

**EKG SİNYALLERİNİN KABA KÜME TEORİSİ YARDIMIYLA
SINIFLANDIRMA ANALİZİ VE YENİ BİR SINIFLANDIRMA
ALGORİTMA ÖNERİSİ (FWRSC)**

Rasım ÇEKİK

Yüksek Lisans Tezi

Bilgisayar Mühendisliği Anabilim Dalı

Mayıs, 2015

JÜRİ VE ENSTİTÜ ONAYI

Rasım ÇEKİK'in “**EKG Sinyallerinin Kaba Küme Teorisi Yardımıyla Sınıflandırma Analizi ve Yeni Bir Sınıflandırma Algoritma Önerisi (FWRSC)**” başlıklı **Bilgisayar Mühendisliği** Anabilim Dalındaki, Yüksek Lisans Tezi 27.05.2015 tarihinde, aşağıdaki jüri tarafından Anadolu Üniversitesi Lisansüstü Eğitim Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

	Adı-Soyadı	İmza
Üye (Tez Danışmanı) :	Yard. Doç. Dr. Sedat TELÇEKEN
Üye :	Yard. Doç. Dr. Muzaffer DOĞAN
Üye :	Yard. Doç. Dr. Mustafa M. ATANAK

Anadolu Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
..... tarih ve sayılı kararıyla onaylanmıştır.

Enstitü Müdürü

ÖZET

Yüksek Lisans Tezi

**EKG SİNYALLERİNİN KABA KÜME TEORİSİ YARDIMIYLA
SINIFLANDIRMA ANALİZİ VE YENİ BİR SINIFLANDIRMA ALGORİTMA
ÖNERİSİ (FWRSC)**

Rasım ÇEKİK

Anadolu Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Yard. Doç. Dr. Sedat TELÇEKEN

2015, 79 sayfa

Kalp krizi gibi kalp rahatsızlıklarının tedavisinde erken tanı ve teşhis çok önemlidir. EKG sinyalleri, kalp ile ilgili çeşitli hastalıkların ve rahatsızlıkların teşhisinde ve erken tanı konmasında kullanılan en etkili yöntemlerden biridir. Sağlık alanında eksik ve tutarsız veri kümesi sıklıkta olduğundan EKG sinyallerinin doğru yorumlanması çok önemlidir. Kaba kümeler teorisi (KKT), bulanık kümeler gibi uzman sistemler için analiz ve bilgi çıkarımında kullanılan kural tabanlı bir yöntemdir. Kaba kümeler eksik, tutarsız ve belirsiz veri kümelerini düzenleyerek değerlendirme için uygun hale getirmektedir. Bu çalışmada ilk olarak KKT yardımı ile EKG sinyalleri sınıflandırılmıştır. Burada en az parametre kullanarak, doktorların daha hızlı ve erken teşhisine yardımcı olacak bir model oluşturulmuştur. Bu model ile EKG sinyalleri % 85'e yakın bir doğruluk ile sınıflandırılmıştır. İkinci olarak KKT kullanılarak puanlama sistemi ile oluşturulan ağırlık matrisi ve benzerlik tabanlı Nitelik Ağırlıklı Kaba Küme Sınıflandırma (Feature Weighted Rough Set Classification <FWRSC>) olarak adlandırılan yeni bir sınıflandırma metodu geliştirilmiştir. Bu metod 5 farklı veri kümesi için WEKA programındaki sınıflandırma metotları ile kıyaslanmıştır. Deney sonuçlarında FWRSC'nin WEKA programındaki birçok metottan daha yüksek performans gösterdiği görülmüştür. Ayrıca FWRSC, doğruluk ölçütü bakımından 5 veri kümesi için 67,47 genel ortalama ile en yüksek performansı göstermiştir.

Anahtar Kelimeler: Kaba kümeler teorisi, EKG, Sınıflandırma, Uzman sistemler, Büyük veri analizi, Veri madenciliği.

ABSTRACT

Master of Science Thesis

CLASSIFICATION OF ECG SIGNALS ANALYSIS USING ROUGH SETS THEORY AND A NEW CLASSIFICATION ALGORITHM APPROACH (FWRSC)

Rasım ÇEKİK

Anadolu University

Graduate School of Sciences

Computer Engineering

Supervisor: Asst. Prof. Dr. Sedat TELÇEKEN

2015, 79 pages

Early diagnosis and treatment of heart diseases such as heart attack is very important. ECG signals, for the diagnosis of various diseases and disorders of the heart is the most effective method used for early diagnosis. Because the incidence of incomplete and inconsistent data sets in the health field, the proper interpretation of the ECG signal is very important. Rough set theory (RST) is a rule-based method used in extracting information and analysis for expert analysis such as fuzzy sets. Rough sets make it suitable by organizing incomplete, inconsistent and uncertain data sets for evaluation. In this study, ECG signals classified with the RST firstly. At this point, using the minimum parameters, which would help doctors for more rapid and early detection of a pattern has been established. This model has been classified the ECG signals with an accuracy close to 85%. Later, a new classification that the weight matrix created scoring system by using RST and similarity-based method has been developed and it's been named as Feature Weighted Rough Set Classification (FWRSC). This method is compared with the classification method in WEKA for 5 difference data sets. The experimental results showed that FWRSC showed higher performance than many methods in WEKA. In addition, in terms of accuracy, it showed the highest performance with the overall average of 67.47 for 5 difference data sets.

Keywords: Rough sets theory, ECG, Classification, Expert systems, Big data analysis, Data mining.

TEŞEKKÜR

Yüksek lisans tezim süresince benden yardımlarını esirgemeyen, karşılaştığım sorunlarda bilgi ve deneyimlerini benimle paylaşan değerli hocam ve tez danışmanım Yard. Doç. Dr. Sedat TELÇEKEN'e teşekkürü bir borç bilirim.

Yüksek lisans tezimde bana büyük yardımları ve desteği olan Eskişehir Osman Gazi Üniversitesi Kardiyoloji Ana Bilim Dalı doktorlarından Sayın Doç. Dr. Hüseyin Uğur YAZICI hocama teşekkürü bir borç bilirim.

Yüksek lisans öğrenimim boyunca beni destekleyen Araş. Gör. M. Yasin PAK başta olmak üzere Bilgisayar Mühendisliğinde görev yapmakta olan tüm hocalarıma ve çalışma arkadaşlarıma teşekkürlerimi sunarım.

