

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**İNCE BAĞIRSAK GÖRÜNTÜLERİ ÜZERİNDE
SEZGİSEL ALGORİTMA TEKNİKLERİ İLE POLİP
TEŞHİSİ**

YÜKSEK LİSANS TEZİ

Fatma AKALIN

Enstitü Anabilim Dalı : BİLGİSAYAR MÜHENDİSLİĞİ

Tez Danışmanı : Prof. Dr. Nejat YUMUŞAK

Haziran 2020

BEYAN

Tez içindeki tüm verilerin akademik kurallar çerçevesinde tarafımdan elde edildiğini, görsel ve yazılı tüm bilgi ve sonuçların akademik ve etik kurallara uygun şekilde sunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezde yer alan verilerin bu üniversite veya başka bir üniversitede herhangi bir tez çalışmasında kullanılmadığını beyan ederim.

Fatma AKALIN

10.06.2020

TEŐEKKÜR

Bu alıŐmanın konusunun belirlenilmesinde, veriseti sađlanılma aŐamasında ve hazırlanma sürecinin her basamađında; deđerli bilgilerini ve zamanını benden esirgemeyerek mősait olduđu her fırsatta alıŐmam ile yakından ilgilenip yaptıklarım ve yapmam gerekenler konusunda bana yol gösteren, her Őeyden önce bilim insanı olmanın ve bilim dőnyasına bir Őeyler katmanın mutluluđunu bizzat örnek olarak bana hissettiren danıŐman hocam Prof. Dr. Nejat YUMUŐAK'a teŐekkür ve minnetimi özellikle belirtmek istiyorum.

İÇİNDEKİLER

TEŞEKKÜR	i
İÇİNDEKİLER	ii
SİMGELER VE KISALTMALAR LİSTESİ	v
ŞEKİLLER LİSTESİ	vi
TABLolar LİSTESİ	vii
ÖZET	viii
SUMMARY	ix
BÖLÜM 1.	
GİRİŞ	1
BÖLÜM 2.	
KAYNAK ARAŞTIRMASI	4
2.1. İnce Bağırsak	4
2.2. Polip	5
2.3. Bilgisayar Destekli Sistemler Nedir? Polipleri Teşhis Etmek Husususunda Niçin İhtiyaç Duyulur?	5
BÖLÜM 3.	
MATERYAL VE YÖNTEM	7
3.1. Veritabanı	7
3.2. Çalışmada Kullanılan Yazılım ve Donanım	7
3.3. Yöntem	8
3.3.1. Görüntü işleme yöntemleri	8
3.3.1.1. Grayscale filtresi	8
3.3.1.2. Unsharp filtresi.....	8

3.3.1.3. Özellik çıkarımı (Feature extraction).....	9
3.3.1.3.1. Yerel İkili Örüntü yöntemi (Local Binary Pattern yöntemi)	9
3.3.1.3.2. Temel Bileşenler Analizi yöntemi (Principal Component Analysis yöntemi)	11
3.3.2. Dengesiz veriseti	12
3.3.2.1. SMOTE yöntemi (Synthetic minority oversampling technique)	13
3.4. Sınıflandırmada Kullanılan Algoritmalar	15
3.4.1. Naif Bayes algoritması (Naive Bayes algorithm)	15
3.4.1.1. Bayes teoremi	15
3.4.2. Destek Vektör Makineleri algoritması (Support Vector Machine algorithm)	17
3.4.2.1. Doğrusal olarak ayrılması mümkün veriler için destek vektör makineleri	18
3.4.2.2. Doğrusal olarak ayrılması mümkün olmayan veriler için destek vektör makineleri	21
3.4.3. Rastgele Orman algoritması (Random Forest algorithm)	22
3.4.4. Karar Ağaçları algoritması (Decision Trees algorithm)	25
3.4.4.1. Tekrarlı İkilikçi Ağaç algoritması (Iterative Dichotomiser 3)	27
3.4.5. Resnet Mimarisi	29
3.5. Analizler	31
3.5.1. Model değerlendirme yöntemleri	31
3.5.1.1. Hold-Out yöntemi	31
3.5.1.1.1. Eğitim verisi (Training data)	31
3.5.1.1.2. Test verisi (Test data)	32
3.5.1.2. Çapraz Geçerlilik yöntemi (Cross validation method).	31
3.6. Performans Değerlendirme Ölçütleri	33
3.6.1. Doğru pozitif oranı (TP rate)	33
3.6.2. Yanlış pozitif oranı (FP rate)	34
3.6.3. Doğru negatif oranı (TN rate)	34

3.6.4. Yanlış negatif oranı (FN rate)	34
3.6.6. Kesinlik (Precision)	35
3.6.7. F ölçütü (F score)	35
3.6.8. Alıcı İşletim Karakteristiği eğrisi (Receiving Operating Characteristic curve)	35
3.6.9. Başarı yüzdesi	36
3.6.9.1. Karışıklık matrisi (Confusion matrix)	36
BÖLÜM 4.	
ARAŞTIRMA BULGULARI	38
4.1. Deneysel Çalışmalar	38
BÖLÜM 5.	
TARTIŞMA VE SONUÇ	43
5.1. Gelecek Çalışmalar	45
KAYNAKLAR	46
ÖZGEÇMİŞ	49

SİMGELER VE KISALTMALAR LİSTESİ

CART	: Classification and Regression Tree–Sınıflandırma ve Regresyon Ağacı
FN	: False Negative–Yanlış Negatif
FP	: False Positive–Yanlış Pozitif
ID 3	: Iterative Dichotomiser 3–Tekrarlı İkili Ağaç 3
LBP	: Local Binary Pattern–Yerel İkili Örüntü
PCA	: Principal Component Analysis–Temel Bileşenler Analizi
ROC	: Receiving Operating Characteristic–Alıcı İşletim Karakteristiği
SMOTE	: Synthetic Minority Oversampling Technique–Sentetik Azınlık Aşırı Örneklemme Tekniği
SVM	: Support Vector Machine–Destek Vektör Makineleri
TN	: True Negative–Doğru Negatif
TP	: True Positive–Doğru Pozitif

ŞEKİLLER LİSTESİ

Şekil 2.1. Bağırsak dokusunda yer alan polip görüntüsü	5
Şekil 3.1. Orijinal görüntülere grayscale ve unsharp filtresinin uygulanılması ...	9
Şekil 3.2. Görüntüden alınan bir kesit üzerinde LBP yönteminin uygulanılması .	10
Şekil 3.3. Ortalamadan çıkartılmak suretiyle normalize edilen verilerin grafik üzerinde görülmesi	11
Şekil 3.4. Kovaryans matrisinin öz vektörler ve normalize edilen veri ile ifade edilmesi	12
Şekil 3.5. Her iki özvektöre PCA yöntemi uygulandıktan sonra grafik üzerinde gösterilmesi	12
Şekil 3.6. SMOTE yöntemi üzerinden sentetik verilerin üretilmesi	14
Şekil 3.7. SMOTE yöntemi uygulanılmadan önceki ve uygulandıktan sonraki veriseti	14
Şekil 3.8. Doğrusal olarak ayrılabilen veriler için Destek Vektör Makineleri	18
Şekil 3.9. Mesafelerine göre doğrusal olarak ayrılabilen veriler için Destek Vektör Makineleri	19
Şekil 3.10. Doğrusal bir şekilde ayrılabilen ve ayrılamayan veri kümeleri için hiper düzlemin belirlenilmesi	22
Şekil 3.11. Destek Vektör Makineleri için kullanılan bazı kernel fonksiyonları	22
Şekil 3.12. Rastgele Orman algoritmasına ait ağaç yapısı	23
Şekil 3.13. Karar Ağacı algoritmasının kaba kodu	26
Şekil 3.14. Karar ağacına uygulanılacak olan veri seti	28
Şekil 3.15. Veri setinin uygulanılması sonucunda elde edilen karar ağacı	29
Şekil 3.16. Resnet Mimarisi içerisinde sunulan blok yapısı.....	30
Şekil 3.17. Çalışmada kullanılan Resnet Mimarisi içerisinde yer verilen süreçler..	30
Şekil 3.18. İterasyon sayısının 10 alınması suretiyle veri setine Çapraz Geçerlilik yönteminin uygulanması	32

Şekil 3.17.Karışıklık matrisinin genel çerçevesi	36
Şekil 4.1. Mevcut sistemin modellenmesi	39
Şekil 4.2. Yöntemler bazında sınıflandırma yüzdeleri.....	43

TABLULAR LİSTESİ

Tablo 4.1. İyileştirilmiş görüntü verilerinin Yerel İkili Ö rüntü yöntemi ile sayısallaştırılmasının ardından sınıflandırma algortmalarına tabi tutulması suretiyle elde edilen performans ölçütleri ve başarı yüzdesi değerleri	40
Tablo 4.2. Sayısallaştırılmış verilere Temel Bileşenler analizi yönteminin uygulanılmasının ardından sınıflandırma algoritmalarına tabi tutulması suretiyle elde edilen performans ölçütleri ve başarı yüzdesi değerleri	40
Tablo 4.3. PCA yönteminin uygulanılmasının ardından elde edilen verilerin SMOTE yöntemiyle dengeli veri seti haline getirilmesinin ardından sınıflandırma algortmalarına tabi tutulması suretiyle elde edilen performans ölçütleri ve başarı yüzdesi değerleri	41

ÖZET

Anahtar Kelimeler: İnce Bağırsak, Polip Tespiti, Görüntü Ön İşleme, Özellik Çıkarımı, SMOTE, Sınıflandırma

İnce bağırsak, sindirim kanalının mideden sonra başlayan, yaklaşık 5-6 metre sonra kalın bağırsak ile sonlanan en büyük organıdır. Sindirim sistemi içerisinde en çok işlem gerçekleşen ince bağırsakta sindirim işleminin gerçekleşme oranı %90'dır. Aynı zamanda ince bağırsak sindirim ve emilimin ana merkezi, vücudun en büyük endokrin organı ve bağışıklık sisteminin en büyük kısmıdır. İnce bağırsak bu kadar önemli işlevlere sahipken, ince bağırsağın herhangi bir tabakasında köklenen ve dışarıya doğru tümör şeklinde itilen poliplerin oranı, sindirim kanalının diğer organlarından çok daha düşüktür.

Bu amaçla ince bağırsaktaki polipleri saptamak için başlatılan bu çalışmada, görüntüler matlab programında çeşitli filtreler kullanılarak ön işleme yöntemlerine tabi tutulmuştur. Akabinde matlab programında LBP özellik çıkarım yöntemi kullanılarak görüntüler sayısallaştırılmış ve sonuçlar excel formatında bir dosyaya kaydedilmiştir. Excel dosyası, dönüştürücü programı yardımıyla csv dosyasına dönüştürüldükten sonra bu dosya weka programında açılmıştır. Dosya içerisindeki veriler Destek Vektör Makineleri Algoritması, Rastgele Orman Algoritması, Karar Ağacı Algoritması ve Naive Bayes Algoritması tarafından sınıflandırılmış ve sırasıyla %70, %71, %63 ve %67 başarı oranları elde edilmiştir. Daha sonra bu veriler PCA özellik çıkarım yöntemi kullanılarak daha detaylı incelenmiş ve veriler kesişme ve çakışmalardan arındırılarak, algoritmalar tarafından yeniden sınıflandırılmıştır. Sonuçlar sırasıyla %73, %73, %68 ve %69 başarı oranları göstermiştir. Diğer yandan ince bağırsakta görülen polip oranının seyrek olması, mevcut veri setinde polip içeren ince bağırsak görüntülerinin, polip içermeyen ince bağırsak görüntülerinden daha az olmasına neden olmuştur. Poliplerin tespit oranını olumsuz etkileyen dengesiz veriler SMOTE yöntemiyle daha dengeli hale getirilmiş ve algoritmalar tarafından yeniden sınıflandırılmıştır. Sonuçlar sırasıyla %79, %76, %72 ve %68 başarı oranları göstermiştir. Bu nihai başarı oranlarının ince bağırsakta polip olup olmadığına dair karar vermede doktorlar için destek sağlayacağı planlanmaktadır.

Ek olarak geleneksel algoritmaların yanında son zamanlarda kullanılmaya başlanmış ve derin öğrenme üzerinde temellenmiş olan Resnet Mimarisi kapsamında veriler yeniden sınıflandırılmış ve başarı oranı %83'e çıkmıştır.

POLYP DIAGNOSTIC WITH HEURISTIC ALGORITHM TECHNIQUES ON SMALL BOWEL IMAGES

SUMMARY

Keywords: Small Bowel, Polyp Detection, Image Preprocessing, Feature Extraction, SMOTE, Classification

The small bowel is the longest organ of the digestive tract, starting after the stomach and ending with the large bowel at the end of about 5-6 meters. In the small bowel which has the most processing in the digestive system, the rate of digestion is 90%. Also it is the main center for digestion and absorption, the largest endocrine organ of the body and the biggest part of the immune system. While the small bowel has such important functions, the proportion of polyps rooted in any layer of the small bowel and pushed outwards in the form of tumors is much lower than the other organs of the digestive track.

For this purpose, this study was started to detect polyps in the small intestine and images were pre-processed using various filters in the matlab program. Subsequently, the images were digitized using LBP feature extraction method in matlab program and the results were recorded in a file in excel format. After excel file is converted to csv file with the help of converter program, this file is opened in weka program. The data in the file were classified by Support Vector Machines, Random Forest Algorithm, Decision Tree Algorithm and Naive Bayes algorithm and achieved success rates of 70%, 71%, 63% and 67% were obtained respectively. Afterwards, these data were examined in detail by using PCA feature extraction method and the data were re-classified by the algorithms after being cleared of overlaps and conflicts. Results showed success rates of 73%, 73%, 68%, and 69% respectively. On the other hand, being sparse polyp rate seeing in small bowel, cause small bowel images with polyp be more less than small bowel images without polyp. Unbalanced data that adversely affect the detection rate of polyps were made more stable by SMOTE method and reclassified by algorithms. Results showed success rates of 79%, 76% ,72% 68%, respectively. These final success rates are planned to provide support for doctors in deciding whether polyps are present in the small bowel.

In addition, besides the traditional algorithms, data has been reclassified and the success rate has increased to %83 within the scope of Resnet Architecture, which has been used recently and based on deep learning.

BÖLÜM 1. GİRİŞ

Ağız ile anüs arasında oluşum gösteren gastrointestinal sistemin %90'lık bir kısmını meydana getiren ince bağırsak, sahip olduğu yapı itibari ile üzerinde inceleme yapılması zor bir organ konumunda olmuştur. Ancak 2000'li yılların başında ortaya çıkan kablosuz kapsül endoskopi yöntemi sayesinde bu zorluğu aşarak üzerinde net bir şekilde inceleme yapılması mümkün kılınmıştır [1]. Bahsedilen bu yöntemde yutulan bir kapsül ile insan vücuduna giren ve ortalama 8 saat boyunca büyük görüntü yığınlarını bilgisayarlar aracılığıyla uzmanlara sunan kablosuz kapsül endoskopi, tıp dünyasında önemli bir yere sahiptir. Fakat bu görüntüleri değerlendiren kişilerin de insani bir hak olarak bir süre sonra dikkatlerinin dağılabilmesi ihtimali polip teşhisi sürecine negatif olarak yansımış ve bilgisayar destekli sistemlerin gerekliliğini ortaya koymuştur [2].

Aynı zamanda ince bağırsak; gastrointestinal sistem içerisindeki tüm anomali miktarının yalnızca %3'ünü barındırmakla birlikte anomali durumları içerisinde oluşum gösteren polip oranının az miktarda seyrettiği hayati bir organdır. İnce bağırsakta poliplere benzeyen villus ile mikrovillus çıkıntılarının olması ve bağırsağın kıvrımlı bir yapıya sahip olması da ince bağırsak içerisinde nadir oranlarda oluşum gösteren poliplerin teşhis konusunda önemini ortaya koymaktadır [3].

