, 119:** 1523-1537.
- Huynh, B. L., Wallwork, H., Stangoulis, J. C. R., Graham, R. D., Willsmore, K. L., Olson, S., Mather, D. E., 2008. Quantitative trait loci for grain fructan concentration in wheat (*Triticum aestivum* L.). ***Theor Appl Genet*, 117:** 701-709.

- Jaccoud, D., Peng, K., Feinstein, D., Kilian, A., 2001. Diversity arrays: a solid state technology for sequence information independent. *Nucl. Acids Res.*, **29**: 25-31.
- Jain A. K., Mao, J., Mohiuddin, K. M., (1996). Artificial Neural Networks: A Tutorial. *Computer*, **29**: 31-44.
- Kohonen, V., 1992. Experiential language learning: second language learning as cooperative learner education. (Ed.), *Collaborative language learning and teaching* (Editör: D. Nunan) Cambridge University Press, Cambridge. 284.
- LeCun, Y., 1986. Learning processes in an asymmetric threshold network, Chap 24. *Disordered Systems and Biological Organization* (Editörler: E. Bienenstock, F. Fogelman Soulié, G. Weisbuch) Berlin: Springer Verlag. 233.
- Legarra, A., Robert-Granie, C., Manfredi E., Elsen J. M., 2008. Performance of genomic selection in mice. *Genetics Society of America*, **180**: 611-618.
- Li, H., Vaillancourt, R., Mendham, N., Zhou, M., 2008. Comparative mapping of quantitative trait loci associated with waterlogging tolerance in barley (*Hordeum vulgare* L.). *BMC Genomics*, **9**: 401-412.
- Lin, E., Hwang, Y., Wang, S. C., Gu, Z. C., Chen E. Y., 2006. An artificial neural network approach to the drug efficacy of interferon treatments. *Pharmacogenomics*, **7**: 1017-1024.
- Liu, L. W., Wang, Y., Gong Y. Q., Zhao T.M., Liu, G., Li X. Y., Yu, F. M., 2007. Assessment of genetic purity of tomato (*Lycopersicon esculentum* L.) hybrid using molecular markers. *Sci Hort*, **115**: 7-12.
- Long, N., Gianola, D., Rosa G. J. M., Weigel, K. A., 2011. Application of support vector regressions to genome-assisted prediction of quantitative traits. *Theoretical and Applied Genetics*, **123**: 1065-1074.
- Mace, E. S., Xia, L., Jordan. D. R., Halloran. K., Parh., D. K., Huttner, E., Wenzl, P., Kilian, A., 2009. DArT markers: diversity analyses and mapping in *Sorghum bicolor*. *BMC Genomics*, **9**: 26-37.
- MacKay, D. J. C., 1991. Bayesian interpolation. *Neural Computation*, **4**: 415-447.
- Mantovani, P., Maccaferri, M., Sanguineti, M. C., Tuberosa, R., Catizone, I., Wenzl, P., Thomson, B., Carling, J., Huttner, E., DeAmbrogio, E., Kilian, A., 2008. An integrated DArT-SSR linkage map of durum wheat. *Mol Breed*, **22**: 629-648.

- McCulloch, W. S. and Pitts, W., 1943. A logical calculus of ideas immanent in nervous activity. ***Bulletin of Mathematical Biophysics***, 5: 115-133.
- McLaren, C. G., Bruskiewich, R., Portugal, A. M., Cosico, A. B., 2005. The international rice information system. a platform for meta-analysis of rice crop data. ***Plant Physiol***, 139: 637-642.
- Meuwissen, T. H., Hayes, B. J., Goddard, M. E., 2001. Prediction of total genetic value using genome-wide dense marker maps. ***Genetics***, 157: 1819-829.
- Minsky, M., Papert, S. A., 1969. ***An Introduction to Computational Geometry***. MIT Press, 308.
- Mochida, K., Yamazaki, Y., Ogihara, Y., 2004. Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. ***Mol. Genet. Genomics***, 270: 371-377.
- Moolman, D. W., Aldrich, C., Van Deventer, J. S. J., Strange W. W., 1994. Digital image processing as a tool for on-line monitoring of froth in flotation plants. ***Miner. Eng***, 7: 1149-1164.
- Oakey, H., Verbyla, A., Pitchford, W., Cullis B., Kuchel, H., 2006. Joint modeling of additive and non-additive genetic line effects in single field trials. ***Theor. Appl. Genet.***, 113: 809-819.
- Okut ,H., Gianola, D., Rosa, G. J. M., Weigel, K. A., 2011. Prediction of body mass index in mice using dense molecular markers and a regularized neural network. ***Genetics Research***, 93: 189-201.
- Öztemel, E., 2006. ***Yapay Sinir Ağları***. Papatya Yayıncılık. İstanbul. 232.
- Parh, D. K., Jordan, D. R., Aitken, E. A. B., Mace E. S., Jun-ai P., McIntyre, C. L., Godwin, I. D., 2008. QTL analysis of ergot resistance in sorghum. ***Theor Appl Genet***, 117: 369-382.
- Parker, D. B., 1985. ***Learning Logic***. Technical Report, TR-47. Sloan School of Management, MIT, Cambridge, Mass.
- Peleg., Z., Cakmak, I., Ozturk, L., Yazici, A., Jun, Y., Budak, H., Korol, A. B., Fahima, T., Saranga, Y., 2009. Quantitative trait loci conferring grain mineral nutrient concentrations in durum wheat × wild emmer wheat RIL population ***Theor. Appl. Genet***, 119: 353-369.