Yüksek lisansımı tamamlamamda bana büyük destek veren sevgili ağabeyim Cemalettin ÇEKİK'e ve aileme sonsuz teşekkürler.

Rasım ÇEKİK

Mayıs, 2015

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	ii
ABSTRACT	iii
TEŞEKKÜR	iv
ŞEKİLLER DİZİNİ	viii
ÇİZELGELER DİZİNİ	ix
SİMGELER ve KISALTMALAR DİZİNİ	x
1. GİRİŞ	1
2. ELEKTROKARDİYOĞRAFI	5
2.1. EKG Ne İçin Kullanılır?.....	6
2.2. EKG Üzerindeki Dalga, Segment ve Aralıklar	8
2.2.1.Dalgalar.....	8
2.2.2.Segmentler	9
2.2.2.1 PR segment	9
2.2.2.2 ST segment.....	10
2.2.3.Aralıklar	11
2.2.3.1 QT aralık	11
2.2.3.2 PR aralık.....	11
2.2.3.3 RR aralık	11
2.3. EKG Derivasyonları	12
2.3.1.Etraf derivasyonları.....	12
2.3.1.1 Bipolar derivasyonlar	12
2.3.1.2 Unipolar derivasyonlar	13
2.3.2.Göğüs derivasyonları	14
3. HASTALIKLAR	16
3.1. Miyokardiyal Enfarktüs.....	16
3.2. Miyokardiyal İskemi	17
3.3. Sol Ventriküler Büyümesi	18
3.4. Atrial Büyümesi	18

4. SINIF, NİTELİK VE DEĞER KÜMESİNİN BELİRLENMESİ	20
4.1. Sınıf Belirleme	20
4.2. Sınıfların EKG Belirtileri	20
4.3. Nitelik Belirleme	22
4.4. Niteliklerin Değer Kümesinin Belirlenmesi	24
4.4.1. Niteliklerin değer kümesinin sayısallaştırılması	27
5. EKG SİNYALLERİNİN SINIFLANDIRMASI	28
5.1. Kaba Kümeler Teorisi	28
5.1.1. Bilgi sistemi	28
5.1.2. Ayırt edilemezlik	28
5.1.2.1 Ayırt edilebilirlik matrisi	29
5.1.2.2 Ayırt edilebilirlik fonksiyonları	29
5.1.3. Karar sınıfları	30
5.1.4. Yaklaşım kümeleri	30
5.1.4.1 Alt yaklaşım kümesi	30
5.1.4.2 Üst yaklaşım kümesi	31
5.2. Önerilen Model	32
5.3. Uygulama	37
5.3.1. Veri kümesi hakkında	37
5.3.2. ROSE yardımı ile veri kümesinde gereksiz nitelik tespiti	39
5.3.3. WEKA’da verilerin sınıflandırılması	41
5.3.4. Uygulama sonuçları	43
6. YENİ SINIFLANDIRMA YÖNTEMİ	49
6.1. Nitelik Ağırlıklı Kaba Küme Sınıflandırma	49
6.2. Ağırlık Matrisi Hesaplanması	51
6.3. Sınıf Belirleme	52
6.4. FWRSC Algoritması	58
6.5. Uygulama	58
6.5.1. Veri kümeleri	58
6.5.2. Değerlendirme ölçütleri	58
6.5.3. Uygulama sonuçları	60

7. TARTIŞMA, SONUÇ VE ÖNERİLER	65
KAYNAKLAR	68
EK-1 FWRSC Algoritması.....	73
Ek-2 ROSE programında üretilen tüm kurallar.....	76

ŞEKİLLER DİZİNİ

	Sayfa No
2. 1. Bir EKG sinyali ve üzerindeki dalga, aralık ve segmentler	5
2. 2. Bir EKG sinyali üzerindeki dalgalar	9
2. 3. ST segmentinde çökme ve yükselme	10
2. 4. Einthoven üçgeni ve standart ekstremite derivasyonları.....	13
2. 5. Cabrera dairesi üzerinde etraf derivasyonlarının yeri	14
2. 6. Göğüs derivasyonları görünümü	14
2. 7. Etraf ve göğüs derivasyonlarının görünümü	15
3. 1. Miyokardiyal Enfarktüs	16
3. 2. Miyokardiyal İskemi	18
3. 3. Atrial Büyümesi	19
5. 1. Düşük ve Yüksek küme yaklaşımları.....	32
5. 2. Önerilen model.....	33
5. 3. ROSE programında yeni bir proje açmak	39
5. 4. ROSE programında dosya eklemek	39
5. 5. ROSE programında nitelik indirgeme.....	40
5. 6. ROSE programında kural çıkarımı	40
5. 7. WEKA programında proje açmak.....	41
5. 8. WEKA programında sınıflandırma işlemi	42
5. 9. WEKA'da (a) 15 nitelik (b) 12 nitelik için sonuçları göstermektedir	42
5. 10. Zaman bazında iyileşme sonuçları	46
6. 1. FWRSC modeli	50
6. 2. Sınıflara göre veri kümesinin bölünmesi	52
6. 3. Sınıf hesaplamaları için bir gösterim	53

ÇİZELGELER DİZİNİ

	Sayfa No
4. 1. Sınıflar için EKG belirtileri.....	21
4. 2. Seçilen nitelikler.....	23
4. 3. Her nitelik için normal değer veya değer aralığı.....	25
4. 4. Niteliklerin değer kümeleri	26
4. 5. Değer kümesinin sayısal karşılıkları	27
5. 1. Bilgi sistemi	34
5. 2. Ayırt edilemezlik matrisi.....	35
5. 3. D-çekirdek nitelikler için bilgi sistemi.....	36
5. 4. Ayırt edilemezlik matrisi.....	36
5. 5. D-çekirdek nitelik değer kümesi indirgenmiş bilgi sistemi	37
5. 6. Veri kümesinden bir kesit	38
5. 7. 15 nitelik için sınıflandırma sonuçları.....	43
5. 8. 12 nitelik için sınıflandırma sonuçları.....	44
5. 9. Nitelik sayısına bağlı değişmeler	45
6. 1. Bir nesne için farklı sınıftaki nesnelere ile ayırt edilebilirliği	50
6. 2. Bilgi sistemi	54
6. 3. Ayırt edilmezlik matrisi hesaplanması.....	54
6. 3. (Devam) Ayırt edilmezlik matrisi hesaplanması.....	55
6. 4. Veri kümeleri	58
6. 5. Karışıklık Matrisi	59
6. 6. Sınıflandırma metotları için doğruluk değerleri.....	61
6. 7. Sınıflandırma metotları için MAE değerleri	62
6. 8. Sınıflandırma metotları için f-ölçümü değerleri	63
6. 9. Ortalama doğruluk yüzdesi % 60 üzerinde olan sınıflandırma metotları	64