Ek olarak ince bağırsak polipleri, çok spesifik belirti göstermeyen yapılar oldukları için erken teşhisi önem taşımakta olup aksi taktirde kanser hastalığı ile karşı karşıya kalınması ihtimalini barındırmaktadır [4]. Aynı zamanda erken teşhisin sağlanamaması sağlık alanında ülkenin ve kişinin mali anlamda bir yük altında kalmasına neden olurken hastanın da zorlu bir sürece girmesine ve sonucun hasta aleyhine vuku bulmasına sebebiyet vermektedir [5]. Bu çalışmanın hastaya getirdiği

zorlu süreçlerin kısaltılmasının yanısıra görüntülere bakan nitelikli uzmanlarımız açısından da üzerlerindeki yükün hafifletilmesi hususunda artışı olacağı düşünülmektedir.

Bu bağlamda sunulan çalışma kapsamında verisetinde mevcut olan görüntülere ilk olarak filtreler uygulanarak iyileştirilme sağlanılmıştır. İyileştirilmiş bu görüntüler özellik çıkarım yöntemi olan Yerel İkili Örüntü vasıtasıyla sayısallaştırılmış akabinde diğer özellik çıkarım yöntemi olan Temel Bileşenler Analizi ile alakasız verilerin uzaklaştırılması sağlanarak daha etkin bir çalışma ve gürültüye karşı daha az hassasiyet elde edilmiştir. Nihai duruma getirilen tüm veriler verisetinin dengesiz olma durumu göz önünde bulundurularak veri dengeleme yöntemlerinden biri olan SMOTE yöntemine tabi tutulmuş ve verilerin dengeli bir duruma getirilmesi sağlanmıştır. Bu aşamaya kadar yapılan tüm işlemler kapsamında sınıflandırılacak veriler için kullanılacak geleneksel algoritmalar hem tekrar tekrar denenerek hem de algoritmaların doğası baz alınarak titizlikle seçilmiştir. Seçilen algoritmalar içerisinde bulunan ilk algoritma Naive Bayes Algoritması olup hızlı çalışabilen bir algoritma özelliği göstermektedir. Ancak çalışmada kullanılan tüm algoritmalar içerisinde değerlendirildiğinde olasılıksal ilkeler üzerinde temellenmesi sebebi ile verisetinin dengesizliğinden çok etkilenmemesi SMOTE yönteminin bu algoritma açısından maximum bir getiri sunmaması ile sonuçlanmaktadır. Çalışmada kullanılan 2. algoritma olan Destek Vektör Makineleri Algoritması, Naive Bayes Algoritmasının aksine olasılıksal sonuçlar üretememektedir. Ayrıca Destek Vektör Makineleri Algoritması içerisindeki işlem yükünden dolayı örnek sayısının fazla olması halinde yüksek performans da gösterememektedir. Ek olarak verisetinde kesişme ya da çakışmaların fazla miktarlarda bulunması özetle verisetinin parazit barındırması durumunda da istenilen sonuç elde edilememektedir. Bu tür olumsuz özelliklerin yanısıra yüksek boyutlu uzaylarda özellikle boyutun örnek sayısından yüksek olduğu hallerde algoritmanın mükemmel sonuçlar verdiği bilinmektedir. Aynı zamanda aşırı öğrenme durumundan etkilenmeyen Destek Vektör Makineleri Algoritması çarpık verilere karşı da dirençlidir. Çalışmada kullanılan diğer algoritmalar, ağaç temelli algoritmalar içerisinde değerlendirilen Karar Ağaçları Algoritması ile Rastgele Orman Algoritmasıdır. Rastgele Orman Algoritmasının Karar Ağacı Algoritmasına

bakış en net ayrımı, tek bir ağacın istenilen sonuca ulaşmada yeterli görülmediği durumlarda birden çok ağacın gücü kullanılarak istenilen sonucu elde etme bağlamında çalışan bir algoritma olmasıdır. Genel anlamda incelendiğinde ağaç temelli algoritmalar; uygulanması, değerlendirilmesi ve istatistiksel temellere dayanmaması sebebi ile ilgili özellikler arasında bir ilişki kurulmasının kolay olması açısından kullanışlı olma özelliği göstermektedir. Aynı zamanda verisetinin eksik olma durumu karşısında fazla etkilenmemekle birlikte istenilen özellikleri kategorilere ayırdıktan sonra sınıflar arasında kesişme ya da çakışmanın görülmemesi ağaç temelli algoritmaların artılarından. Aşırı öğrenme göstermesi ise algoritmanın bir eksisidir. Bunun yanısıra aç gözlü bir yaklaşım benimsenmesi bu tür algoritmaların gelecekte çok iyi bir ağaç üretmek suretiyle mükemmel bir sonuca ulaşmasının maskelenmesine de neden olabilmektedir. Çünkü aç gözlü yaklaşım mevcut durumda en iyi sonucu üretmek üzerinde bir tutum sergilemektedir. Ek olarak kökten yapraklara doğru gerçekleşecek olan bölünme sürecinde, tüm özellikler kapsamında ilgili özelliğin seçilmesi sırasında uygun yöntemin hedef verisetine göre doğru bir şekilde seçilmesi de büyük önem arz etmektedir.

Sunulan çalışma kapsamında teşhis amaçlı kullanılan diğer yöntem ise çok uzun yıllardır çalışmalarda yerini alan geleneksel algoritmaların aksine son dönemlerde kullanılmaya başlanmış ve özellik çıkarım sürecini kendi içerisinde barındırıyor olan Resnet Mimarisi'dir. Resnet Mimarisi, derin öğrenme içerisinde gelişim gösteren bir teknik olan Convolutional Neural Network Mimarisindeki gradyanların yok olması probleminde karşı çözüm niteliğinde sunulmuş başarılı bir yöntem olma özelliği göstermektedir.

Sonuç olarak yapılan çalışma ince bağırsak lokasyonlarından elde edilen görüntüler kullanılarak 2 farklı yöntem üzerinden değerlendirilmiştir. İlk yöntem de KID ATLAS Veriseti içerisinde mevcut olan polipli ve polipsiz ince bağırsak görüntüleri iyileştirilmiş daha sonra farklı özellik çıkarım yöntemlerinin kullanılması ile geleneksel algoritmalar üzerinden sınıflandırılmaya tabi tutulmuştur. İkinci yöntem de ise ilgili görüntüler son dönemlerde sıklıkla kullanılan derin öğrenme temelli Resnet Mimarisi ile sınıflandırılmak sureti ile bir başarı oranı elde etmiştir.

BÖLÜM 2. KAYNAK ARAŞTIRMASI

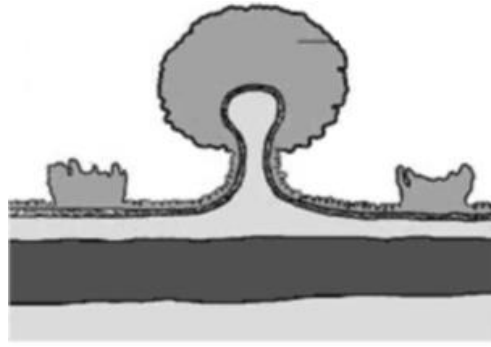
2.1. İnce Bağırsak

Sindirim kanalının en büyük organı olan ince bağırsak, mide ve kalın bağırsak arasına yerleşen yaklaşık 5 metre uzunluğuna sahip bir sindirim organıdır. Vücuttaki besin emilimi için birincil organ vazifesi gören ince bağırsak; duodenum, jejunum ve ileum olarak 3 bölgeye ayrılmıştır. Mideden gelen yarı sindirilmiş yiyecekleri, pankreas vasıtasıyla enzimleri ve karaciğerden de safrayı alan duodenum; bu besinleri karıştırarak sindirir ve yağ asitlerini çözünür hale getirir. Ardından duodenum bölgesinden jejunum bölgesine yiyeceklerin iletimi sağlanıp emilim işlemi gerçekleştirildikten sonra ileum bölgesine geçiş olmaktadır. İleum bölgesinde de jejunum bölgesinde emilemeyen besinlerin emilim işlemi gerçekleştirilerek önemli bir safha tamamlanmış olur [6]. Bahsedilen bu emilim işlemleri ince bağırsakta bulunan villus ile mikrovillus yapıları üzerinden gerçekleşmekte olup bu yapıların varlığı hayati önem arz etmektedir.

Aynı zamanda önemli bağışıklık fonksiyonlarını barındırarak bağışıklık sistemi üzerinde önemli bir paya sahip olan ince bağırsak, vücudun en büyük endokrin organı ve bağışıklık sisteminin en büyük kısmı olma özelliği göstermektedir. Sindirim işlemlerinin yaklaşık %90'ının gerçekleştiği yer olan ince bağırsağın önemli birçok görevi olmasına rağmen içerisinde meydana gelip zaman zaman bir tehdit görevi üstlenen polip oranı, diğer sindirim organları olan mide ve kalın bağırsağa göre oldukça az seyretmektedir. Bu sebepten dolayı ince bağırsak üzerinde yer alan polipleri algılamak zor bir hal almış ve üzerinde bilgisayar destekli sistemler ile çalışılarak aydınlatılması gereken önemli bir organ konumuna yükselmiştir.

2.2. Polip

Polipler insan vücudunun dokularında gelişim gösteren yapılar olup iyi huylu ve kötü huylu olarak değerlendirilmektedir [7]. İyi huylu poliplerin doku üzerinde bir kabartı şeklinde oluşup herhangi bir zararlı etki göstermediği gözlemlenirken; kötü huylu poliplerin çoğaldığı, büyüdüğü hatta kansere dönüşüm gösterdiği gözlemlenmiştir. Şekil 2.1.'de bağırsak dokusunda yer alan polipli bir görüntü verilmiştir.



Şekil 2.1. Bağırsak dokusunda yer alan polip görüntüsü [7]

Şekil, davranış ve büyüklük gibi çeşitli başlıklar altında sınıflandırılan polipler genellikle bir belirti göstermeyip tesadüfî bir şekilde tanı konulabilen bir rahatsızlık olmanın yanı sıra kısmî ya da tam ince bağırsak tıkanıklığı ile demir eksikliği anemisine bağlı açık kanama görülebilmesi olasılıklar arasında yer almaktadır [4]. Aynı zamanda sigara alışkanlığı, kilo fazlalığı, 50 yaş ve üzerinde bulunulması, yağ içeriğinden zengin beslenilmesi, hareketsiz bir yaşam sürülmesi, sıklıkla alkol kullanılması, et gıdasının fazla tüketilmesi, ailede bu tür hastalıkla daha önceden karşılaşmış olması gibi genetik ya da zararlı alışkanlıkların oluşturduğu etkenler ile polip riski daha da artmaktadır.

2.3. Bilgisayar Destekli Sistemler Nedir? Polip Teşhisi Hususunda Niçin İhtiyaç Duyulur?

Bilgisayar destekli sistemler ilk zamanlarda tıp dünyasında kullanılmamaktayken mühendislik bölümü bu alanı konusu içerisine dahil ettikten sonra tıp dünyasında da

kullanılmaya başlanmıştır. Hatta günümüzde doktorlar, inceleme yöntemlerinin birçoğunda gelişmiş bilgisayar sistemlerinden faydalanmaktadır.

Bu çalışma kapsamında bilgisayar destekli sistemler, görüntüleme üzerinde yoğun olarak kullanılmakta olup bilgisayarlar aracılığıyla hastaların durumu ya da problemi karşısında olayın anlaşılır hale gelmesini sağlayan akabinde doktorların birçok konuda karar vermesinde yardımcı olan bir yapı haline gelmiştir. Aşağıda verilen ifadeler, polip teşhisi hususunda bilgisayar destekli sistemler ile görüntüler üzerinde yapılan çalışmaların niçin gerekli olduğuna dair bir açıklama niteliği taşımakla birlikte yapılan çalışmanın amacını ifade etmektedir.

- 1-) İnce bağırsakta görülen poliplerin belirtileri çok genel belirtiler olduğu için doktorlar tarafından doğru kararların güç alınmasına neden olmaktadır.
- 2-) İnce bağırsakta görülen polip oranı, diğer sindirim organları olan mide ve kalın bağırsağa göre düşük miktarlarda seyrettiği için bu durum veritabanında dengesiz verisetinin varlığına yol açmakta ve üzerinde çalışılması gereken bir konu haline gelmektedir.
- 3-) İnce bağırsağın emilim yüzeyini arttıran ve miktar olarak da 5 milyondan fazla oluşum gösteren villus ile mikrovillus yapıları poliplere benzediği için teşhisi zorlaştırmaktadırlar.
- 4-) Görüntüleme koşullarını zorlaştıran çeşitli güçlüklerin olması görüntülerin netliğine negatif olarak yansiyarak süreci etkilemektedir.

Belirtilen bu sebeplerden ötürü tıp dünyasında hastaların en fazla yarısına tanı konulabilmekte olup aksi durumda hasta geç dönemde bağırsak tıkanıklığı ile ameliyata alınabilmekte ya da ilerlemiş kanser hastalığı ile karşı karşıya kalabilmektedir.

BÖLÜM 3. MATERYAL VE YÖNTEM

3.1. Veritabanı

Bu çalışmada kullanılan KID Atlas veriseti kablosuz kapsül endoskopi yöntemi vasıtası ile gastrointestinal bölgelerden elde edilen video ve görüntülerden oluşmaktadır. Bu veriler Mirocam Kapsül Endoskop ile elde edilen 360x360 çözünürlüğe sahip görüntü ve videolardır [5]. Mirocam, Kore Bilim ve Teknoloji Bakanlığı tarafından kurulmuş ve Akıllı Mikrosistem Merkezi ile geliştirilmiş kapsül endoskop sistemidir. Standart modeli MC1000-W iken en yeni modeli MC2000 olup üzerine yerleştirilen kameralar ile ince bağırsağın iki yönlü görüntülenmesine olanak sağlamaktadır [8].

Tıp ve mühendislik dünyası arasındaki etkileşime yardımcı olan KID atlas veriseti bu çalışmada kullanılmış polipli ve polipsiz ince bağırsak görüntüleri üzerinde çalışılmak suretiyle bulgular elde edilmiştir.

3.2. Çalışmada Kullanılan Yazılım ve Donanım

Bu çalışmada görüntülere uygulanan filtreler ile görüntülerin sayısallaştırılması için kullanılan yöntem, MATLAB programında gerçekleştirilir. Akabinde sayısallaştırılan bu görüntüler üzerinde iyileştirilmelerin yapılması ve performans durumlarının belirlenilmesi aşamaları, WEKA programı aracılığıyla sağlanır. Çalışmada yapılan tüm işlemler Windows 10 üzerinde 2GHz i7 işlemcili ve 8 Gb RAM bulunduran bilgisayar kullanılarak gerçekleştirilir.

3.3. Yöntem

3.3.1. Görüntü işleme yöntemleri

Görüntü işleme; orijinal görüntü değerlerinin pixel değerlerini, komşuluk değerlerini, pozisyonunu ve mekânsal ilişkisini dikkate alıp çeşitli işlemler uygulamak suretiyle yeni bir görüntü sunma işlemidir [9]. Bu işlem sırasında sayısal hale getirilen görüntülere çeşitli işlemlerin uygulanması ya da görüntülerden istenilen bilgilerin çıkartılması şeklinde inşa edilen bu yöntem, yeni nesil teknolojiler arasında yerini almaktadır. Bilgisayar bilimleri disiplinlerinin arasında olan görüntü işleme yönteminin amaçları; görünmesi zor varlıkları gözleme, gürültülü görüntüler üzerinde iyileştirmelerde bulunma, görüntüdeki çeşitli nesnelere tanıma ve ayırt etme şeklinde sayılabilir. Tıbbi alanlarda da sıklıkla kullanılan görüntü işleme, doktorlara birçok konuda destek sağlayarak doktorların karar vermesinde yardımcı bir görev üstlenmektedir.