- Peleg, Z., Saranga, Y., Suprunova, T., Ronin, Y., Röder, M., Kilian, A., Korol, A., Fahima, T., 2008. High-density genetic map of durum wheat×wild emmer wheat based on SSR and DArT markers. *Theor Appl Genet*, **117**:103-115.
- Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sanchez-Villeda, H., Sorrells, M., and Jannink J.L., 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Gen.*, **5**: 103-113.
- Raman, H., Raman, R., Kilian, A., Detering., F, Long, Y., 2013. A consensus map of rapeseed (*Brassica napus* L.) based on diversity array technology markers: Applications in genetic dissection of qualitative and quantitative traits. *BMC Genomics*, **14**: 277.
- Refenes, A.N., Zapranis, A., Francis, G., 1994. Stock performance modeling using neural networks: A comparative study with regression models. *Neural Networks*, **7**: 375-388.
- Risterucci, A. M., Hippolyte, I., Perrier, X., Xia, L., Caig, V., Evers, M., Huttner, E., Kilian, A., Glaszmann, J.C., 2009. Development and assessment of Diversity Arrays Technology for high-throughput DNA analyses in *Musa*. *Theor Appl Genet*, **119**: 1093-1103.
- Rosenblatt, F., 1958. The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain. *Psychological Review*, **65**: 386-408.
- Rowley, H., Baluja, S., Kanade, T., 1998. Neural network-based face detection. *IEEE Patt. Anal. Mach. Intell.*, **20**: 22-38.
- Schouten, H. J., van de Weg, W. E., Carling, J., 2012. Diversity arrays technology (DArT) markers in apple for genetic linkage maps. *Molecular Breeding*, **29**: 645-660.
- Sessittsch. A., Hackl, E., Wenzl, P., Kilian, A., Kostic, T., Stralis-Pavese. N., Tankouo Sandjong, B., Bodrossy, L., 2006. Diagnostic microbial microarrays in soil ecology. *New Phytol*, **171**: 719-736.
- Shaneh, A., Butler, G., 2006. Bayesian learning for feedforward neural network with application to proteomic data: the glycosylation sites detection of the epidermal growth factor-like proteins associated with cancer as a case study. *In Canadian*

- AI LNAI 4013.** (Editörler: L. Lamontagne, M. Marchand). Berlin-Heidelberg: Springer-Verlag. 110.
- Sorich, M. J., Miners, J. O., McKinnon, R. A., Winkler, D. A., Burden, F. R., Smith, P. A., 2003. Comparison of linear and nonlinear classification algorithms for the prediction of drug and chemical metabolism by human UDP-glucuronosyltransferase isoforms. *J. Chem. Inf. Comput. Sci.*, **43**: 2019-2024.
- Stodart, B. J., Mackay, M. C., Raman, H., 2007. Assessment of molecular diversity in landraces of bread wheat (*Triticum aestivum* L.) held in an ex situ collection with Diversity Arrays Technology (DArT). *Aust J Agric Res*, **58**: 1174-1182.
- Taylor J. W., Buizza, R., 2002. Neural network load forecasting with weather ensemble predictions. *IEEE Trans. Power Syst.*, **17**: 626-632.
- Tibshirani, R., 1996. Regression shrinkage and selection via the LASSO. *J. R. Stat. Soc.*, **58**: 267-288.
- Tomida, S., Hanai, T., Koma, N., Suzuki, Y., Kobayashi, T., Honda, H., 2002. Artificial neural network predictive model for allergic disease using single nucleotide polymorphisms data. *J Biosci Bioeng*, **93**: 470-478.
- Tomita, Y., Tomida, S., Hasegawa, Y., Suzuki, Y., Shirakawa, T., Kobayashi, T., Honda, H., 2004. Artificial neural network approach for selection of susceptible single nucleotide polymorphisms and construction of prediction model on childhood allergic asthma. *BMC Bioinformatics*, **5**: 120-132.
- Tu, J. V., 1996. Advantages and Disadvantages of Using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes. *J. Clin. Epidemiol*, **49**: 1225-1231..
- Van Raden, P. M., 2007. Genomic measures of relationship and inbreeding. *Interbull Bull.*, **37**: 33-36.
- Van Raden, P. M., 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.*, **91**: 4414-4423
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Fritjers, A., Pot, J., Peleman, J., Kuiper, M., Zabeau, M., 1995. AFLP: a new technique for DNA Wngerprinting. *Nucleic Acids Res.*, **23**: 4407-4414.
- Weber, J., May, P., 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet.*, **44**: 388-396.