SİMGELER ve KISALTMALAR DİZİNİ

EKG	: Elektrokardiyografi
ME	: Miyokardiyal enfarktüsü
MİS	: Miyokardiyal iskemi
KKT	: Kaba kümeler teorisi
SVB	: Sol ventriküler büyümesi
AB	: Atrial büyümesi
ESOGÜ	: Eskişehir Osmangazi Üniversitesi
FWRSC	: Feature Weighted Rough Set Classification
$M_{i,j}$: Ayırt edilemezlik matrisi
N	: Bir ayırt edilebilirlik fonksiyonundaki 'V' işlemi frekansı
$d_{i,j}$: Ağırlık matrisindeki nitelik ağırlık değeri
$W_{i,j}$: Ağırlık matrisi
f_i	: i. nesne için ayırt edilemezlik fonksiyonu
a_j	: j. nitelik
C_x	: x. Sınıf
U	: Sonlu nesne kümesi
A	: Sonlu nitelik kümesi
\hat{R}	: Nesne kümesi eleman sayısı
\hat{r}	: Nitelik kümesi eleman sayısı
β	: Test verisi kümesi
γ	: Eğitim verisi kümesi

Alt ve üst indisler

i	: Ağırlık matrisi için nesne indeksi
j	: Ağırlık matrisi için nitelik indeksi
x	: Sınıf indeksi
y	: Sınıf indeksi

1. GİRİŞ

Kalp sağlığı, sağlık alanının en önemli konularından birisidir. Kalbin fonksiyonel görevinin öneminden dolayı kalp sağlığı ile ilgili gelişmeler önem arz etmektedir. Sağlık alanında erken teşhis birçok hastalığın tedavisinin yapılması için son derece önemlidir. Kalp hastalıklarının erken teşhisi için birçok yöntem geliştirilmiştir. Bu yöntemlerin en yaygın ve en etkili olanı Elektrokardiyografi (EKG)'dir.

Kalbin hem kas hem de sinirsel iletim sisteminin çalışmasını incelemek üzere tasarlanan bu yöntem, kalp genişlemesi, kalp büyümesi, kalp ritim bozuklukları gibi kalp ile ilgili hastalıklar hakkında önemli bilgiler verebilmektedir. Bir EKG sinyali; dalga, aralık ve segmentlerden oluşur. Dalgalar; *P*, *Q*, *R*, *S* ve *T*, segmentler; *ST* ve *PR* ve aralıklar; *RR*, *QT* ve *PR* olarak bilinir. EKG sinyali bir sinüs sinyalıdır. Normal bir EKG sinyalinde dalgalar belirli bir sıra ve aralıklarla oluşurlar. Dalgaların bu sıra ve aralıklarla oluşumu dalganın şekli, süresi, *ST segmenti*, *RR aralık* gibi parametrelerini içermektedir. Bu parametrelerdeki anormallikler, tek başına ya da başka tahlillerle birlikte kalp sorunları hakkında bilgi vermektedir. Örneğin, *RR aralığının* anormal olması kalpte ritim bozukluğu olduğunu gösterirken *ST segmentinin* çökmesi veya yükselmesi Miyokardiyal Enfarktüsü ya da Miyokardiyal İskemi hastalıklarına işaretir [1].

Eksik, belirsiz ve tutarsız veriler gerçek uygulamalarda sık rastlanan bir durumdur. Sağlık alanı da eksik, belirsizlik tutarsız verilerin çok olduğu bir alandır. Bu alanında sezgisel teşhislerin olması belirsizliği, tutarsızlığı artırdığında, verilerin rasyonel olarak tanımlanabilmesi sıklıkla güç olmaktadır. Kalbin çalışması, fonksiyonel görevi ve özel yapısı; yukarıda ifade edilen parametreler arasındaki mantıksal bağlantıyı karmaşıklığa sürüklemektedir [2]. Eğer bir veri kümesinde bazı nitelik değerleri bilinmiyorsa, bu veri kümesine eksik veri kümesi denir [3]. Veri kümelerinden çıkarım yapan birçok metot bulunmaktadır. Örneğin, eksik nitelik değerleri içeren nesnelerin doğrudan silinmesi, eksik nitelik değerini istatistiksel analiz sonucu elde edilen bir değer ile veya aynı giriş değerlerine sahip başka bir nesnenin değerleri ile bu nesnenin değerlerini değiştirmek gibi metotlar mevcuttur [4, 5]. Bu metotlar eksik veri kümesini tam veri kümesine dönüştürürler.

Ancak bu metotlardan elde edilen çıkarımlar “Kesin kurallar içermektedir” denilemez. Bu yöntemlerin yanında yüksek başarı gösteren bulanık kümeler (fuzzy sets) ve kaba kümeler (rough sets) gibi yöntemler; özellikle eksik, yetersiz ve belirsiz verileri düzenleyerek, veri analizi için yeterli hale getirmektedir [7]. Pawlak [8] tarafından ortaya atılan Kaba Kümeler Teorisi (KKT) günümüzde çoğu bilim dalında uygulama olarak kendisini kabul ettirmiştir.