3.3.1.1. Grayscale filtresi

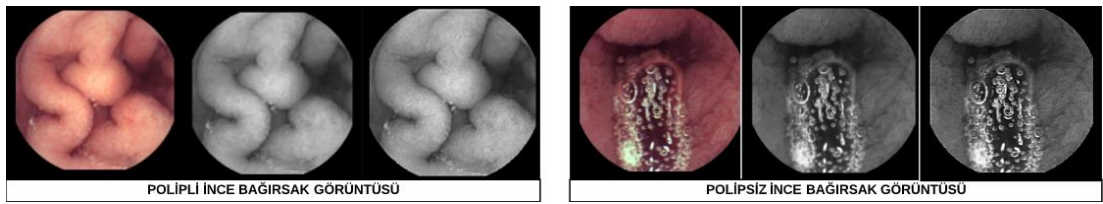
Görüntü işleme yöntemleri içerisinde yer alan grayscale filtresi sayesinde orijinal görüntü gri tonlarına çevrilmektedir. Bu bağlamda ince bağırsak lokasyonlarından alınan polipli ve polipsiz görüntüler, ilk olarak matlab programında görüntü işleme yöntemlerinden biri olan grayscale filtresine tabi tutulmakta ve bu sayede bir sonraki aşamada uygulanacak olan unsharp filtresi için 2 boyutlu görüntü formatına dönüşümü gerçekleştirilmektedir.

3.3.1.2. Unsharp filtresi

Unsharp filtresinin temel mantığı, orijinal görüntü üzerinde ölçeklendirilmiş artışlar ve parlaklıklar sunmak suretiyle iyileştirmeler yapmaktır [10]. Tıbbi çalışmalarda da sıklıkla kullanılan unsharp filtresi, görüntüleri keskin bir hale getirerek, net olmayan görüntülerdeki ayrıntıları açığa çıkartıp görüntüler üzerinde sağlıklı bir çalışma yapmayı mümkün kılan yararlı bir filtredir. Matlab programında grayscale

filtresinden hemen sonra kullanılan unsharp filtresi ile saptanılması istenilen poliplerin yeri daha belirgin hale getirilmekte ve polipleri tespit etme hususunda başarı sağlanmaktadır.

Bu bilgiler dikkate alındığında çalışmada kullanılan görüntülere filtreler uygulanarak görüntülerin netliği keskinleştirilmiş ve polipin sınırını oluşturan alanda bir yoğunluk meydana getirilmek suretiyle polip tespitinde başarı sağlanmıştır [11]. Matlabta görüntülere uygulanan bu filtreler Şekil 3.1.'de gösterilmektedir.



Şekil 3.1. Orijinal görüntülere grayscale ve unsharp filtresinin uygulanması

3.3.1.3. Özellik çıkarımı (Feature extraction)

Özellik çıkarımı görüntüyü bir bütün olarak işlemek yerine görüntüyü oluşturan özelliklerden bazılarının ele alınması ve görüntünün ele alınan bu özellikler üzerinden inşa edilmesidir. Bu cümleden hareketle görüntülerde var olan polipler göz önüne alındığında genellikle çevrelerinden farklı bir renk tonu ile ayrımının mümkün olması çalışmamızda kullandığımız özellik çıkarımının gerekliliğine bir cevap niteliği taşımaktadır [12]. Diğer taraftan özellik çıkarım işlemi bir boyut azaltma yöntemi olarak karmaşık olan verinin boyutlarının azaltılması ve daha basit bir hale indirgenilmesi anlamına da gelmektedir. Özellik çıkarımı sayesinde performansın bu durumdan pozitif yönde etkilendiği görülmekle birlikte başarı oranında bir artış sağlanmaktadır.

3.3.1.3.1. Yerel İkili Örüntü yöntemi (Local Binary Pattern yöntemi)

Sınıflandırma için kullanılan bir görsel tanımlayıcı yöntemi olan LBP, merkez pixel ile çevre pixel değerleri arasında karşılaştırmaların yapılması suretiyle aralarındaki farkı ayırt edebilme yetisine sahip olan bir desen analiz yöntemidir [13].

Bu yöntem dahilinde kullanılan LBP operatörü, görüntünün her piksel değeri için birler ve sıfırlardan meydana gelen bir etiket oluşturmaktadır. Bu kapsamda ilk aşama olarak LBP operatörünün kullanılması ile komşuluk analizleri yapılır. Bu komşuluk analizleri 3.1. ve 3.2. denklemlerinde gösterilmekte olup denklem içerisinde yer alan p , komşu sayısını; y , merkez ile komşu piksel arasındaki farkı; x_c , LBP etiketi üzerinden üretilen merkez pikselini; x_p , merkez pikselin komşularını ve $u(y)$, LBP operatörü sonucunda üretilen bitleri ifade etmektedir [14].

$$LBP(x_c) = \sum_{p=0}^{p-1} u(x_p - x_c) 2^p \quad (3.1)[14]$$

$$u(y) = \begin{cases} 1 & y \geq 0 \\ 0 & y < 0 \end{cases} \quad (3.2)[14]$$

Bu bağlamda merkez pixel, komşu pixel değerleri ile karşılaştırılmaya merkezi eşik değeri üzerinden tabi tutulduktan sonra her bir komşuluk bölgesi için 8 bitlik LBP kodu oluşturmaktadır. Bu 8 bitlik kod vasıtasıyla 10 tabanında hesaplanan LBP kodu, 0 ile 255 arasında değişen 256 değerden her birine ne kadar sıklıkla rastlanıldığı ile ilgili bilgi vermektedir [13]. LBP değerinin hesaplanılmasına ilişkin bir örnek Şekil 3.2.'de gösterilmiştir.



Şekil 3.2. Görüntüden alınan bir kesit üzerinde LBP yönteminin uygulanması [13]

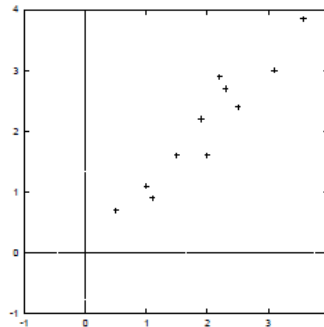
Matlab programında görüntülere uygulanan bu yöntem ile görüntüler sayısallaştırılmış akabinde sayısallaştırılan bu görüntüler excel dosyasına kayıt edilmiştir. Program dönüştürücü ile csv dosyasına dönüştürülen bu excel dosyası, weka programında açılmak üzere hazır hale getirilmiştir.

3.2.1.3.2. Temel Bileşenler Analizi yöntemi (Principal Component Analysis yöntemi)

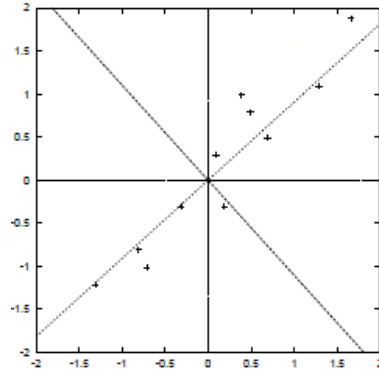
Temel Bileşenler Analizi, mevcut algoritması sayesinde görüntü boyutunun azaltılması için görüntünün özelliklerini kullanan bir analiz yöntemidir. Burada mevcut veri örneklerinin çok boyutlu uzaydan düşük boyutlu uzaya yansıtılması PCA'nın özünü oluşturmaktadır. Aynı zamanda PCA algoritması geniş bilgi miktarına sahip büyük varyanslı bileşenleri korumak suretiyle görüntünün boyutunu azaltırken; küçük varyanslı, yeterli bilgiye sahip olmayan bileşenleri görüntüden uzaklaştırmak şeklinde bir yol izler. Bu çalışmada çok değişkenli analiz yöntemlerinden biri olan PCA algoritması, düşük boyutlu bir uzayda n örnekle verinin en iyi nasıl temsil edileceği üzerinde kapsamlı bir değerlendirme sunmaktadır [15].

Temel bileşenler analizi yönteminde izlenecek yol, kullanılacak olan verilerin elde edilmesinin ardından veriye ait ortalama değer her elemandan sırayla çıkarılması ile başlar. Kovaryans matrisi kullanılarak iki veya daha fazla rastgele değişken kümesi arasındaki ilişkinin miktarı hesaplandıktan sonra kovaryans matrisinin öz vektörleri ile öz değerlerinin hesaplanması sağlanır. Seçilen bileşenler vasıtasıyla özellik vektörünün oluşturulması ile en fazla öz değere sahip olan öz vektörün temel bileşen olarak kabul edilmesi uygun görülür ve yeni boyuta sahip bir veri seti elde edilir [16].

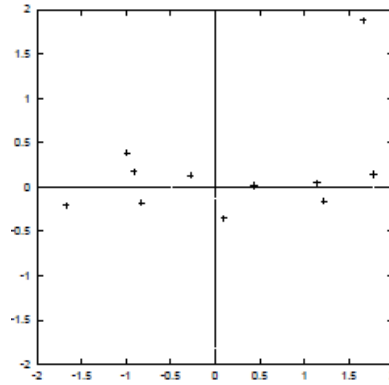
Aşağıda verilen Şekil 3.3. Şekil 3.4. ve Şekil 3.5. ile veri setine uygulanan temel bileşenler yönteminin grafik üzerindeki gösterimi verilmiştir.



Şekil 3.3. Ortalamadan çıkartılmak suretiyle normalize edilen verilerin grafik üzerinde gösterilmesi [17]



Şekil 3.4. Kovaryans matrisinin öz vektörler ve normalize edilen veri ile ifade edilmesi [17]



Şekil 3.5. Her iki özvektörün PCA yöntemi uygulandıktan sonra grafik üzerinde gösterilmesi [17]

Sayısallaştırma işleminin gerçekleştirildiği veri seti olan mevcut csv dosyası weka programında açılmıştır. Bu program sayesinde mevcut verilere PCA yönteminin uygulanılmasının ardından veriler üzerinde daha etkin çalışma, gürültüye karşı daha az hassasiyet sağlandığı görülmüştür. Kesişme ve çakışmalardan arındırılan nihai veriseti ile başarı oranında bir artış meydana gelmiştir.

3.3.2. Dengesiz veriseti

Algoritmalar, verisetlerinin dengeli bir şekilde dağıldığı varsayımı üzerinden çalıştıkları için dengeli olmayan bir veriseti gördüklerinde verilerin dağılımlarını net bir şekilde tanımlayamaz ve sonuçların büyük hata oranı göstermelerine neden olabilirler.

Dengeli olmayan sınıflandırma problemleri özellikle tıp alanındaki teşhislerde büyük problem haline gelmiştir. Negatif örnek sınıflarını yanlış tahmin etmenin yanı sıra pozitif örnekler üzerinde de yanlış tahmin oranının fazla olması dengesiz öğrenmenin varlığını ortaya koymuştur. Ek olarak daha az örnek miktarı üzerinden öğrenmeye çalışan azınlık sınıflarının durumu, algoritmaların doğru bir tahmin ortaya koyabilmeleri için güç bir ortam oluşturmuştur [18].

Bu hususta sınıfların dengesizliği ile ilgili problemleri çözmek için çeşitli yöntemler geliştirilmiştir. Sınıf dağılımlarını dengeli hale getirmek için geliştirilen yöntemlerden biri olan synthetic minority over-sampling tekniği, çalışma kapsamında ele alınmıştır [18].

3.3.2.1. SMOTE (Synthetic Minority Oversampling)

SMOTE yöntemi k-nearest neighbors algoritmasını baz alarak çalışan ve rastgele seçilen veri noktalarını en yakın komşularına bağlayarak bu alan dahilinde sentetik veriler oluşturan bir yöntem olup geniş kullanım alanına sahiptir [19].

Aşırı örnekleme tekniği olarak da ifade edilen SMOTE yönteminin avantaj ve dezavantajları aşağıda belirtilmiştir.

Avantajları

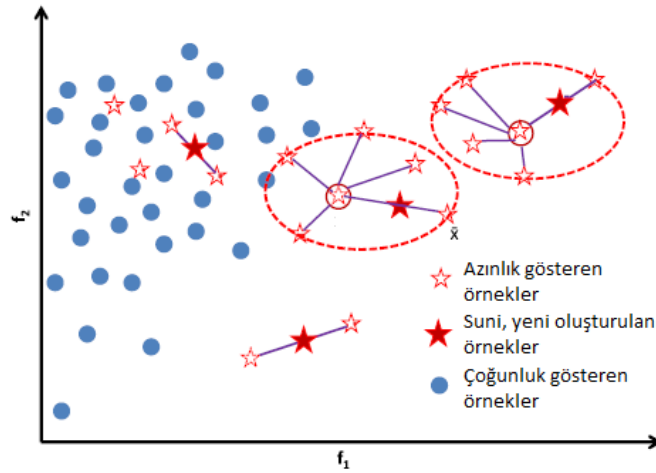
- 1-) Bilgi kaybı yoktur.
- 2-) Uygulanılması kolay bir metottur.
- 3-) Örneklerin aynısının çoğaltılmaması nedeniyle aşırı uyumu azaltır.

Dezavantajları

- 1-) SMOTE yüksek boyutlu veriler için uygun değildir.

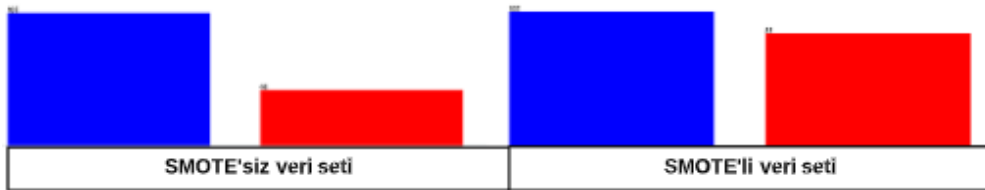
2-) Sentetik örneklerin üretildiği esnada SMOTE, komşu örneklerin diğer sınıflardan olabileceği ihtimalini göz ardı eder ve bu ihmal sınıfların üst üste gelme olasılığını yükseltmek suretiyle gürültü miktarını artırabilir.

Şekil 3.6.'da verilen grafikte, SMOTE yöntemi üzerinden sentetik verilerin üretilme durumları gösterilmiştir.



Şekil 3.6. SMOTE yöntemi üzerinden sentetik verilerin üretilmesi [18]

Bu çalışmada dengesiz veriseti, weka programı kullanılarak SMOTE yöntemi ile dengeli bir hale getirilmiş ve başarı oranında artış sağlanmıştır. Weka programında açılan verisetinin SMOTE yöntemi uygulanmadan önce ve uygulandıktan sonraki durumları Şekil 3.7.'de gösterilmektedir.



Şekil 3.7. SMOTE yöntemi uygulanılmadan önceki ve uygulandıktan sonraki veri seti

Bazı algoritmalar için verisetinin dengeli ya da dengesiz olması önem arz etmemekte iken bir çok algoritma için verilerin dengeli bir şekilde dağılması önemli bir konumda olmuş ve performans ölçütleri bakımından pozitif anlamda artışı mümkün kılmıştır.

3.4. Sınıflandırmada Kullanılan Algoritmalar

Sınıflandırma, verisetini var olan sınıflar kapsamında dağıtma işlemi iken sınıflandırma algoritmaları, mevcut eğitim kümesinin zaruri bilgileri öğrenmesinin akabinde gelen test kümesine bu bilgileri uygulayarak test kümesindeki verilerin sınıfını doğru bir şekilde tahmin etme sürecini sağlayan kısımdır.

Bu çalışmada sayısallaştırılan görüntülerin üzerinde gerekli iyileştirmeler yapıldıktan sonra sınıflandırma algoritmaları tarafından analiz edilebilir duruma gelmiştir. Naif Bayes (Naive Bayes), Destek Vektör Makineleri (Support Vector Machine), Rastgele Orman (Random Forest) ve Karar Ağaçları (Decision Tress-J48) algoritmalarına tabi tutulan veriler, cross validation geçerliliği yöntemi ile sınıflandırıldıktan sonra karşılaştırmaları yapılmıştır.