- Wenzl, P., Carling, J., Kudrna, D., Jaccoud, D., Huttner, E., Kleinhofs, A., Kilian A., 2004. Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proc Natl Acad Sci*, **101**: 9915-9920.
- Wenzl, P., Li, H., Carling, J., Zhou, M., Raman, H., Paul, E., Hearnden, P., Maier, C., Xia, L., Caig, V., Ovesná, J., Cakir, M., Poulsen, D., Wang, J., Raman, R., Smith, K.P., Muehlbauer, G. J., Chalmers, K. J., Kleinhofs, A., Huttner, E., Kilian, A., 2006. A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. *BMC Genomics*, **7**: 206-227.
- Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., Hopper, C., Aschenbrenner-Kilian, G., Evers, M., Hok, P., Duncan, M., Miller, K., Uszynski, G., Kilian A., 2008. Diversity Arrays Technology (DArT): A generic high-density genotyping platform. *7th Safflower Conference*. 2008. Wagga Wagga, Avusturalya.
- Widrow, B., Hoff, M. E., 1960. *Adaptive switching circuits*. 1960 IRE WESTCON Convention Record, New York.
- Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A., Tingey, S. V., 1990. DNA Polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.*, **18**: 6531-6535.
- Winkler, D. A., Burden, F. R., 2004. Modelling Blood-Brain Barrier Partitioning Using Bayesian Neural Nets. *J. Mol. Graphics Modell*, **22**: 499-505
- Wittenberg, A. H. J., Cayla T. L. C., Killian, A., Visser, R. G. F., Schouten, H. J., 2005. Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. *Mol Gen Genomics*, **274**: 30-39.
- Wu, Z., Giger, M. L., Doi, K., Vyborny, C. J., Schmidt, R. A., Metz, C. E., 1993. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. *Radiology*, **187**: 81-87.
- Xia, L., Peng, K., Yang, S., Wenzl, P., Vicente, M. C., Fregene, M., Kilian, A., 2005. DArT for high-throughput genotyping of Cassava (*Manihot esculenta*) and its wild relatives. *Theor Appl Genet*, **110**: 1092-1098.
- Xu, M., Zengi, G., Xu, X., Huang, G., Jiang, R. & Sun, W., 2006. Application of Bayesian regularized BP neural network model for trend analysis, acidity and

- chemical composition of precipitation in North. ***Water, Air, and Soil Pollution***, **172**: 167-184.
- Yang, S., Pang, W., Ash, G., Harper, J., Carling, J., Wenzl, P., Huttner, E., Kilian, A., 2006. Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by Diversity Arrays Technology (DArT). ***Theor Appl Genet***, **113**: 585-595.
- Young, N. D., Menancio-Hautea, D., Fatokun, C. A., Danesh, D., 1992. RFLP technology, crop improvement and international agriculture, Chap 5. ***Biotechnology: Enhancing research on tropical crops in Africa*** (Editörler: G. Thottappilly, L.M. Monti, D.R. Moham, A.W. Moore). Technical Center for Agriculture and Rural Cooperation, International Institute of Tropical Agriculture. 364.
- Zhang, L. Y., Liu, D. C., Guo, X. L., Yang, W. L., Sun, J. Z., Wang, D. W., Sourdille, P., Zhang, A. M., 2011. Investigation of genetic diversity and population structure of common wheat cultivars in northern China using DArT markers. ***BMC Genetics***, **12**: 42-52.
- Zhong, S., Dekkers, J. C. M., Fernando, R. L., Jannink, J. L., 2009. Factors Affecting Accuracy From Genomic Selection in Populations Derived From Multiple Inbred Lines: A Barley Case Study. ***Genetics***, **182**: 355-364.

ÖZGEÇMİŞ

1988 yılında Çanakkale’de doğdu. İlköğretim ve liseyi Çanakkale’de tamamladı. İstanbul Üniversitesi Moleküler Biyoloji ve Genetik Bölümü’nde 2006 yılında başladığı yükseköğrenimini 2010 yılında tamamladı. YYÜ Fen Bilimleri Enstitüsü Zootekni Anabilim Dalı’nda yüksek lisans eğitimine 2013 yılında başladı. Halen aynı fakültede araştırma görevlisi olarak görev yapmaktadır.