Yapılan arařtırmalarda ve literatür çalıřmasında, EKG'nin tek başına bir kalp hastalıęının kesin teřhisi için yeterli olmadığı saptanmıřtır. EKG ile birlikte transözofageal ekokardiyografi, B-M mode doppler ve renkli doppler ekokardiyografi, 24 saat kan basıncı izlemi ambulatuar kan basıncı monitörizasyonu (tansiyon holteri) gibi başka tetkiklerin de bulunması gereklidir. Bütün bunlara raęmen EKG sinyalleri, teřhis için çok deęerli bilgiler tařımaktadır. Bu nedenle EKG'nin doęru yorumlanması çok önemlidir. EKG ile tanı konulabilecek en genel durumlar; Ritimler, Aktivasyon serisindeki bozukluklar, Atria ve Ventrices duvarında kalınlařma veya boyutunda büyüme, Miyokardiyal enfarktüs ve iskemi vb. gibi ana başlıklar altında toplanabilir. Uzmanlar tarafından bu tanılar, EKG kâğıdındaki bilgiler ışığında ön teřhisi konulur. Ancak bir tanıyı koymak için birden çok EKG belirtisini deęerlendirmek gerekebilir. Bu işlemler zaman ve dikkat isteyen işlemlerdir. İnsanoęlunun hata yapmaya müsait yapısı düşünöldüğünde, doktorların dikkatsizliğinden kaynaklanan hatalar yanlış tanı konulmasına ve önemli detayların gözden kaçmasına neden olabilir. Bu çalıřma ile bu problemlere çözümler bulma ve doktorların zamandan kazanım saęlaması amaçlanmıřtır.

Tezde Sol ventriküler büyümesi, Atrial büyümesi, Miyokardiyal iskemi ve Miyokardiyal enfarktüs hastalıkları üzerinde durulmuřtur. Bu hastalıkların kalp krizin oluşmasında en önemli hastalıklar olduęu arařtırmalar sonucunda görölmüřtür. Yürütölen çalıřma ile en az parametre kullanılarak, doktorların daha hızlı ve erken teřhisine yardımcı olacak bir model oluşturulmuřtur. KKT ve sınıflandırma yöntemlerinden oluşun model, gereksiz, tekrarlı ve eksik verinin sık olduęu EKG alanındaki verileri tutarlı hale getirmektedir. Bu tutarlı veri kümesi en etkili sınıflandırma yöntemi ile sınıflandırmaktadır. Bu sınıflandırma doktorlara daha hızlı ve daha az hata ile tanı koyma imkânı sunar [2].

Bu çalışmada Eskişehir Osmangazi Üniversitesi (ESOGÜ) Kardiyoloji anabilim dalı doktorlarının katkılarıyla elde edilen hastaların verileri kullanılmıştır. Toplamda 689 hastanın yorumlanmış gerçek EKG sinyalleri üzerinde çalışılmıştır. Yapılan deneylerde, önerilen model ile sınıflandırma yöntemlerinin yürütme zamanlarında (Running Time) iyileşmeler sağlanmıştır. Ayrıca kural çıkarmaya dayalı çalışan JRip algoritmasının aynı zamanda doğruluğunda da iyileşmeler olduğu görülmüştür. Önerilen model ile % 85'e yakın bir sınıflandırma başarısı elde edilmiştir. Uzman hekimlerle yapılan görüşmelerde elde edilen sonuçların yüksek oranda doğru sınıflandırdığı bilgisine varılmıştır.

Bu çalışmada, üzerinde durulmuş diğer bir konu ise sınıflandırma konusudur. Günümüzde gelişen bilgisayar teknolojileri ile birlikte kullanılmakta olan veri miktarı hızla büyümekte; büyüyen veriyi de analiz etmek zorlaşmaktadır. Bir çalışmaya göre her 20 ayda bir dünyadaki veri miktarının ikiye katlandığı ifade edilmektedir [8]. Bu kadar hızla büyüyen veri miktarında gereksiz, eksik, tutarsız, tekrarlı verinin olması kaçınılmazdır. İstatistik, yapay zeka ve makine öğrenme temeline dayalı veri madenciliği (data mining), verileri analiz ederek yararlı ve anlamlı bilgilere ulaşmaktır. Veri madenciliğinin bu işlevi; onu endüstri, ekonomi, ve iş çevrelerinden gelen yoğun ilgi yanında akademik alanda da ilgi odağı olmasına neden olmuştur. Sağlık, temel bilimler, bankacılık, finans, pazar araştırmaları gibi pek çok alanda kullanılan veri madenciliği yöntemleri bünyesinde çok sayıda algoritma barındırmaktadır [45].

Veri madenciliğinde kullanılan modeller, tahmin edici (predictive) ve tanımlayıcı (descriptive) olmak üzere iki ana başlık altında incelenmektedir [46]. Tahmin edici modellerde, sonuçları bilinen verilerden bir model geliştirilir ve bu model kullanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerleri tahmin edilmesi amaçlanmaktadır.

Sınıflandırma işlemi veri madenciliğinde tahmin edici modellerden bir tanesidir. Bu çalışmada sınıflandırma üzerinde durulmuş ve yeni bir sınıflandırma yöntemi geliştirilmiştir. Bir sınıflandırma, en basit tanımıyla bir nesnenin hangi sınıfa ait olduğunu bulma işlemidir. Sınıflandırma işleminde sınıf sayısı ve sınıfın etiketi bellidir. Ancak Sınıflandırma yapmada bazı problemler söz konusudur. Örneğin veri kümesinin eksik, tutarsız ve tekrarlı veri içermesi sınıflandırma

işlemini zorlaştırmaktadır. Ayrıca her nesnenin niteliklerinin olması ve bu nitelikler ile sınıf arasındaki ilişkisi sınıflandırma için önemlidir. Veri madenciliğinde bahsi geçen problemlere önışlem (pre-processing) prosedürleri olan veri temizleme (data cleaning), ilişki analizi (relevance analysis), veri dönüşümü ve indirgenmesi (data transformation and reduction) ile çözüm getirmeye çalışılmıştır [45]. Fakat yapılan literatür çalışmasında son zamanlarda KKT'nin bu noktada kullanımı daha yaygın ve etkili olduğu görülmüştür [9, 42-44]. KKT eksik ve tutarsız veri üzerinde çalışarak veri kümesini daha tutarlı hale getirmektedir. Ayrıca KKT, her nesnenin sınıfını belirlemede hangi niteliklerinin daha etkili olduğu bilgisini vermektedir. KKT'nin bu özellikleri kullanılarak daha etkili bir sınıflandırma yöntemi elde edilebileceği hem literatürde hem de yapılan deney sonuçlarında görülmüştür.