3.4.1. Naif Bayes algoritması (Naive Bayes algorithm)

Adını matematikçi Thomas Bayes'den alan sınıflandırma algoritması, istatistiksel çalışma mekanizmasına sahip olduğu için istatistiksel değerlere göre değişen bir sınıflandırma metodudur. Olasılık durumlarına göre sınıflandırma işlemi uygulanan ve veri kümeleri arasında bir ilişki kabul etmeyen bu metot, arka planda bayes teoremini baz alarak çalışmaktadır [20].

3.4.1.1. Bayes teoremi

Bayes teoremi olayların meydana gelme olasılıklarını çeşitli faktörler kapsamında değerlendirerek koşullu olasılıkların hesaplanması suretiyle sonuca ulaşmayı amaçlayan bir kavramdır.

Bu bağlamda birbirinden bağımsız olarak X ve Y şeklinde isimlendirilen iki olay düşünüldüğünde birinci olayın meydana gelmesi durumunda ikinci olayın meydana gelmesi olasılığı $P(X \cap Y)$ şeklinde belirtilir. Aynı zamanda Denklem 3.3.'te gösterildiği gibi mevcut bir ifadesi de bulunmaktadır.

$$P(X \cap Y) = P(X|Y)P(Y) = P(Y|X)P(X) \quad (3.3)$$

Bu teoremin eşiti Denklem 3.4.'te verilmiştir.

$$P(Y|X) = P\left(\frac{P(X|Y)P(Y)}{P(X)}\right) \quad (3.4)$$

Yukarıdaki denklemlerde görülen $P(Y|X)$ ifadesi, X 'in doğru olduğu göz önüne alındığında Y 'nin doğru olma olasılığını, $P(X|Y)$ ifadesi Y 'nin doğru olduğu göz önüne alındığında X 'in doğru olma olasılığını, $P(X)$ ifadesi X 'in doğru olma olasılığını, $P(Y)$ ifadesi ise Y 'nin doğru olma olasılığını belirtmektedir. Bu ifadeler kapsamında bayes teoreminin Naive Bayes sınıflandırma algoritması üzerinde adım adım ilerleyişi aşağıda verildiği gibi çalışmaktadır.

1-) Verisetinin tanımlama grubundan ve ilişkili sınıf etiketlerinden oluştuğu göz önüne alındığında, her bir tanımlama grubunun N boyutlu bir özellik vektörü tarafından temsil edildiği bilinir. N boyutlu özellik vektörü, $X=(x_1, x_2, \dots, x_n)$ şeklinde tanımlama grubunu ve sırasıyla A_1, A_2 ve A_n şeklinde tanımlama grubunun ölçümlerini belirtmektedir [20].

2-) Bu kapsamda C_1, C_2 ve C_m şeklinde m adet sınıfın mevcut olduğu varsayalım. X tanımlama grubu verildiğinde Naive Bayes sınıflandırıcı, X tanımlama grubunun en yüksek olasılığa sahip olduğu sınıfı tahmin edecektir. Bu sebepten dolayı $P(C_i|X)$ ifadesinin maksimum hale getirilmesi gerekmektedir [20].

3-) Denklem 3.5.'te verilen Bayes teoreminde görüldüğü üzere, $P(C_i|X)$ ifadesinin maksimum hale getirilebilmesi için payda kısmında yer alan $P(X)$ ifadesinin sabit bir ifade olması sebebi ile pay kısmında yer alan $P(X|C_i)P(C_i)$ ifadesinin maksimuma çıkartılması gerekmektedir [20].

$$P(C_i|X) = P\left(\frac{P(X|C_i)P(C_i)}{P(X)}\right) \quad (3.5)[20]$$

4-) Bu hususta birçok özelliği olan veri kümelerine bakıldığında Denklem 3.5.'te bulunan $P(X|C_i)$ kavramının değerlendirilmesi yapılırken sınıf şartlı bağımsızlık varsayımı oluşturulur. Grubun sınıf etiketleri göz önüne alındığında özellik değerlerinin birbiri ile ilişki içerisinde olmadığı varsayımının matematiksel hali elde edilir ve Denklem 3.6.'da ifade edilmektedir. Bu denklemde X_k , tanımlama grubunu; C_i ilgili sınıfı ve n , tanımlama grubunun toplam sayısını ifade etmektedir [20].

$$\begin{aligned} P(X|C_i) &= \prod_{k=1}^n P(X_k|C_i) \\ &= P(x_1|C_i) * P(x_2|C_i) * \dots * P(x_n|C_i) \end{aligned} \quad (3.6)[20]$$

5-) X durumunun sınıf etiketini tahmin edebilmek amacıyla her C_i sınıfı için $P(X|C_i)P(C_i)$ değerlendirilmesi yapılır ve nihai karar elde edilir [20].

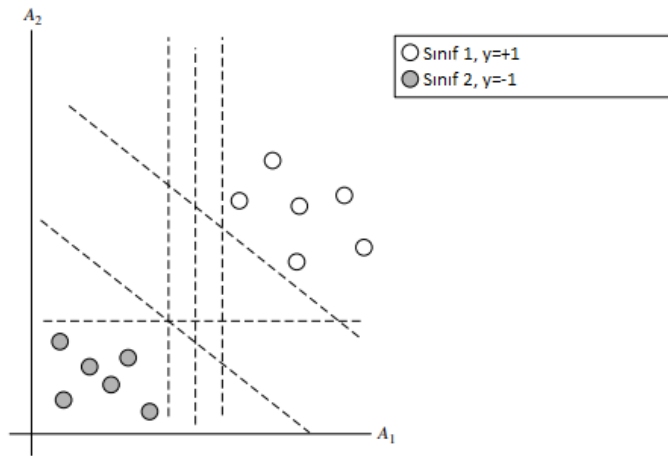
Bu çalışmada sayısallaştırılan veriler bahsedilen yöntemler ile iyileştirildikten sonra Naive Bayes algoritmasına tabi tutulmuş ve genel anlamda performans ölçütleri ile başarı oranında artışların sağlandığı görülmüştür.

3.4.2. Destek Vektör Makineleri algoritması (Support Vector Machine algorithm)

İstatistiksel öğrenme teorisine dayalı, kontrollü ve etkili sınıflandırma algoritmalarından olan Destek Vektör Makineleri algoritması; ilk başta iki sınıflı doğrusal veriler için kullanılmış daha sonra doğrusal olmayan veriler için de geliştirilmiştir [21]. Bu algoritma, ayrılması gereken iki sınıfı birbirinden en makul şekilde ayırarak istenilen formatta elde edilmesi temeline dayanır ve çok sayıda bağımsız değişken ile çalışmayı mümkün kılar.

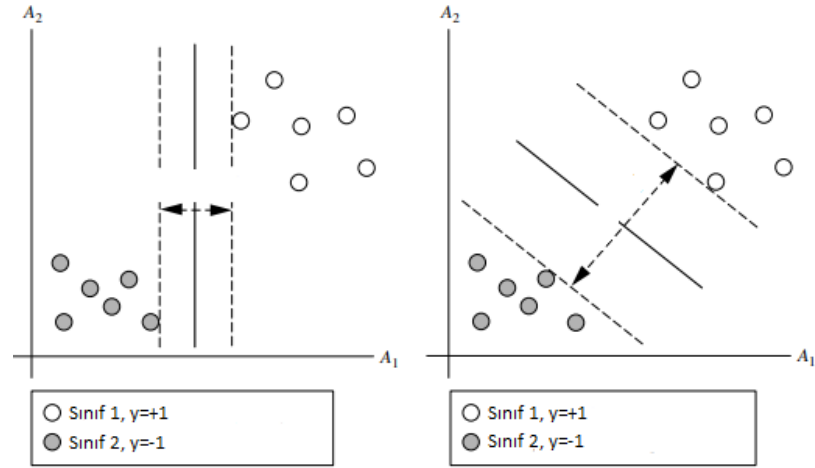
3.4.2.1. Doğrusal olarak ayrılması mümkün veriler için Destek Vektör Makineleri

Doğrusal olarak ayrılması mümkün olan iki sınıflı bir veri setinde verilerin $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ olarak dağıldığı varsayılır. Bu veri setinde x_i , eğitim gruplarını ifade ederken; y_i , +1 ya da -1 şeklinde mevcut iki sınıftan sadece bir sınıfın seçilmesi şeklinde bir sonuç döndürmektedir. Şekil 3.8.'de verilerin doğrusal olarak ayrılmasını mümkün kılan bir grafik verilmiştir [20].



Şekil 3.8. Doğrusal olarak ayrılabilen veriler için Destek Vektör Makineleri [21]

Destek Vektör Makineleri algoritmasında verileri ayırmak için sonsuz sayıda ayırma çizgisi çizilebilir. Ancak minimum sınıflandırma hatasına sahip bir ayırma düzlemi elde etmek istediğimiz için SVM'in en iyi karar düzlemini yani hiper düzlemini bulmak amaçlanır. Şekil 3.9.'da gösterilen mevcut grafiklerde ayırıcı iki hiper düzlem çizilmiştir. Bu hiper düzlemlerden ilki daha küçük kenar boşluklarına sahipken ikincisi daha geniş kenar boşluklarına sahiptir. Bu genişlikler dikkate alındığında yeni veriler için büyük kenar boşluklarına sahip hiper düzlemin küçük kenar boşluklarına sahip hiper düzlemden daha doğru yanıtlar vererek verileri sınıflandırmasının mümkün olduğu yorumlanmaktadır [20].



Şekil 3.9. Mesafelerine göre doğrusal olarak ayrılabilen veriler için Destek Vektör Makineleri [20]

Bu kapsamda ayırıcı hiper düzlem Denklem 3.7.'de, destek vektörleri Denklem 3.8.'de ifade edilmektedir.

$$W.X+b=0 \quad (3.7) [20]$$

$$W.X+b=\pm 1 \quad (3.8) [20]$$

Yukarıdaki denklemlerde görülen X , R^N 'in elemanı olup N boyutlu bir uzayı; $\{-1, +1\}$, sınıf etiketlerini; W , ağırlık vektörünü ve b , eğilim değerini göstermektedir [21].

İki giriş niteliği olarak x_1 ve x_2 özelliklerini ele aldığımızı varsayalım. Bu durumda 3.7. denklemini yeniden yazdığımızda çıkan eşitlik, Denklem 3.9.'da görüldüğü gibi ifade edilmektedir [20].

$$w_0+w_1x_1+w_2x_2=0 \quad (3.9) [20]$$

Bu bağlamda herhangi bir noktanın ayırıcı hiper nokta üzerinde bulunmasının ifadesi Denklem 3.10.'da verilmektedir [20].

$$w_0+w_1x_1+w_2x_2>0 \quad (3.10) [20]$$

Benzer şekilde herhangi bir noktanın ayırıcı hiper düzlemin altında bulunmasının ifadesi Denklem 3.11.'de verilmektedir [20].

$$w_0 + w_1x_1 + w_2x_2 < 0 \quad (3.11) [20]$$

Ağırlıklar yeniden ayarlandığında marjinin kenarları şeklinde tanımlanan hiper düzlem, +1 ve -1 sınıfları için Denklem 3.12. ve Denklem 3.13.'te görüldüğü üzere yeniden yazılabilmektedir [20].

$$H_1 = w_0 + w_1x_1 + w_2x_2 \geq 0 \quad y_i = +1 \text{ için} \quad (3.12) [20]$$

$$H_2 = w_0 + w_1x_1 + w_2x_2 \leq 0 \quad y_i = -1 \text{ için} \quad (3.13) [20]$$

Verilen 3.12. ve 3.13. denklemleri birleştirilmek suretiyle ifade edildiğinde Denklem 3.14. elde edilir [20].

$$y_i(w_0 + w_1x_1 + w_2x_2 \geq 1), \quad \forall_i \quad (3.14) [20]$$

H_1 ya da H_2 hiper düzlem üzerinde yer alan tanımlama grupları, destek vektörleri şeklinde isimlendirilir ve sınıflandırmaya ilişkin en çok bilgiyi veren tanımlayıcı gruplardır.

Aynı zamanda ayırıcı hiper düzleme paralel olan destek vektörünün ayırıcı hiper düzlem ile arasındaki mevcut uzunluğu, $1/|W|$ kadarlık bir mesafeden oluşmaktadır. Dolayısıyla destek vektörlerinin sağ ve sol kenarlarında bulunduğu düşünülen maksimum aralığın uzunluğu, $2/|W|$ kadarlık bir mesafeye denk gelmektedir [20].

Bu hususta tanımlama grubunun hiper düzlemin hangi tarafına düştüğünün araştırılması üzerine bir örnek sunulduğunda işaretin pozitif olduğu durumlarda tanımlama grubunun ayırıcı hiperdüzlemin üzerine düştüğü ve Destek Vektör Makinası tarafından test tanımlama grubunun +1 sınıfına ait olduğu, işaretin negatif olduğu durumlarda ise tanımlama grubunun ayırıcı hiper düzlemin altına düştüğü ve

Destek Vektör Makinesi tarafından test tanımlama grubunun -1 sınıfına ait olduğu tahmin edilmektedir [20].

3.4.2.2. Doğrusal olarak ayrılması mümkün olmayan veriler için Destek Vektör Makineleri

Doğrusal olarak bir ayrımı mümkün kılan veri setlerinde optimum hiper düzlem seçebilmek için $\|w\|$ değişkeninin minimum hale getirilmesi gerekmektedir. Bu bağlamda en uygun hiperdüzlemi belirlemek, aşağıda ifade edilen optimizasyon probleminin çözümüne bağlıdır [21].

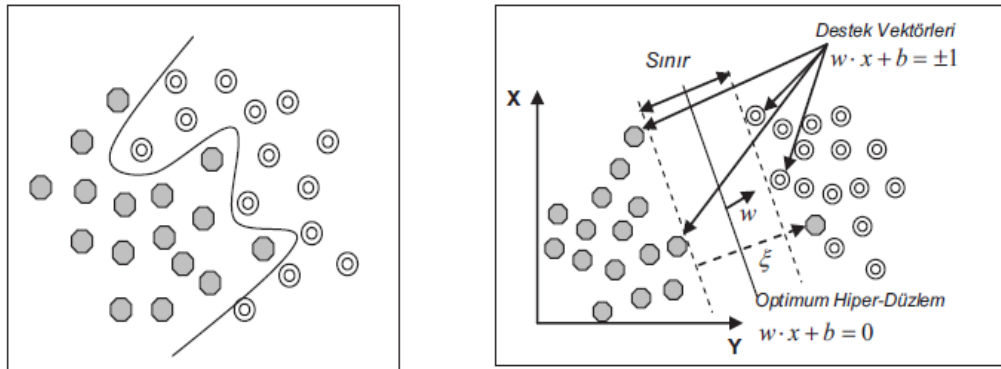
$$\min \left[\frac{1}{2} \|w\|^2 \right] \quad (3.15) [21]$$

Ancak verilerin daima doğrusal bir biçimde ayrılması her zaman mümkün değildir. Bu tür durumlarda verilerin mevcut düzlemden daha yüksek boyuta sahip bir düzleme aktarılıyormuşçasına suni bir değişkenin ilave edilmesiyle incelenmesi gerekmektedir. Doğrusal olarak ayrımı mümkün olmayan bu verilerin değişken ilavesi ile çözümü Denklem 3.16.'da verilmiştir [21].

$$\min \left[\frac{1}{2} \|w\|^2 + C \cdot \sum_{i=1}^r \varepsilon_i \right] \quad (3.16) [21]$$

Yukarıdaki ifadede ε_i değişkeni, pozitif suni bir değişkendir. Bu değişken eğitim verilerinin bir kısmının hiper-düzlemin diğer tarafında kalmasından doğan problemin çözümü için gerekli olmaktadır. C değişkeni bir düzenleme parametresidir, yanlış sınıflandırmadan doğan hataların minimum hale getirilmesi anlamına gelir ve pozitif değerler alır.