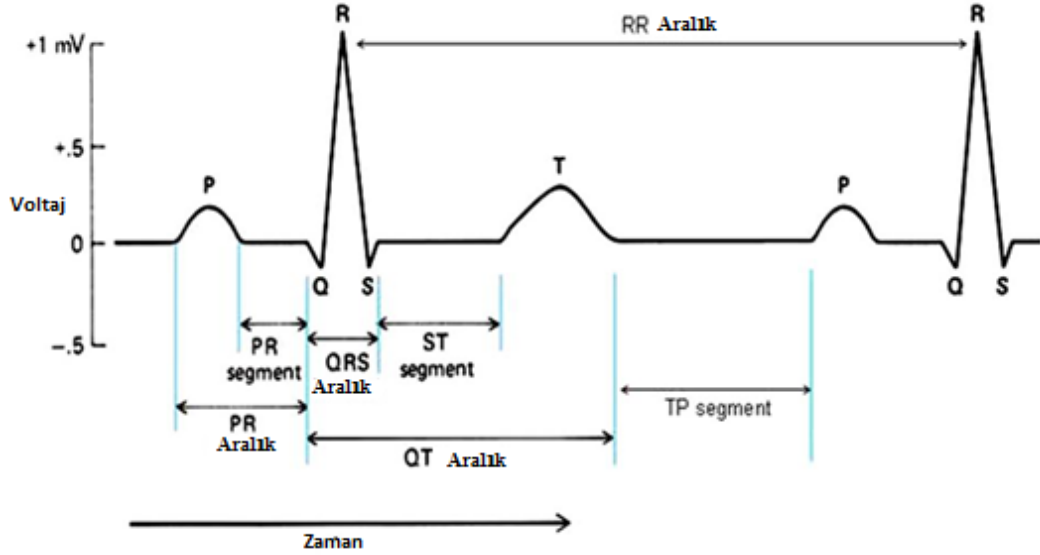
Bu tez, 7 bölümden oluşmaktadır. EKG hakkında bilgiler bölüm 2'de, hastalıklar ile ilgili bilgiler bölüm 3'te verilmiştir. Bölüm 4'te EKG ve hastalıklar ile ilgili bilgiler kullanılarak nitelik ve nitelik değerleri tespiti yapılmıştır. Bölüm 5'te önerilen model hakkında bilgi sunulmuştur. Ayrıca bu bölümde EKG sinyallerinin sınıflandırılması ile ilgili uygulama sonuçları da verilmiştir. Bölüm 6'da ise geliştirilen sınıflandırma yöntemi Nitelik Ağırlıklı Kaba Küme Sınıflandırma (Feature Weigted Rough Set Classification <FWRSC>) açıklanmış ve deney sonuçları verilmiştir.

2. ELEKTROKARDİYOĞRAFI

Kalpdeki ritimler, sinüs düğümünden çıkan elektrik uyarıları ile meydana gelir. Bir dakikada 60–100 defa tekrarlanan bu uyarılar yardımıyla önce kulakçıklar kasılır ve içindeki kan karıncıklara aktarılır, ardından karıncıklar kasılır ve kulakçıklardaki kan bu işlem ile aorta ve akciğer atardamarına pompalanır. Kalpte bu şekilde oluşan elektrik akımlarının kâğıda yazdırılma işlemine EKG denir [10].

EKG, kalbin çalışması sırasında oluşan elektriksel uyarıların elektrotlar aracılığıyla kaydedilmesi esasına dayanır. Yani bir EKG temel olarak kalpteki elektriksel işlevleri gösterdiğinde mekanik işlevler hakkında bilgi vermez. Genelde saniyede 25 mm dönme hızına ayarlanmış bir kâğıt üzerine kaydedilir. Bu kâğıdın üzerinde 1x1 mm'lik küçük ve 5x5 mm'lik büyük kareler mevcuttur. EKG sinyallerindeki dalga, aralık ve segmentlerin hesaplanmasında bu kareler dikkate alınır.

Bir EKG sinyali; dalga, segment ve aralıklardan oluşan bir sinüs sinyalidir.



Şekil 2.1. Bir EKG sinyali ve üzerindeki dalga, aralık ve segmentler

2.1. EKG Ne İin Kullanılır?

“Elektrokardiyografi kalbin iřlevinin deęerlendirilmesinde kullanılan en önemli yöntemlerden biridir. Doğrudan doğruya kalp kasının kasılma řeklini gösterir. EKG genellikle, göęüs ağrısı ile başvuran hastalarda kalp krizi tanısını koyabilmek, düzenli veya düzensiz kalp ritim bozukluklarını (aritmi) ortaya çıkarmak, efor testi sırasında kalbin elektriksel aktivitesini kaydetmek, kalbi etkileyebilecek deęişik ilaçların etkinliklerini veya yan etkilerini ve mekanik aletlerin (kalıcı kalp pili vb.) fonksiyonlarını deęerlendirmek amacıyla yapılmaktadır. EKG testi, göęüs ağrısı, çarpıntı, nefes darlığı, baş dönmesi, bayılma gibi kalp hastalığı řikâyeti olan hastaların ya da kalp hastalığı açısından risk (yüksek kan basıncı ve kolesterol, sigara içimi, diyabet, aile öyküsü veya ileri yaş) taşıyan hastaların deęerlendirilmesi amacıyla da yapılmaktadır” [10].

Kardiyoloji uzmanı Dr. Süleyman Aysel’e göre bir EKG ařağıdaki durumların tespitini yapmak için kullanılabilir:

- EKG ile kalbin ritim ve iletim bozuklukları (kalbin hızlı alışması (tařikardi) veya yavaş alışması (bradikardi) ile seyreden hastalıkları, atrial fibrilasyon, SVT) belirlenebilir.
- Kalp damar hastalığı tanısında: Kalp damarlarında daralma veya kalbin beslenmesinde problemleri görülebilir. Yeni kalp krizinin tanısında veya daha önceden kalp krizi geçirip geçirmedięinin araştırılmasında kullanılır.
- Kalp duvarlarında kalınlaşma ve kalp boşluklarında genişleme bulguları saptanabilir.
- Kardiyomiyopatiler (kalp kasının çeřitli nedenlere baęlı hastalıkları).
- Perikarditler (kalp zarının çeřitli nedenlere baęlı iltihapları).
- Miyokarditler (kalp kasının çeřitli nedenlere baęlı iltihapları).
- Bazı doğumsal kalp hastalıkları deęerlendirilebilir.
- Elektronik kalp pilinin iřlevleri deęerlendirilebilir.
- Bazı kalp ilaçlarının etkileri (digoksin, aritmi ilaçları) kontrol edilebilir.
- Vücutta sıvı ve elektrolit dengesinin bozukluklarında (özellikle serum potasyum eksikliği veya fazlalığı) yararlıdır.