Şekil 3.10.'da doğrusal olarak ayrılan ve doğrusal olarak ayrılamayan veri kümeleri için hiper düzlemin belirlenmesi ile ilgili iki grafik verilmiştir.



Şekil 3.10. Doğrusal bir şekilde ayrılabilen ve ayrılamayan veri kümeleri için hiper düzlemin belirlenmesi [21]

Ek olarak Destek Vektör Makineleri doğrusal bir şekilde ayırım imkanı sağlamayan veri kümeleri için kernel fonksiyonları kullanmak suretiyle yüksek boyutta ayırım imkanı da sunmuştur. Bu fonksiyonlardan bazıları; Polinom Kerneli, Normalleştirilmiş Polinom Kerneli, Radyal Tabanlı Fonksiyon Kerneli ve Pearson VII Kerneli olup Şekil 3.11.'de ifade edilmiştir [21].

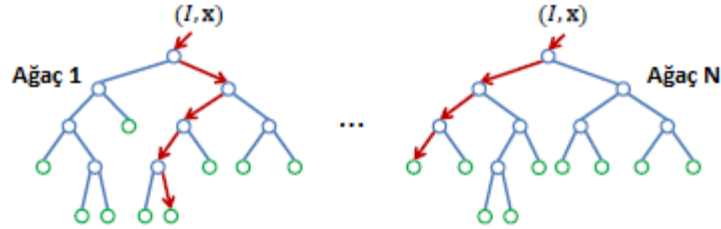
Kernel Fonksiyonu	Matematiksel İfadesi	Parametre
Polinom Kerneli	$K(x,y)=((x \cdot y)+1)^d$	Polinom Derecesi(d)
Normalleştirilmiş Polinom Kerneli	$K(x,y)=((x \cdot y)+1)^d / (((x \cdot x)+1)^d ((y \cdot y)+1)^d)^{1/2}$	Polinom Derecesi(d)
Radyal Tabanlı Fonksiyon Kerneli	$K(x,y)=e^{-\gamma(x-x_j)^2}$	Kernel Boyutu(γ)
Pearson VII (PUK) Kerneli	$K(x,y)=1/[1+((2(x-y ^2 (2^{(1/w)}-1)^{1/2})/a)^2)^w]$	Pearson Genişliği Parametreleri (a,w)

Şekil 3.11. Destek Vektör Makineleri için kullanılan bazı kernel fonksiyonları [21]

Bu çalışmada sayısallaştırılan veriler bahsedilen yöntemler ile iyileştirildikten sonra Destek Vektör Makinesi algoritmasına tabi tutulmuş ve genel anlamda performans ölçütleri ile başarı oranında artışların olduğu görülmüştür.

3.4.3. Rastgele Orman algoritması (Random Forest algorithm)

2001 yılında Leo Breiman tarafından geliştirilmiş olan Random Forest algoritması, denetimli bir sınıflandırma algoritması olup eğitilen karar ağaçları topluluğundan meydana gelmektedir. Şekil 3.12.'de verilen görüntü, her biri bölünmüş ve yaprak düğümlerinden oluşan T adet karar ağacının meydana getirdiği ormanı göstermektedir. Bu ormanda kökten başlamak suretiyle t eşiği üzerinden sağa ve sola dallanmalar yaparak istenilen özellik üzerinden öğrenilmiş bir dağılım depolanmaktadır [22].



Şekil 3.12. Rastgele Orman Algoritmasına ait ağac yapısı [22]

Algoritma, ağaçları gelişigüzel bir şekilde büyütmenin yanı sıra parçalara bölme sürecinde de istenilen özelliği araştırmak yerine rastgele bir özelliğin mevcut kümesi içerisinde istenilen özelliği bulmak suretiyle çalışan ve bu şekilde daha iyi bir model elde eden geniş bir yelpazedir. Bu bakımdan sınıflandırma sırasında bağımsız ve birden fazla karar ağacı üretmek için bölümlenmeden faydalanma özelliği göstermektedir [23].

Rastgele orman algoritması kullanıcı tarafından 2 parametrenin tanımlanmasıyla başlar. Bu parametrelere ek olarak geliştirilecek ağaç sayısı (N) ve m olarak ifade edilen değişken sayısı, düğümde kullanılmak amacıyla edilir. %66'sı eğitim için ayrılan verilerin her bir düğümündeki m değişkenleri var olan bütün değişkenler arasından rastgele olarak seçilip en iyi dalın belirlenmesini sağlar. Rastgele orman algoritması ağaç üretmek için CART (Classification and Regression Tree) algoritmasını kullanır. Algoritma kullanılırken belli bir ölçüt göz önüne alınarak bölünme yapılır ve seçme işlemi gerçekleştirilir. Bu hususta bölünme esnasında

düğümün homojen olarak dağılması istendiğinden homojenlik ölçümünde CART algoritmasının kriteri olan GINI indeksi hesaplanır. Sınıf homojenliğini ölçen GINI indeksinin formülü Denklem 3.17.'de ifade edilmiştir [23].

$$\sum_{j \neq i} (f(C_i, T)/|T|)(F(C_i, T)/|T|) \quad (3.17) [23]$$

Yukarıda verilmiş mevcut denklemde T değişkeni eğitim verisetini, C_i değişkeni rasgele seçilen özelliğin ait olduğu sınıfı, f(C_i,T)/|T| değişkeni ise seçilen örnek için C_i sınıfına ait olma durumunu göstermektedir. İşlemler yapıp ağırlıklı GINI değeri hesaplandıktan sonra bölünme kriteri belirlenmiş olmaktadır. GINI indeksi ile sınıfın homojenliğinin ters orantılı olduğu bilinen Rastgele Orman algoritmasında, GINI indeksi azaldıkça sınıfın homojenliği artmakta ve alt düğümün GINI indeksinin üst düğümün GINI indeksinden daha az olması o dalın başarılı olarak kabul görmesini sağlamaktadır [23].

Bu bağlamda inşa edilen ağaç toplulukları vasıtasıyla girdi verisinden yeni bir nesne sınıflandırılmak istendiğinde girdi verisi ormanda mevcut tüm ağaçlara yerleştirilir ve mevcut tüm ağaçlardan bir sınıflandırma sonucu elde edilir. GINI indeksi sıfıra ulaşır tüm ağaçlar üretildikten sonra aralarında oylama gerçekleştirilir ve en çok oyu alan bir sınıfa atanır [23].

Farklı dallar için kullanım alanı oluşturan rastgele orman algoritmasının tıp dünyasında; hastalıkları tanımlamak, ilaç bileşenleri üzerinde seçimler yapmak gibi birçok hususta yeri vardır. Aşağıda bu algoritma ile ilgili bilgiler maddeler halinde sıralanmıştır.

- 1- Esnek bir algoritmadır.
- 2- Kısa süren model oluşturma sürecinden dolayı, zamanın kısıtlı olduğu durumlar için mükemmel bir tercih olarak karşımıza çıkmaktadır.
- 3- Regresyon ve sınıflandırma aşamaları için kullanılabilir olması algoritmaya pozitif bir fayda sağlamaktadır.

- 4- Öngörü çerçevesine yerleşmiş, açıklama sağlamayan bir algoritma durumundadır.
- 5- Hızlı ve basit olma özelliği göstermesi, birçok alanda sıklıkla kullanılabiliyor olmasını sağlamıştır.

Bu çalışmada sayısallaştırılan veriler bahsedilen yöntemler ile iyileştirildikten sonra rastgele orman algoritmasına tabi tutulmuş ve genel anlamda performans ölçütleri ile başarı oranında artışların olduğu görülmüştür.

3.4.4. Karar Ağaçları algoritması (Decision Trees algorithm)

Karar Ağacı algoritması, ağaç yapısı şeklinde oluşum gösteren bir sınıflandırma algoritmasıdır. Kuralları kolay bir şekilde ifade edebilmesi, kategorik ve numerik veriler ile çalışabilmesi bu algoritmaya basit ve yaygın bir metot olma özelliği sunmuştur. Karar ağaçları, mevcut verisetini çeşitli karar kurallarına tabi tutarak daha küçük veriseti gruplarına bölmek suretiyle çalışan bir algoritmadır. Karar ağaçlarının temelini oluşturan iç karar düğümleri, dal ve uç yapraklar terimleri aşağıda tanımlanmıştır.

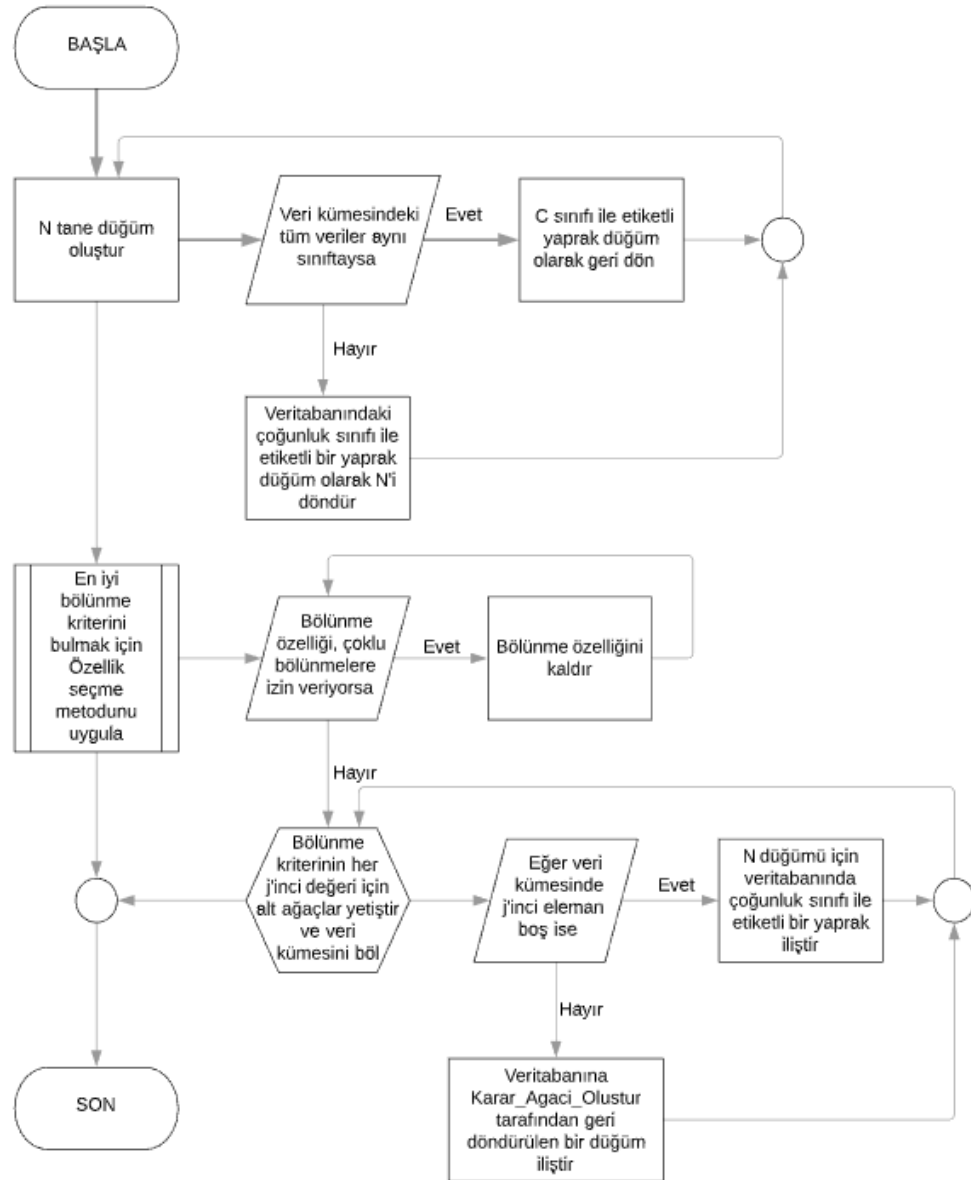
İç karar düğümleri: Mevcut şart ifadelerini baz alarak verilerin test edildiği ve sonuca göre dallanacağı yer ile ilgili nihai kararların verildiği kısımdır.

Dal: Düğümler arasındaki ilişkilerin bağlantı kısmıdır.

Uç yapraklar: Verilerin test edilmediği son kısımdır.

Ağaç yapısı şeklinde sınıflandırma özelliğini gösteren Karar Ağacı algoritmasının akış diyagramı Şekil 3.13.'te verilmiştir.

Algoritma: Karar_Agaci_Olustur (Veri kümesinin eğitim setinden bir karar ağacı oluştur)



Şekil 3.13. Karar Ağacı algoritmasının kaba kodu [20]

Karar Ağacı algoritmasının avantajları aşağıda belirtilmiş olup akabinde Karar Ağacı algoritması kapsamında bu çalışmada kullanılan algoritmalarından biri olan ID3 algoritmasına açıklık getirilmiştir.

- 1-) Tercih edilen bir algoritmadır.
- 2-) Ürettiği kuralların anlaşılabilirliği kolay bir şekilde sağlanmaktadır.

- 3-) Kategorik ve numerik verilerle işlem yapabiliyor olması algoritma için ekstra bir avantaj sağlamaktadır.
- 4-) Tümevarım mantığını kullanan basit bir metottur.
- 5-) Ayırık değerlerle çalışabilmesi gürültüye karşı çok dayanıklı olmasını sağlamıştır.

3.4.4.1. Tekrarlı ikilikçi ağaç algoritması (Iterative dichotomiser 3 algorithm)

Makine öğrenmesi araştırmacısı olan J. Ross Quinlan, ID3 olarak bilinen bir karar ağacı algoritması geliştirdi. Akabinde öğrenme algoritmalarını karşılaştırmak için sıklıkla kullanılan ID3'ün ardılı şeklinde gelişim gösteren C4.5 algoritmasını sundu. Bu sırada L. Breiman, J. Friedman, R. Olshen ve C. Stone adlı araştırmacılar da ikili karar ağaçları üzerindeki çalışmalarını gösteren Sınıflandırma ve Regresyon Ağaçları (CART) kitabını yayınladılar. Bahsedilen ID3 ve CART algoritmaları aynı süreçte bağımsız olarak üretilmiş benzer yaklaşımlar üzerinden inşa edilen algoritmalar olma özelliği sergilemektedirler. Aynı zamanda karar ağacı algoritması altında mevcut olan ID3, C4.5 ve CART isimli tüm algoritmalar; yukarıdan aşağıya yinelenen böl ve fethet tarzında bir yaklaşım sergileyen algoritmalar kapsamında değerlendirilme özelliği göstermektedir [20].

ID3 algoritması, baştan başlayarak sona doğru çalışan ve kolay bir şekilde kavranabilen bir algoritmadır. Belirsizliğin ortaya çıkma ihtimali için entropi kavramı vasıtasıyla sınıflandırma hususunda çözüm sunar. Bu bağlamda Denklem 3.18.'de ifade edilen Shannon'un entropi formülünde S harfi mevcut durumu, a harfi entropisi hesaplanacak mevcut durumların adetini, p_i harfi ise i'inci durumun olasılığını göstermektedir.

Aynı zamanda Denklem 3.19.'da belirtilen Bilgi Kazanımı(S,A) ifadesi, A niteliğinin bilgi kazancını ifade etmekte olup sistemin toplam entropi değerinden herbir özelliğın entropi değerinin tek tek çıkarılması ile elde edilir. Bu kapsamda kök düğümünden itibaren belirlenen veri kümelerinin her biri için gerekli hesaplar yapılarak karar ağacı elde edilmektedir. Algoritmanın daha net bir şekilde anlaşılabilmesi için bahsedilen bu işlemlerin Şekil 3.14.'te verilen veriseti ile belirli durumlar baz

alınarak belirli bir amaç doğrultusunda karar verilmesi işlemleri bir örnek kapsamında karar ağacı üzerinde somut bir şekilde uygulanarak Şekil 3.15.'te ayrıntılı bir şekilde gösterilmiş ve sürecin anlatımı gerçekleştirilmiştir.

$$\text{Entropi}(S) = -\sum_{i=1}^a p_i \log_2 p_i \quad (3.18)$$

$$\text{Bilgi Kazanım}(S, A) = \text{Entropi}(S) - \sum_{V \in \text{Değerler}(A)} \frac{|S_V|}{|S|} \text{Entropi}(S_V) \quad (3.19)$$

Sıra	Yaş	Gelir	Öğrenci	Kredi Oranı	Bilgisayar Alındı Mı?
1	Genç	Yüksek	Hayır	Makul	Hayır
2	Genç	Yüksek	Hayır	Mükemmel	Hayır
3	Orta Yaş	Yüksek	Hayır	Makul	Evet
4	Yaşlı	Orta	Hayır	Makul	Evet
5	Yaşlı	Düşük	Evet	Makul	Evet
6	Yaşlı	Düşük	Evet	Mükemmel	Hayır
7	Orta Yaş	Düşük	Evet	Mükemmel	Evet
8	Genç	Orta	Hayır	Makul	Hayır
9	Genç	Düşük	Evet	Makul	Evet
10	Yaşlı	Orta	Evet	Makul	Evet
11	Genç	Orta	Evet	Mükemmel	Evet
12	Orta Yaş	Orta	Hayır	Mükemmel	Evet
13	Orta Yaş	Yüksek	Evet	Makul	Evet
14	Yaşlı	Orta	Hayır	Mükemmel	Hayır

Şekil 3.14. Karar ağacı algoritmasına uygulanacak veri seti [21]

$$\text{Entropi}(S) = -9/14 \log_2 9/14 - 5/14 \log_2 5/14 = 0.940$$

$$\text{Entropi}_{\text{Yaş}}(S) = 5/14 \times (-2/5 \log_2 2/5 - 3/5 \log_2 3/5)$$

$$+ 4/14 \times (-4/4 \log_2 4/4)$$

$$+ 5/14 \times (-3/5 \log_2 3/5 - 2/5 \log_2 2/5) = 0.694$$

$$\text{Bilgi Kazanımı}(\text{Yaş}) = \text{Entropi}(S) - \text{Entropi}_{\text{Yaş}}(S) = 0.940 - 0.694 = 0.246$$

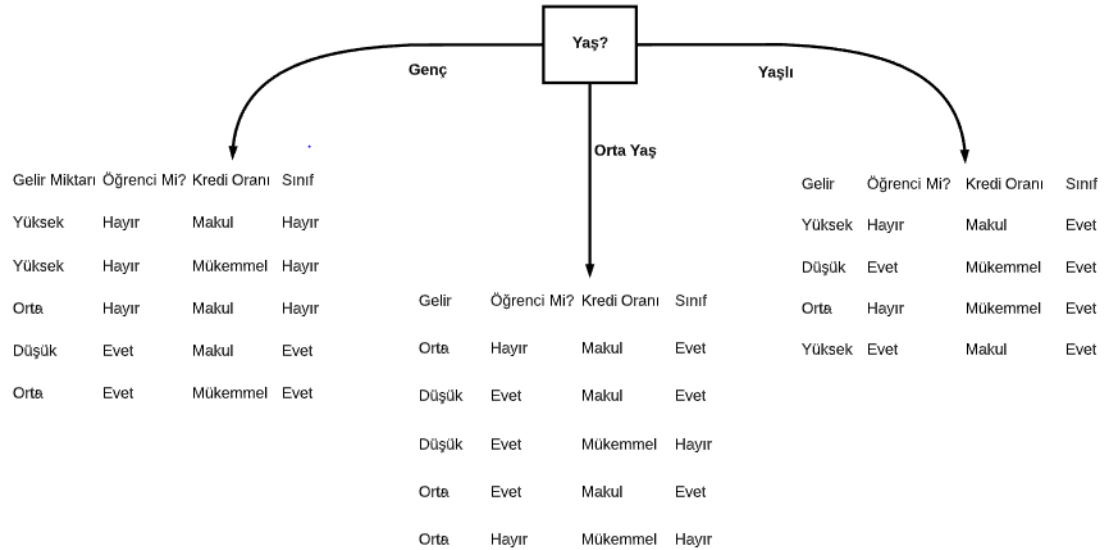
Benzer şekilde diğer kriterler için de hesaplamalar yapıldıktan sonra aşağıda belirtilen sonuçlar elde edilmektedir.

$$\text{Bilgi Kazanımı}(\text{Gelir Miktarı}) = 0.029$$

$$\text{Bilgi Kazanımı}(\text{Öğrenci Mi?}) = 0.151$$

$$\text{Bilgi Kazanımı}(\text{Kredi Oranı}) = 0.048$$

Sonuçlar arasından yaş en yüksek bilgi kazanımı değerine sahip olan bir kriter olduğu için karar ağacının öncelikle yaş kriterine göre bölüneceği yorumu yapılabilmektedir. Diğer kriterlerin de belirtildiği gibi hesaplanıp karar ağacında yerleştirilmesi suretiyle ağaç tamamlanmış olacaktır.

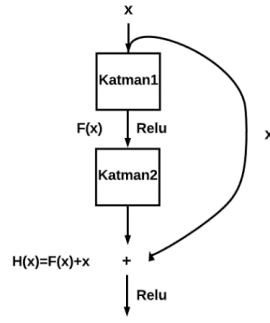


Şekil 3.15. Veri setinin uygulanılması sonucunda elde edilen karar ağacı [21]

Bu çalışmada sayısallaştırılan veriler bahsedilen yöntemler ile iyileştirildikten sonra karar ağacı algoritması altında bulunan ID3 algoritmasına tabi tutulmuş ve genel anlamda performans ölçütleri ile başarı oranında artışların olduğu görülmüştür.

3.4.5. RESNET Mimarisi

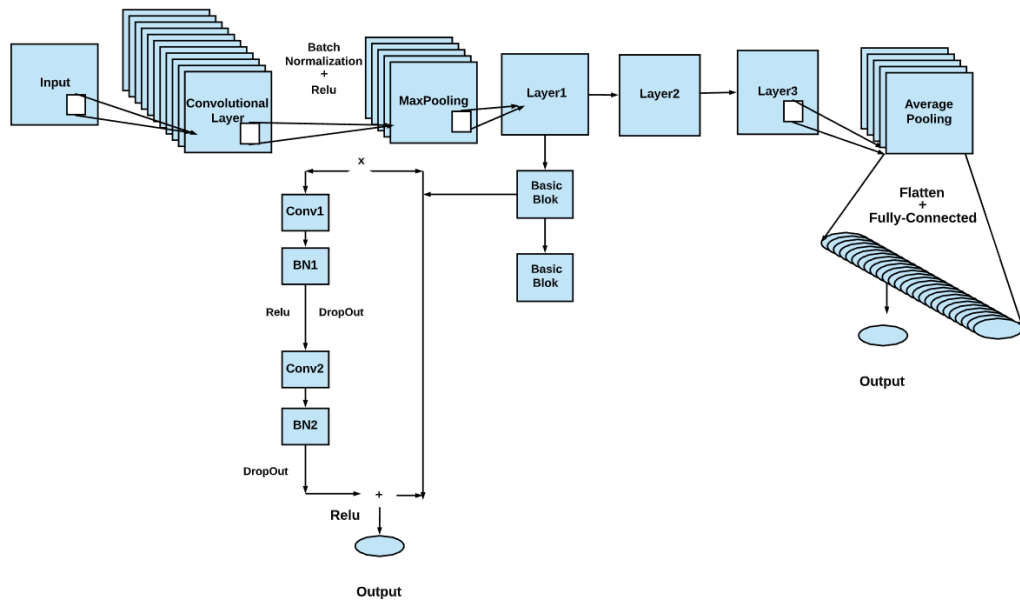
Resnet Mimarisi, veri setindeki artışa bağlı olarak geleneksel algoritmalar ile çözülemeyen durumları derin öğrenme kapsamında çözen Convolutional Neural Network mimarisinin gradyanların kaybolması problemine karşı sunulmuş bir yanıttır. Gradyanların kaybolması sinir ağının öğrenememesi ve doğru bir şekilde sınıflandıramamasına neden olan problemdir. Bu bağlamda verilerin sınıflandırıldığı Resnet mimarisinde her bir katman sonraki katmanın girişi olacak şekilde çalışır. Katmanlarda bulunan basic blokların içerisindeki süreçlere Şekil 3.16.'da yer verilmiştir [24].



Şekil 3.16. Resnet mimarisi içerisinde sunulan blok yapısı [25]

Resnet mimarisinde bulunan bloklar $H(x)=F(x)+x$ formülasyonunun sonucunu hesaplamakta olup denklemde belirtilen $F(x)+x$ ifadesi shortcut bağlantılı ileri beslemeli sinir sistemleri tarafından ayırt edilebilen bir durumdur [24]. Shortcut bağlantıları, bir veya daha fazla katman atlayarak kimlik haritalaması yapan ve sonucu evrişim işlemlerine tabi tutarak elde eden çıktılara eklemek üzerine kurulu bir yapıdır. Aynı zamanda shortcut bağlantıları herhangi bir hesaplama karmaşıklığı ya da parametre barındırmamaktadır [25].

Çalışmada kullanılan Resnet Mimarisinin içerisinde yer verilen süreçler Şekil 3.17.'de gösterilmiştir.



Şekil 3.17. Çalışmada kullanılan Resnet Mimarisi içerisinde yer verilen süreçler

Sunulan çalışmada verisetine uygulanılan Resnet Mimarisi, geleneksel algoritmalara kıyasla başarı oranında pozitif yönde bir artış sağlamaktadır.

3.5. Analizler

Analiz, bir bütünü oluşturan parçaları ayırmak ve bu parçalardan hareketle durumun niteliği ya da niceliği hakkında bilgi sahibi olmak şeklinde temellenen kavramdır. Bu çalışmada iyileştirilmiş görüntüler üzerinden aşama aşama çözümlenmeler yapılmıştır.

3.5.1. Model değerlendirme yöntemleri

Model değerlendirme yöntemleri, modeller arasında seçim yapmak suretiyle en iyi modeli bulmak ve gelecekte ne derece iyi çalışacağını belirlemek için kullanılır. Aynı zamanda aşırı uyumdan (overfitting) sakınmak için eğitim (training) kümesinin yanında test (test) kümesini kullanmak, ilgili çalışma için makul olacaktır. Yöntemlerden biri olan Hold-Out aşağıda açıklanmaktadır

3.5.1.1. Hold-Out yöntemi

Çalışmada mevcut olan veri kümesinin eğitim verisi ve test verisi şeklinde ayrılması ile 2 farklı veri seti elde etmeyi mümkün kılan bu metot, veri kümesinin daha doğru bir şekilde değerlendirilmesini sağlar.

3.5.1.1.1. Eğitim verisi (Training data)

Algoritmaya, algoritmanın öğrenmesi için verilen bir veri kümesidir. Algoritma bu veri kümesi üzerinden model oluşturmak suretiyle öğrenmeyi sağlamaktadır.

3.5.1.1.2. Test verisi (Test data)

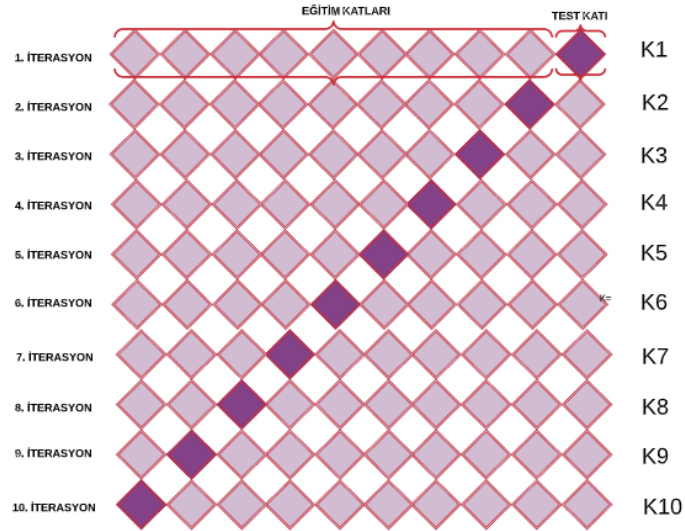
Algoritmanın oluşturduğu modelin doğruluğunu test etmek için kullanılmakta olan bir veri kümesidir.

3.5.1.2. Çapraz Geçerlilik yöntemi (Cross validation method)

Bir tahmin sisteminin uygulanması aşamasında, sistemin ne kadar doğrulukla çalışacağına bulunması üzerine inşa edilen bir tekniktir. Bu bağlamda ilgili çalışmanın performansı sadece uygulanan algoritmalara bağlı olmamakla birlikte eğitim ve test kümesinin örnekleri ile de ilişkili olabilmektedir. Bahsi geçen model performans değerlendirme tekniği, veri kümesi üzerinde örnekleme işleminin ne şekilde yapılması gerektiğine dair bilgi verir. Performans değerlendirme ölçütlerinden biri olan Çapraz Geçerlilik yöntemi, Denklem 3.20.'de verilen formül üzerinden uygulanmaktadır. Bu denklemde N, iterasyon sayısını ve K_i , ele alınan i'inci parçayı ifade etmektedir.

$$K = \frac{1}{N} \sum_{i=0}^N (K_i) \quad (3.20)$$

Tüm verinin kullanılmasını sağlamak amacıyla bir model oluşturmak için uygulanan bir yöntemdir. K adet parçaya bölünen verinin bir kısmı test, geri kalanı ise eğitim için kullanılır. Süreç ayrılan eğitim verisi üzerinden model oluşturup test verisi üzerinden ilgili modelin denenmesi ile şekillenir. Bu işlem de her defasında k kere model seçilir ve seçilenler arasından en uygun olanı alınır. İterasyon sayısının k defa olması, hata oranının düşük bir oran çerçevesinde seyretmesini mümkün kılar. Çapraz Geçerlilik yöntemi Şekil 3.18.'te gösterilmektedir.



Şekil 3.18. İterasyon sayısını 10 alınması suretiyle veri setine Çapraz Geçerlilik yönteminin uygulanılması

3.6. Performans Değerlendirme Ölçütleri

Sınıflandırma algoritmalarına tabi tutulan çalışmalardan elde edilen başarı oranı, tek başına bir ölçüt olarak baz alındığında sağlıklı bir bilgi vermemektedir. Bu nedenle algoritmalar kullanılarak sınıflandırılan veri setini değerlendirmek amacıyla TP oranı, FP oranı, kesinlik(precision), F ölçütü ve ROC alanı değerleri de kriter olarak alınmaktadır. Bu kapsamda çalışmanın çeşitli basamakları göz önünde bulundurulmak suretiyle karşılaştırma işleminin uygulanılması ile değerlendirme safhasına geçilmektedir.

3.6.1. Doğru pozitif oranı (TP rate)

Doğru sınıflandırılmış pozitif örneklerin sayısını vermekte olup Denklem 3.21.'de formülasyonu gösterilmektedir.

$$\text{TP Oranı} = \frac{TP}{TP+FN} \quad (3.21)$$

Bahsedilen bu oran hasta olan kişilere doğru bir şekilde hastasın diyebilme şeklinin matematiksel olarak ifade edilmesidir.

3.6.2. Yanlış pozitif oranı (FP rate)

Yanlış sınıflandırılmış pozitif örnek sayısını vermekte olup Denklem 3.22.'de formülasyonu gösterilmektedir.

$$\text{FP Oranı} = \frac{FP}{FP+TN} \quad (3.22)$$

Bahsedilen bu oran hasta olmayan hastalara yanlış bir şekilde hastasın diyebilme şeklinin matematiksel olarak ifade edilmesidir.

3.6.3. Doğru negatif oranı (TN rate)

Doğru sınıflandırılmış negatif örnek sayısını vermekte olup Denklem 3.23.'te formülasyonu gösterilmektedir.

$$\text{TN Oranı} = \frac{TN}{TN+FN} \quad (3.23)$$

Bahsedilen bu oran hasta olmayan hastalara doğru bir şekilde hasta değilsin diyebilme şeklinin matematiksel olarak ifade edilmesidir.