Yukarıda bahsi geçen bilgiler ışığında, EKG ile tanı konulabilecek durumlar genel olarak Malmivuo ve Plonsey'ye [11] göre aşağıdaki gibi sınıflandırılabilir.

1. Kalbin elektrik eksenini
2. Kalp hızını izleme
3. Ritimler
 - a. Supraventriküler ritimler
 - b. Ventriküler ritimler
4. Aktivasyon serisindeki bozukluklar
 - a. Atriyoventriküler iletim bozuklukları (atrioventricular conduction defects)
 - b. Dal bloğu (bundle-branch block)
 - c. Wolff-parkinson-white sendromu
5. Atria ve ventricles duvarında kalınlaşma veya boyutunda büyüme
 - a. Atrial büyümesi (hypertrophy)
 - b. Ventriküler büyümesi (hypertrophy)
6. Miyokard iskemisi ve enfarktüs
 - a. İskemi
 - b. Enfarktüs
7. İlaç Etkileri
 - a. Yüksükotu
 - b. Kinidin
8. Elektrolit dengesizliği
 - a. Potasyum
 - b. Kalsiyum
9. Kardit
 - a. Kâlp zarı iltihabı
 - b. Kâlp kası iltihabı
10. Kalp pili izleme

Bu maddelerden 3., 4., 5. ve 6. maddeler kalp sağlığı ile ilgili diğer maddelerden daha önemli bilgiler taşımaktadırlar. Çünkü bu maddelerin alt başlıkları olan Miyokardiyal Enfarktüs (ME) [12 - 14], Miyokardiyal İskemi (MİS) [15, 16] ve Sol Ventriküler Büyümesi (SVB) [17] hastalıkları kalp krizinin en önemli nedenleri

olarak görülmektedir [18, 19]. Bu hastalıkların bu öneminden dolayı sınıf olarak seçilmesinde önemli bir etkisi olmuştur.

2.2. EKG Üzerindeki Dalga, Segment ve Aralıklar

Normal bir EKG'de P, Q, R, S ve T diye adlandırılan dalgalar ve bunlar arasında aralık diye tanımlanan düz çizgiler ve segmentler vardır. Bu dalga, aralık ve segmentlerdeki değişiklikler normalden sapmaları gösterir.

2.2.1.Dalgalar

EKG dalgaları P, Q, R, S, T dalgalarıdır. Bu dalgaların tanımlamalarını ilk defa 1895 yılında Willem Einthoven tarafından yapılmıştır.

“EKG tek bir hücrede değil, bir bütün hâlinde atriyal ve ventriküler hücrelerde oluşan elektrik akımını kaydeder. EKG’de iki temel olay kaydedilir. Birincisi depolarizasyon, kalp kasında uyarının yayılması; ikincisi repolarizasyon, uyarılmış kalp kasının normal dinlenme durumuna dönmesidir. Kalp kası hücrelerinde depolarizasyon ve repolarizasyon eş zamanlı gerçekleştiğinden bu elektriksel akımlar dalgalar hâlinde kaydedilir” [20].

Bir EKG sinyalinde P, Q, R, S, T dalgaları belirli bir sıra ve aralıklarla meydana gelirler. Aşağıda her dalganın anlamı verilmiştir.

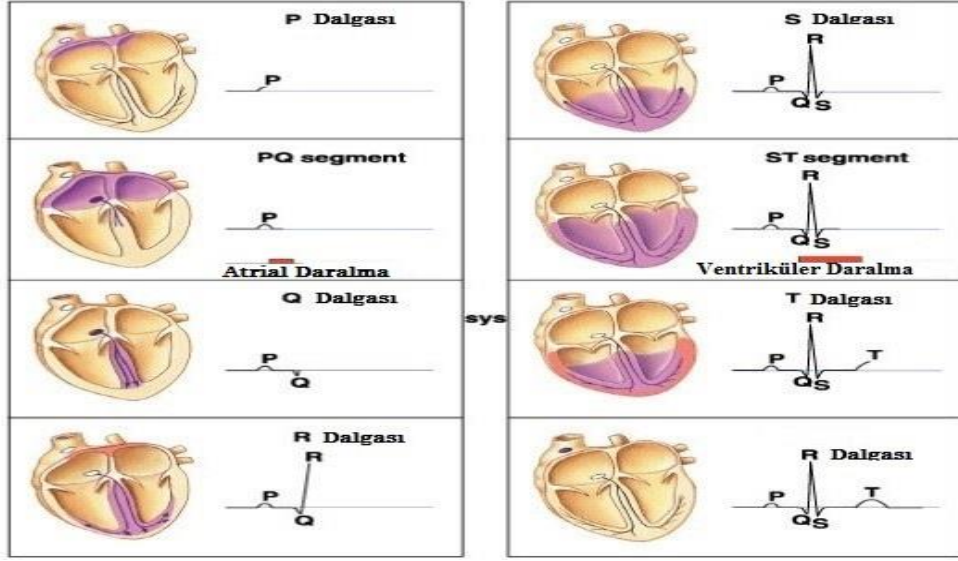
P: Atriyum depolarizasyonu.

Q: P dalgasından sonraki ilk negatif dalga.

R: P dalgasından sonraki ilk pozitif dalga.

S: P dalgasından sonraki ikinci negatif dalga.

T: Ventrikül repolarizasyon dalgası.



Şekil 2.2. Bir EKG sinyali üzerindeki dalgalar

“P-QRS-T dalga dizisi kalbin bir sinüsünü gösterir. EKG’de bir dalganın başlangıcından bir sonraki aynı dalganın başlangıcına kadar olan bölüm bir kalp sinüsüdür. İki kalp sinüsü arasında elektriksel aktivitenin olmadığı zamanı gösteren yatay düz çizgiye izoelektrik hat (bazal çizgi) denir. İzoelektrik hattın üzerine olan voltaj sapmasına, pozitif defleksiyon; izoelektrik hattın altına olan voltaj sapmasına da negatif defleksiyon denir. Defleksiyonun negatif ya da pozitif oluşunu elektrik akımının yönü belirler. Kalp kasındaki elektrik akımı, ilgili elektroda doğru yaklaşıyor ise pozitif defleksiyon, ilgili elektrottan uzaklaşıyor ise negatif defleksiyon oluşur. Kalpteki uyarı ve iletinin yönü gereği, normal bir EKG’de aVR derivasyonunda her zaman negatif defleksiyon izlenir” [20].