3.6.4. Yanlış negatif oranı (FN rate)

Yanlış sınıflandırılmış negatif örnek sayısını vermekte olup Denklem 3.24.'te formülasyonu gösterilmektedir.

$$\text{FN Oranı} = \frac{FN}{FN+TP} \quad (3.24)$$

Bahsedilen bu oran hasta olan hastalara yanlış bir şekilde hasta değilsin diyebilme şeklinin matematiksel halidir.

3.6.6. Kesinlik (Precision)

Model seçiminde önemli bir kriter olan Kesinlik, Yanlış Pozitif durumlarının yüksek seyrettiği zamanlarda önemi artan bir parametre olup formülü Denklem 3.25.'te gösterilmiştir.

$$\text{Kesinlik} = \frac{TP}{TP+FP} \quad (3.25)$$

Bahsedilen bu oran hasta denilen kişilerin gerçekten kaçının hasta olduğu tespitinin matematiksel halidir.

3.6.7. F ölçütü (F score)

F Ölçütü, hassaslık ve kesinliğin harmonik ortalaması olarak tanımlanmaktadır ve formülü Denklem 3.26.'da gösterilmiştir.

$$\text{F Ölçütü} = 2 * \frac{\text{Kesinlik} * \text{Hassasiyet}}{\text{Kesinlik} + \text{Hassasiyet}} \quad (3.26)$$

Bahsedilen bu oran hassaslık ve kesinlik değerleri arasında ikilem olması sebebiyle tanımlanma ihtiyacı hissedilen bir oran olarak performans değerlendirme kriterleri arasında yer almaktadır.

3.6.8. Alıcı İşletim Karakteristiği eğrisi (Receiving Operating Characteristic curve)

ROC eğrisi başarı hesaplamaları için kullanılmakta olup doğru pozitif oranı ile yanlış pozitif oranı arasındaki ilişkiden doğmaktadır. ROC eğrisinin formülü Denklem 3.27.'de gösterilmektedir.

$$\text{ROC Eğrisi} = \frac{TP}{FP} \quad (3.27)$$

Bahsedilen bu oran grafik üzerinde çizildiği taktirde altında kalan alanın artışı ile başarı oranındaki artışın doğru orantılı olduğu yorumu yapılabilmektedir.

3.6.9. Başarı yüzdesi

Model üzerinde yapılan çalışmalarda modelin doğru tahmin ettiği durumlar yeterli bir ölçüt olarak değerlendirilmezler. Bu bakımdan tüm tahminleri başarı yüzdesi kapsamında ele almak bu ölçütün değerlendirilme sırasında daha anlamlı hale gelmesini sağlamaktadır. Ek olarak başarı yüzdesi karışıklık matrisi ile daha net bir şekilde açıklanabilmektedir.

3.6.9.1. Karışıklık matrisi (Confusion matrix)

Sınıflandırıcı algoritmamızın farklı sınıflardaki veri setlerini ne kadar iyi tanıdığını analiz ederek gerçek durum ile tahmin edilen durum arasındaki bağı anlamak üzerine kurulan bu matrisin genel çerçevesi Şekil 3.19.'da gösterilmiştir [20].

		Gerçek Durumlar	
		Pozitif Durumlar(A)	Negatif Durumlar(B)
Tahmini Durumlar	Pozitif Durumlar	TP 1. Kutu	FP 2. Kutu
	Negatif Durumlar	FN 3. Kutu	TN 4. Kutu

Şekil 3.19. Karışıklık matrisinin genel çerçevesi

Yukarıda 2x2'lik olan matrisin boyutu, hedef sayısına göre istenilen nxn'lik bir boyuta dönüşüm gösterebilmektedir. Bu karışıklık matrisinde yer alan TP değeri gerçek durumda pozitif olan bir durumun tahmininin pozitif olduğunu, FP değeri gerçek durumda negatif olan bir durumun tahmininin pozitif olduğunu, FN değeri gerçek durumda pozitif olan bir durumun tahmininin negatif olduğunu, TN değeri ise gerçek durumda negatif olan bir durumun tahmininin negatif olduğunu ifade etmektedir.

Çalışmamızda ifade edilen model çeşitli algoritmalara tabi tutulduktan sonra elde edilen başarı yüzdesi değeri Şekil 3.19.'da gösterilen confusion matrix üzerinden hesaplanmaktadır. Şöyle ki; 1. kutu A olması gerekirken A olanları, 2. kutu A olması gerekirken B olanları, 3. kutu B olması gerekirken A olanları, 4. kutu B olması gerekirken B olanları şeklinde ifade edilir ve başarı yüzdesi değeri, 1. Ve 4. kutudaki değerlerin çarpımından, 2. ve 3. kutudaki değerlerin çarpımının çıkarılması ile elde edilmektedir.

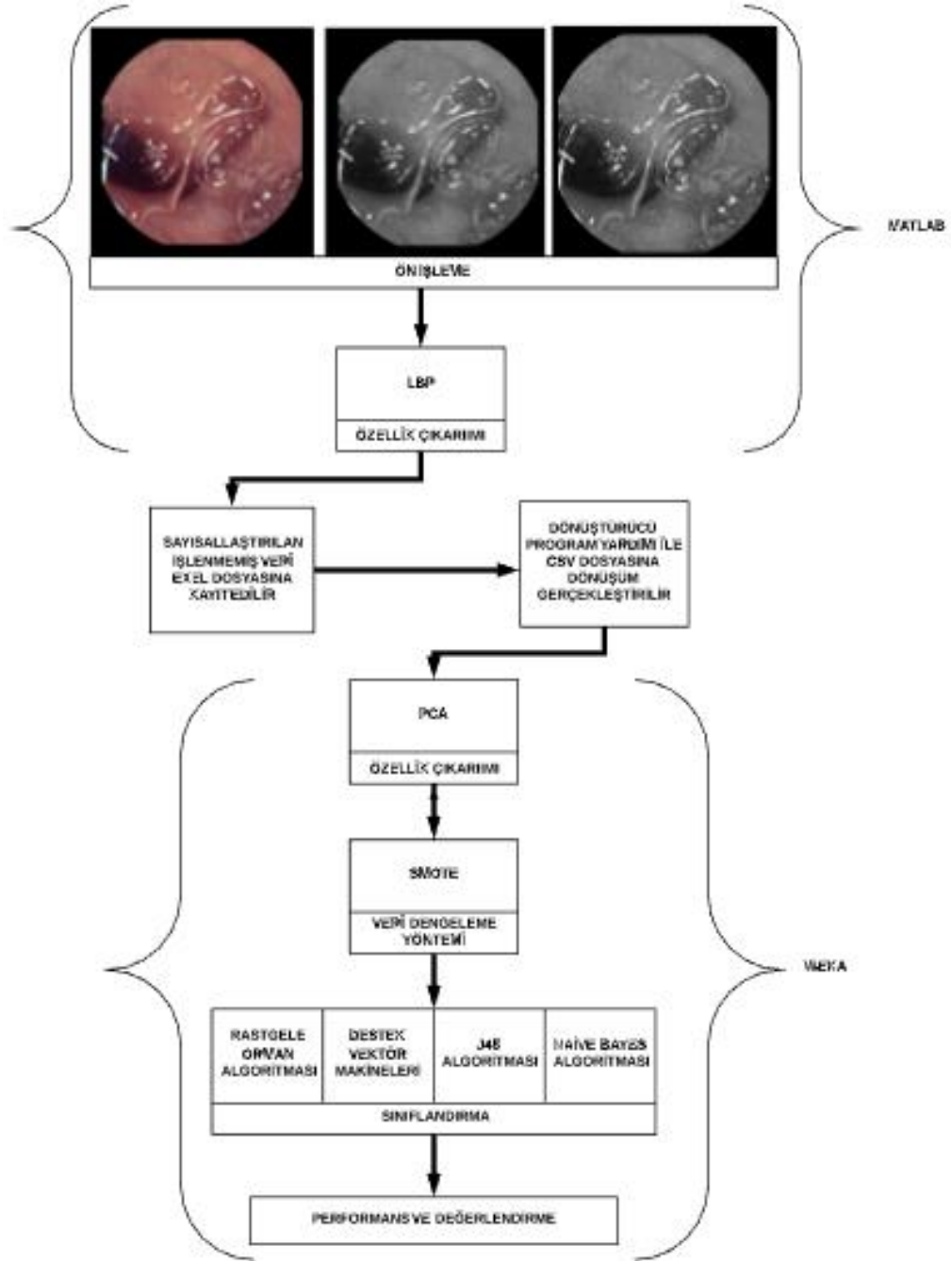
BÖLÜM 4. ARAŞTIRMA BULGULARI

4.1. Deneysel Çalışma

Şekil 4.1.'de gösterilen mevcut sistemin modellenmesinde ilk aşama olarak ince bağırsak lokasyonları üzerinden alınan polipli ve polipsiz görüntüler, toplanan görüntü verilerinde iyileştirmelerin yapılması anlamına gelen görüntü ön işleme sürecine tabi tutulmaktadır. Bu süreçte 3 boyutlu bir yapıya sahip olan orijinal görüntü, grayscale filtresi ile 2 boyutlu görüntü formatına çevrilmekte olup polipleri algılama hususunda orijinal görüntüye kıyasla bir başarı sağlamaktadır. Akabinde elde edilmiş olan 2 boyutlu görüntüler; geçişleri keskinleştirmek, bulanıklaşmış görüntülerdeki ayrıntıları yeniden ortaya çıkarmak amacı ile tıp dünyasında sıklıkla kullanılan unsharp filtresine tabi tutulmakta ve başarı oranında ciddi miktarda artış meydana getirmektedir. Matlab programı vasıtasıyla bu filtrelerin uygulandığı iyileştirilmiş görüntüler, özellik çıkarım yöntemlerinden biri olan Yerel İkili Örüntü görsel tanımlayıcısı ile sayısallaştırılarak üzerinde kullanılması gereken çeşitli tekniklerin uygulanmasını mümkün kılmıştır. Sayısallaştırılan bu görüntü verileri, matlab programından excel dosyasına kopyalanılmış ve program dönüştürücü vasıtası ile csv dosyasına dönüşümü sağlanmıştır. Bu aşamaya kadarki süreçlerde matlab programı kullanılarak yapılan çalışma, sonraki aşamalar için weka programı üzerinden devam etmiştir. Bu bağlamda csv dosyası weka programında açılarak görüntü verilerine Temel Bileşenler Analizi uygulanmıştır. Bu şekilde kesişme ve çakışmalardan arındırılmış daha küçük boyuta sahip bir veriseti elde edilerek başarı oranında artış sağlanmıştır. Ancak çalışmanın giriş kısmında bahsedildiği üzere ince bağırsaklarda görülen mevcut polip oranının azlığı, doğrudan verisetine yansıdığı için polipli ve polipsiz görüntüler üzerinde bir dengesizliği meydana getirmiştir. Bu problem ile başa çıkmak amacıyla weka programında KNN algoritması üzerinden sentetik veriler oluşturulmuştur. Böylelikle daha dengeli hale gelen veri seti

sayesinde başarı oranında yükselme görülmüştür. Son aşamada ise nihai duruma getirilen görüntü verileri, olasılık ilkelerine göre tanımlanmış matematiksel hesaplar kullanılarak sınıflandırma işlemini gerçekleştiren Naif Bayes (Naive Bayes) algoritmasına, iki sınıf arasında karar sınırı bularak sınıflandırmayı sağlayan Destek Vektör Makineleri (Support Vector Machine) algoritmasına, birden fazla karar ağacı üreterek sınıflandırma sürecini gerçekleştiren Rastgele Orman (Random Forest) algoritmasına ve ağaç yapısı şeklinde model oluşturmak suretiyle sınıflandırmayı hedefleyen Karar Ağaçları (Decision Tress) algoritmasına; Çapraz Geçerlilik yöntemi üzerinden çeşitli aşamalarda tabii tutularak performans değerlendirme ölçütleri ve başarı yüzdelerinin hesaplanması ile nihai sonuçlar üzerinden değerlendirilmiştir.

Tablo 4.1. Tablo 4.2. ve Tablo 4.3. incelendiğinde aşama aşama yapılan tüm işlemlerin başarı yüzdesine ve performans ölçütlerine pozitif bir katkısı olduğu görülmektedir.



Şekil 4.1. Mevcut sistemin modellenilmesi

Özetle; Şekil 4.1.'de modellenen sistemin ilk aşamasında, matlab programı vasıtasıyla görüntü verileri iyileştirilmiş ve yerel ikili örüntü yöntemine tabi tutulmak suretiyle görüntülerin sayısallaştırılması sağlanmıştır. Akabinde sayısallaştırılan bu veri seti; Naif Bayes (Naive Bayes), Destek Vektör Makineleri (Support Vector Machine), Rastgele Orman (Random Forest) ve Karar Ağaçları (Decision Tress-J48) algoritmaları tarafından cross validation geçerliliği üzerinden sınıflandırılarak

performans ölçütleri ve başarı yüzdesi değerleri Tablo 4.1.'de görüldüğü gibi elde edilmiştir.

Tablo 4.1.İyileştirilmiş görüntü verilerinin Yerel İkili Örüntü yöntemi ile sayısallaştırılmasının ardından sınıflandırma algoritmalarının uygulanması suretiyle elde edilen performans ölçütleri ve başarı yüzdesi değerleri

Sınıflandırıcılar	TP Oranı	FP Oranı	Hassaslık	Kesinlik	F Ölçütü	ROC Alanı	Başarı Yüzdesi
Destek V.M. Alg.	0,705	0,705	0,705	0,705	0,827	0,500	70,4698%
Rastgele Or. Alg.	0,711	0,557	0,711	0,678	0,675	0,664	71,1409%
Karar Ağacı Alg.	0,631	0,551	0,631	0,619	0,624	0,496	63,0872%
Naive Bayes Alg.	0,671	0,468	0,671	0,669	0,670	0,648	67,1141%

Tablo 4.1. incelendiğinde en başarılı algoritmanın, rastgele orman algoritması olduğu görülmektedir.

Sayısallaştırılan görüntülere temel bileşenler analizi uygulandıktan sonra meydana gelen verisetinin; sınıflandırma algoritmalarına tabi tutulmak suretiyle elde edilen nihai başarı oranı, performans ölçütleri ve başarı yüzdesi değerleri Tablo 4.2.'de verilmiştir.

Tablo 4.2.Sayısallaştırılmış verilere Temel Bileşenler Analizi yönteminin uygulanmasının ardından sınıflandırma algoritmalarına tabi tutulması suretiyle elde edilen performans ölçütleri ve başarı yüzdesi değerleri

Sınıflandırıcılar	TP Oranı	FP Oranı	Hassaslık	Kesinlik	F Ölçütü	ROC Alanı	Başarı Yüzdesi
Destek V.M. Alg.	0,738	0,598	0,738	0,748	0,671	0,570	73,8255%
Rastgele Or. Alg.	0,738	0,559	0,738	0,723	0,690	0,699	73,8255%
Karar Ağacı Alg.	0,685	0,541	0,685	0,655	0,663	0,638	68,4564%
Naive Bayes Alg.	0,691	0,565	0,691	0,654	0,659	0,632	69,1275%

Tablo 4.2. incelendiğinde en başarılı algoritmaların; aynı başarı yüzdesine sahip olan Destek Vektör Makineleri algoritması ve Rastgele Orman algoritması olduğu görülür. Tablo 4.1.'e kıyasla artması gereken hassaslık, kesinlik, F ölçütü, ROC alanı ve başarı yüzdesi değerlerinde artış görülürken, azalması gereken FP oranı değerinde ise bir azalış meydana gelmiştir.

Temel bileşenler analizi uygulandıktan sonra elde edilen veri seti, SMOTE yöntemi üzerinden azınlık veriye sahip sınıfın artan sentetik verileri ile daha dengeli bir

duruma getirilmek suretiyle tekrar sınıflandırma algoritmalarına tabi tutulmuş ve elde edilen nihai başarı oranı, performans ölçütleri ve başarı yüzdesi değerleri Tablo 4.3.'te verilmiştir.

Tablo 4.3.PCA yönteminin uygulanılmasının ardından elde edilen verilerin SMOTE yöntemiyle dengeli veri seti haline getirilmesinin ardından sınıflandırma algoritmalarının tabi tutulması suretiyle elde edilen performans ölçütleri ve başarı yüzdesi değerleri

Sınıflandırıcılar	TP Oranı	FP Oranı	Hassaslık	Kesinlik	F Ölçütü	ROC Alanı	Başarı Yüzdesi
Destek V.M. Alg.	0,798	0,210	0,798	0,798	0,797	0,794	79,7927%
Rastgele Or. Alg.	0,762	0,242	0,762	0,762	0,762	0,868	76,1658%
Karar Ağacı Alg.	0,720	0,282	0,720	0,721	0,720	0,732	72,0207%
Naive Bayes Alg.	0,689	0,310	0,689	0,692	0,690	0,749	68,9119%

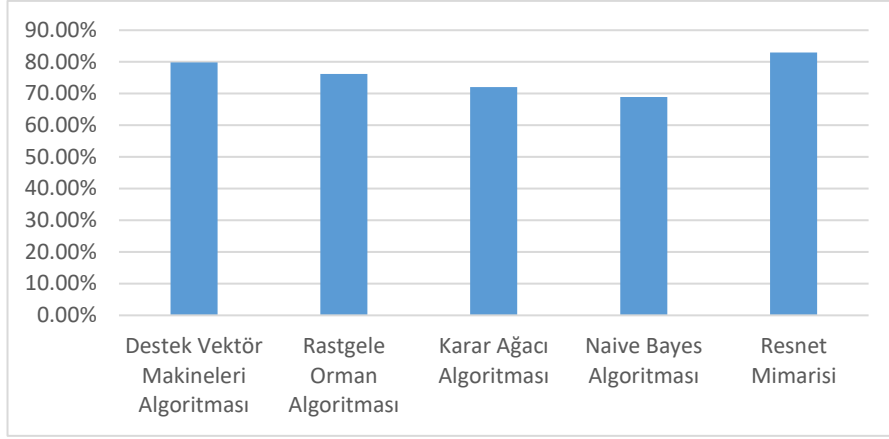
Tablo 4.3. incelendiğinde en yüksek başarı yüzdesine sahip algoritmanın %79.7927 oranı ile Destek Vektör Makineleri algoritması olduğu görülmektedir. Tablo 4.2.'ye kıyasla artması gereken hassaslık, kesinlik, F ölçütü, ROC Alanı ve başarı yüzdesi değerlerinde artış görülürken azalması gereken FP oranı değerinde ise bir azalış meydana gelmiştir. Ek olarak performans ölçütleri ile başarı değerlerinde görülen nihai oranlar, tatmin edici rakamlar olmakla birlikte tablolar arası geçişlerde kademeli bir şekilde değişen oranların miktarı, çalışmada aşama aşama kullanılan yöntemlerin ne kadar doğru olduğunu göstermektedir.

Ek olarak geçmiş yıllardan bu yana sıklıkla kullanılan geleneksel algoritmaların yanında son zamanlarda bir çok çalışmada uygulanılan derin öğrenme üzerinde temellenen ve gradyanların yok olması problemine bir yanıt niteliği taşıyan Resnet Mimarisi, ilgili görüntü verilerine uygulanılmış ve geleneksel algoritmalara kıyasla başarı oranında %83'lük bir oran ile polip teşhisine pozitif şekilde yansımıştır.

BÖLÜM 5. TARTIŞMA VE SONUÇ

Çalışma KID Atlas veriseti içerisinde bulunan polipli ve polipsiz ince bağırsak görüntülerine çeşitli filtreler uygulayarak görüntüler üzerinde bir iyileştirme sunmuştur. Ardından kullanılan özellik çıkarım yöntemlerinden ilki olan Yerel İkili Örüntü Yöntemi ile görüntülerin sayısallaştırılması sağlanmış ve diğer özellik çıkarım yöntemi olan Temel Bileşenler Analizi yöntemi ile de alakasız verilerin uzaklaştırılması sağlanarak daha etkin bir çalışma görülmüş ve performans oranında artış elde edilmiştir. Bir çok algoritmanın dengeli veri üzerinden çok daha iyi performans gösterdiği bilinmesi sebebi ile veri dengeleme yöntemi olan SMOTE'nin uygulanması sağlanarak veriler üzerindeki başarı oranı arttırılmıştır. Ek olarak son zamanlarda sıklıkla kullanılan Resnet Mimarisi de çalışmaya dahil edilerek geleneksel algoritmalara kıyasla polip teşhisi hususunda daha yüksek başarı oranı elde edilmiştir.

Yerel İkili Örüntü Yöntemi, Temel Bileşenler Analizi Yöntemi ve SMOTE Yönteminin uygulanmasından sonra elde edilen çıktılar, seçilen 4 algoritma tarafından sırasıyla sınıflandırılmış ve nihai başarı oranları elde edilmiştir. Aynı zamanda kullanılan yöntemlerden farklı olarak ilgili görüntülere uygulanan Resnet Mimarisi ile de yapılan çalışma üzerinden bu kapsam dahilinde bir değerlendirme sunulmuştur. Şekil 4.2.'de verilen grafik tüm süreçler sonucunda uygulanan mimari ve algoritmalar için başarı kıyaslamaları açısından bir karşılaştırma sunmaktadır.



Şekil 4.2. Yöntemler bazında sınıflandırma yüzdeleri

Sonuç olarak sunulan çalışmada polipli ve polipsiz ince bağırsak lokasyonları üzerinden alınan görüntüler için yapılan iyileştirmelerin ardından uygulanan farklı yöntemler ile elde edilen performanslarda pozitif yönde bir artış sağlandığı görülmüştür. Yöntemler uygulandıkça; hasta olanlara hangi olasılıkla doğru bir şekilde hasta olduğunu söyleyebilme olasılığının formüle edilmiş hali olan TP oranında, hasta olanları tespit etme hususunda yardımcı olan hassaslık oranında, hasta dediğimiz hastaların gerçekten kaçının hasta olduğunu gösteren kesinlik oranında, hassaslık ile kesinliğin harmonik ortalaması olan F ölçütü oranında, başarı ile doğru orantılı olan ROC eğrisi altında kalan alanın oranında bir artış yaşanırken hasta olmayanlara hangi olasılıkla yanlış bir şekilde hasta olduğunu söyleyebilme olasılığının formüle edilmiş hali olan FP oranında bir azalış olması çalışmanın başarılı bir çalışma olduğunun göstergesidir. Ek olarak ilgili görüntüler, son zamanlarda sıklıkla kullanılan Resnet Mimari açısından da değerlendirilmiş ve geleneksel algoritmalara kıyasla daha yüksek bir başarı oranı elde edilmiştir.

Resnet Mimarisinin, kullanılan diğer yöntemlere kıyasla daha yüksek başarı elde etmesinde sinir ağının derinlere inerek daha iyi bir eğitim süreci geçirmesi, kullanılan parametrelerin uzun denemeler sonucunda iyi bir şekilde tune edilmiş olması ve resnet mimarisinin derin öğrenme kapsamında kullanılan Convolutional Neural Network Mimarisine kıyasla gradyanların yok olması problemine bir çözüm niteliği taşıması gösterilebilmektedir.

5.1. Gelecek Çalışmalar

İnce bağırsak görüntüleri üzerinde yapılan bu çalışmada görüntüler matlab programının ön işleme aşamasında, [3] çalışmasını baz almak suretiyle görüntüler üzerinde parlaklıkları yok etme sürecine tabi tutulmasının ardından hesaplanılan başarı yüzdelerinin ve performans ölçütlerinin oranlarında daha fazla bir artış yaşanılabilir.

Algoritma doğaları kapsamında değerlendirildiğinde; olasılıksal sonuç üretemeyen ancak çok boyutlu uzaylarda güzel sonuçlar veren ve aykırı değerlere karşı güçlü bir tutum sergileyen destek vektör makineleri algoritması ile olasılıksal değerler üretebilen verisetinin dengesiz olma durumundan çok etkilenmeyen ve hızlı çalışan naive bayes algoritmasının hibrit bir şekilde kullanımı ile bir algoritmanın artısı diğer algoritmanın eksisi olabilir ve daha güzel sonuçlarda elde edilebilir

Ek olarak problem uzayının genişletilmesi ile beraber morfolojik özellik çıkarımı ile ilgili algoritmalar araştırılarak bu yönde çalışmalar yapılabilir ya da probabilistik ağların kullanılması kapsamında yapılan çalışmalar incelenebilir ve tıp dünyasına faydalı olabilecek çeşitli fikirlerin ortaya çıkması mümkün olabilmektedir.

KAYNAKLAR

- [1] A. Koulaouzidis et al., “KID Project: an internet-based digital video atlas of capsule endoscopy for research purposes,” *Endosc. Int. Open*, vol. 05, no. 06, pp. E477–E483, 2017, doi: 10.1055/s-0043-105488.
- [2] D. K. Iakovidis and A. Koulaouzidis, “Automatic lesion detection in wireless capsule endoscopy - A simple solution for a complex problem,” 2014 IEEE Int. Conf. Image Process. ICIP 2014, pp. 2236–2240, 2014, doi: 10.1109/ICIP.2014.7025453.
- [3] A. Sánchez-González, B. García-Zapirain, D. Sierra-Sosa, and A. Elmaghraby, “Automatized colon polyp segmentation via contour region analysis,” *Comput. Biol. Med.*, vol. 100, no. March, pp. 152–164, 2018, doi: 10.1016/j.combiomed.2018.07.002.
- [4] R. A. de Latour, S. M. Kilaru, and S. A. Gross, “Management of small bowel polyps: A literature review,” *Best Pract. Res. Clin. Gastroenterol.*, vol. 31, no. 4, pp. 401–408, 2017, doi: 10.1016/j.bpg.2017.06.003.
- [5] D. K. Iakovidis, S. V. Georgakopoulos, M. Vasilakakis, A. Koulaouzidis, and V. P. Plagianakos, “Detecting and Locating Gastrointestinal Anomalies Using Deep Learning and Iterative Cluster Unification,” *IEEE Trans. Med. Imaging*, vol. 37, no. 10, pp. 2196–2210, 2018, doi: 10.1109/TMI.2018.2837002.
- [6] D. Qi et al., “Repair and regeneration of small intestine: A review of current engineering approaches,” *Biomaterials*, vol. 240, no. January, p. 119832, 2020, doi: 10.1016/j.biomaterials.2020.119832.
- [7] A. Karargyris and N. Bourbakis, “Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos,” *IEEE Trans. Biomed. Eng.*, vol. 58, no. 10 PART 1, pp. 2777–2786, 2011, doi: 10.1109/TBME.2011.2155064.
- [8] M. Vasilakakis, A. Koulaouzidis, D. E. Yung, J. N. Plevris, E. Toth, and D. K. Iakovidis, “Follow-up on: optimizing lesion detection in small bowel capsule endoscopy and beyond: from present problems to future solutions,” *Expert Rev. Gastroenterol. Hepatol.*, vol. 13, no. 2, pp. 129–141, 2019, doi: 10.1080/17474124.2019.1553616.

- [9] J. C. Russ, *Computer-Assisted Microscopy: The Measurement and Analysis of Images*. .
- [10] S. C. F. Lin et al., “Intensity and edge based adaptive unsharp masking filter for color image enhancement,” *Optik (Stuttg.)*, vol. 127, no. 1, pp. 407–414, 2016, doi: 10.1016/j.ijleo.2015.08.046.
- [11] J. Bernal, J. Sanchez, and F. Vilarino, “Impact of image preprocessing methods on polyp localization in colonoscopy frames,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, no. July, pp. 7350–7354, 2013, doi: 10.1109/EMBC.2013.6611256.
- [12] D. K. Iakovidis and A. Koulaouzidis, “Automatic lesion detection in capsule endoscopy based on color saliency: Closer to an essential adjunct for reviewing software,” *Gastrointest. Endosc.*, vol. 80, no. 5, pp. 877–883, 2014, doi: 10.1016/j.gie.2014.06.026.
- [13] E. Tekeli, M. Cetin, and A. Ercil, “A Local Binary Patterns and Shape Priors Based Texture Segmentation Method,” pp. 1–4, 2007, doi: 10.1109/siu.2007.4298755.
- [14] I. Steganography, “Yerel İkili Örüntü Tabanlı Veri Gizleme Algoritması : LBP- LSB Local Binary Pattern Based Data Hiding Algorithm : LBP-,” pp. 41–47, 2017.
- [15] J. Ma and Y. Yuan, “Dimension reduction of image deep feature using PCA,” *J. Vis. Commun. Image Represent.*, vol. 63, 2019, doi: 10.1016/j.jvcir.2019.102578
- [16] Ş. Yücelbaş and C. Yücelbaş, “Temel Bileşen Analizi Yöntemleri Kullanarak Parkinson Hastalığının Otomatik Teşhisi,” *Eur. J. Sci. Technol.*, no. 16, pp. 294–300, 2019, doi: 10.31590/ejosat.568544.
- [17] S. Narasimhan and S. L. Shah, “Model identification and error covariance matrix estimation from noisy data using PCA,” *IFAC Proc. Vol.*, vol. 37, no. 1, pp. 511–516, 2004, doi: 10.1016/s1474-6670(17)38783-9.
- [18] B. S. Raghuwanshi and S. Shukla, “SMOTE based class-specific extreme learning machine for imbalanced learning,” *Knowledge-Based Syst.*, vol. 187, p. 104814, 2020, doi: 10.1016/j.knosys.2019.06.022.
- [19] D. Elreedy and A. F. Atiya, “A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance,” *Inf. Sci. (Ny.)*, vol. 505, pp. 32–64, 2019, doi: 10.1016/j.ins.2019.07.070.

- [20] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014.
- [21] T. Kavzoglu and I. Colkesen, "Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi," *Harit. Derg.*, no. 144, pp. 73–82, 2010.
- [22] J. Shotton et al., "Real-Time human pose recognition in parts from single depth images," *Commun. ACM*, vol. 56, no. 1, pp. 116–124, 2013, doi: 10.1145/2398356.2398381.
- [23] Ö. Akar and O. Güngör, "Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması," *J. Geod. Geoinf.*, vol. 1, no. 2, pp. 139–146, 2012, doi: 10.9733/jgg.241212.1t.
- [24] L. H. Shehab, O. M. Fahmy, S. M. Gasser, and M. S. El-Mahallawy, "An efficient brain tumor image segmentation based on deep residual networks (ResNets)," *J. King Saud Univ. - Eng. Sci.*, no. xxxx, pp. 1–9, 2020, doi: 10.1016/j.jksues.2020.06.001.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.
- [26] D. K. Iakovidis and A. Koulaouzidis, "Software for enhanced video capsule endoscopy: Challenges for essential progress," *Nat. Rev. Gastroenterol. Hepatol.*, vol. 12, no. 3, pp. 172–186, 2015, doi: 10.1038/nrgastro.2015.13.

ÖZGEÇMİŞ

Fatma Akalın, 14.03.1996'da İstanbul'da doğdu. İlk, orta ve lise eğitimini İstanbul'da tamamladı. 2014 yılında Çamlıca Kız Anadolu Lisesi'nden mezun oldu. 2014 yılında başladığı Sakarya Üniversitesi Bilgisayar Mühendisliği Bölümü'nü 2018 yılında bitirdi. 2018 yılında Sakarya Üniversitesi Bilgisayar Mühendisliği Bölümü'nde yüksek lisans eğitimine başladı. 2020 yılında Sakarya Üniversitesi'ne Araştırma Görevlisi olarak atandı. Halen Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi, Bilişim Sistemleri Mühendisliği Bölümü'nde Araştırma Görevlisi olarak görev yapmaktadır.