2.2.2.Segmentler

Bir dalganın sonundan ikinci dalganın başına kadar geçen süreye segment (bölüm, parça) denir.

2.2.2.1 PR segment

PR segmenti, P dalgası bitişi ve QRS kompleksi başlangıcı arasındaki yassı ve genellikle izoelektrik şeklindeki segmenttir [1].

PR segmenti anormallikleri iki ana durumda ortaya çıkar: Perikardit ve Atriyal iskemi.

2.2.2.2 ST segment

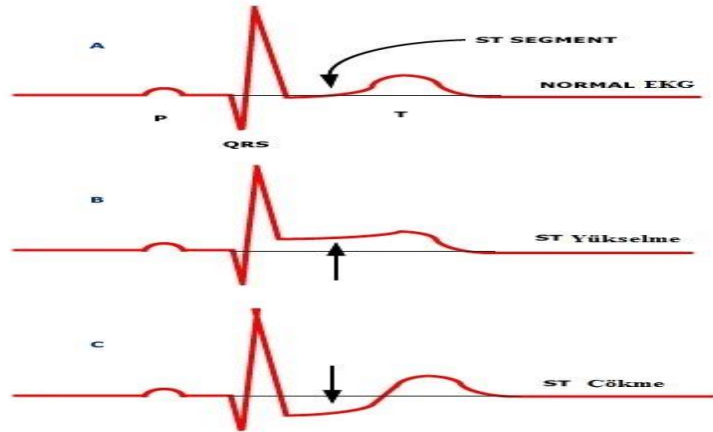
ST segmenti, S dalgası sonu (J noktası) ile T dalgası arasındaki yassı, izoelektrik EKG kısmıdır. Ventriküler depolarizasyon ve repolarizasyonu gösterir. ST segment anormalliklerinin (Yükselme/ Çökme) en yaygın sebebi miyokard iskemisi/enfarktüstür [1].

ST segment çökmesi:

- Yukarıya eğimli, aşağıya eğimli ya da yatay olabilir.
- J noktasında ≥ 2 mm ve birbirini takip eden derivasyonda ≥ 0.5 mm yatay veya aşağı eğimli olması miyokard iskemisine işaret eder.
- Yukarı eğimli olması miyokard iskemisi için nonspesifiktir.
- Resiprokal değişiklik baş aşağı dönmüş ST yükselmeye benzer morfolojidedir.
- Posterior ME'de, V1-3'te dominant R dalgaları ve dik T dalgaları ile birliktelik gösterebilir.

ST segment yükselmesi:

- Akut ST elevasyonlu miyokard enfarktüsü (STEMI) olması.
- Komşu derivasyonlarda Q dalgası ile birlikte oluşur.
- Daha çok karşı derivasyonlarda resiprokal olarak mevcuttur.



Şekil 2.3. ST segmentinde çökme ve yükselme

2.2.3.Aralıklar

Bir dalganın başından ikinci dalganın başına kadar geçen süreye, aralık (interval) denir.

2.2.3.1 QT aralık

Q dalga başlangıcından T dalga sonuna kadar geçen süreyi gösterir. Ventriküler depolarizasyon ve repolarizasyon için geçen süreyi belirtir.

QT aralığı ve kalp hızı ters orantılı bir ilişki var:

- Kalp hızlandıkça QT kısalır.
- Kalp yavaşladıkça QT uzar.

Anormal olarak uzamış bir QT aralığı, artmış ventriküler aritmi riski ile alakalıdır. Konjenital kısa QT sendromu ise, artmış paroksizmal atriyal ve ventriküler fibrilasyon ve ani kardiyak ölüm riski ile ilişkilidir [1].

2.2.3.2 PR aralık

PR aralığı, P dalgasının başlangıcından QRS kompleksinin başlangıcına kadar devam eder. Uyarının atriyumlara yayılmasını ve AV düğüm his demeti ile purkinje liflerinden geçerek ventrikülleri depolarize etmeye başlayıncaya kadar geçen süreyi gösterir [20].

PR aralığı:

- Normal aralığı 120 – 200 ms süreleri arasındadır.
- > 200 ms olması birinci derece kalp bloğu varlığı söz konusudur.
- < 120 ms olması ise pre-eksitasyonu (atriyum ve ventriküller arası aksesuar yolak varlığı) veya AV nodal (junctional) ritmi gösterir [1].

2.2.3.3 RR aralık

İki kompleks QRS arasında geçen süre olarak değerlendirilebilir. Anlık kalp ritmi iki kompleks QRS arasında geçen süre ile hesaplanabilir. Dolayısıyla kalp

ritmi RR aralık ile hesaplanır. Bu yöntemin kötü tarafı ise normal bir insanın kalp ritmi anormal çıkabilir veya tam tersi durum yaşanabilir.

2.3. EKG Derivasyonları

Kalpte ortaya çıkan elektrik değişiklikleri, EKG ile vücudun çeşitli yerlerine yerleştirilen elektrotlar vasıtasıyla güçlendirilerek kaydedilir. Vücudun çeşitli noktaları arasındaki potansiyel farklar kaydedilerek o bölgeye göre isim verilir. Her bir değişik bölge için çizdirilen elektrokardiyogram eğrisine derivasyon denmektedir. Normal EKG'de 12 ayrı derivasyonun kaydı yapılır. EKG konusunda bilinmesi gereken önemli bir nokta şu ki, 12-lead EKG yaklaşık 3 yönden (X, Y, Z) kalbin elektriksel hareketi hakkında uzaysal bilgiler vermektedir [21].

- Sağ - Sol (X) - (Right – Left (X))
- Üst - Alt (Y) - (Superior – Inferior (Y))
- Ön - Arka (Z) - (Anterior – Posterior (Z))

Yukarıdaki yönlere göre derivasyonlar sınıflandırılmıştır.

2.3.1.Etraf derivasyonları

2.3.1.1 Bipolar derivasyonlar

Bipolar (iki kutuplu) derivasyonları; DI, DII ve DIII şeklinde ifade edilir. “D” harfi derivasyon anlamı taşır.

D1: + elektrot sol kola , - elektrot ise sağ kola bağlanır, iki kol arasındaki potansiyel farkı gösterir.

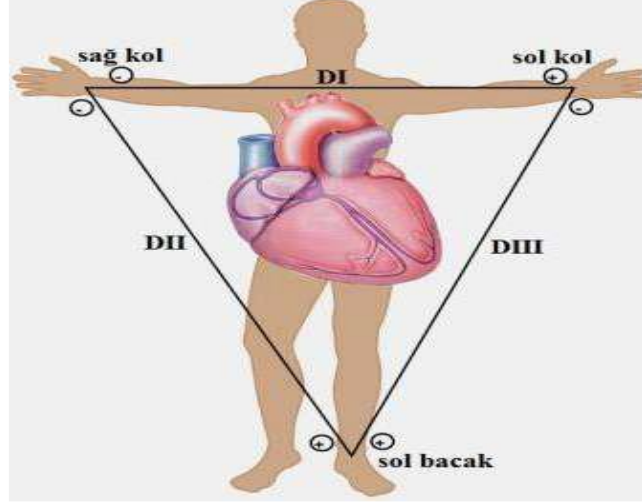
D2: + elektrot sol bacağı , - elektrot ise sağ kola bağlanır, sağ kol ile sol bacak arasındaki potansiyel farkı gösterir.

D3: + elektrot sol bacağı , - elektrot ise sol kola bağlanır, sol kol ile sol bacak arasındaki potansiyel farkı gösterir.

Milli Eğitim Bakanlığı tarafından hazırlanan “Acil Sağlık Hizmetleri EKG” adlı çalışmada bu durum şöyle açıklanmış: Sol kol elektrodu, kalpten sol kola yayılan elektriksel voltajları kaydetmekte; sağ kol elektrodu, sağ kola yayılan

voltajları kaydetmektedir. EKG cihazı içindeki mekanizma ile sağ kol voltajları, sol kol voltajlarından çıkarılmakta ve fark Derivasyon I olarak ortaya çıkmaktadır.

Kalbin sağ kol, sol kol ve sol bacadan oluşan eşkenar bir üçgenin merkezinde bulunduğu kabul edilir ki bu üçgene, einthoven üçgeni denir.



Şekil 2.4. Einthoven üçgeni ve standart ekstremité derivasyonları

Bipolar derivasyonları arasında $DI + DIII = DII$ şeklinde bağıntı eşitliği vardır. Bu bağıntı eşitliğine einthoven eşitliği denir.

2.3.1.2 Unipolar derivasyonlar

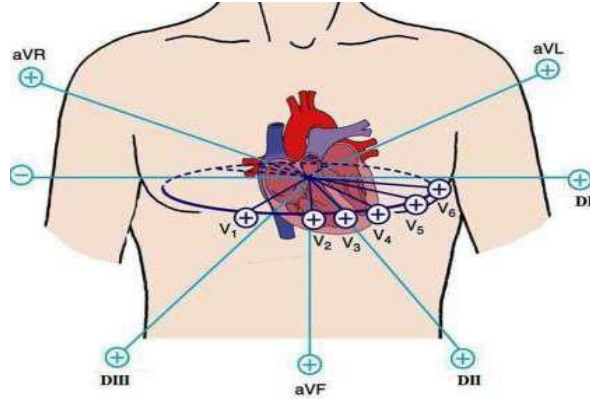
Unipolar (tek kutuplu) derivasyonları; aVR, aVL ve aVF şeklinde ifade edilir. “a” harfi güçlendirilmiş (augmented); “V” harfi voltaj anlamı taşır. R (right) sağ kolu, L (left) sol kolu, F (foot) harfi de sol bacağı göstermektedir.

Başka bir deyişle bu derivasyonlar aşağıdaki denkliği gösterir.

- sağ kol = aVR
- sol kol = aVL
- sağ bacak = aVF

“Etraf derivasyonlarının her biri, kalbi frontal düzlemde farklı açılardan gören eksenlere sahiptir. Bu eksenlerin yerleştiği frontal düzlemdeki alana Cabrera dairesi denir. Bu daire, 30 derecelik 12 eşit parçaya bölünmüştür. Üst yarım daire

Göğüs derivasyonlarından; V1 ve V2 sağ ventrikülü, V3 ve V4 ventriküler septumu, V5 ve V6 ise sol ventrikülü görür. Bu durumda V7 sol arka koltuk altı çizgisi ile 5. interkostal aralığın kesiştiği noktaya, V8 ve V9 da sırtta doğru sol 5. interkostal aralık boyunca yerleştirilir. V7 için V1, V8 için V2, V9 için de V3 elektrodu kullanılır” [20].



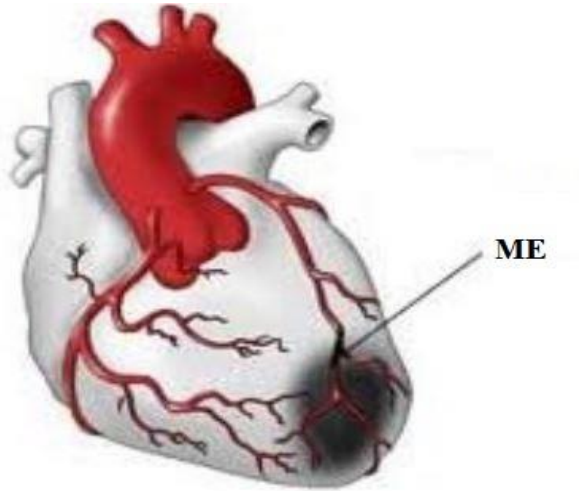
Şekil 2.7. Etraf ve göğüs derivasyonlarının görünümü

Her derivasyon, pozitif ve negatif olmak üzere iki kutba sahiptir. Bipolar derivasyonların bir ucu pozitif kutup iken diğer ucu negatif kutuptur. Unipolar derivasyonlar pozitif kutuptur; ancak bulduklarının eksenin diğer ucu negatif kutup kabul edilir.

3. HASTALIKLAR

3.1. Miyokardiyal Enfarktüs

Miyokardiyal Enfarktüs (ME), kalp kasının iskemik nekrozu olarak bilinen bir rahatsızlıktır. İskemik nekrozu, kalbin koroner kan dolaşımının belli bir bölgede yetersiz kalması sonucu, o bölgedeki kalp kası dokusunun ölmesi anlamına gelir. Sıklıkla koroner damarların atheroskleroz sonucu daralma ya da tıkanmalarına bağlı olan ME, % 80 vakada kendisini göğüs ortasındaki şiddetli, sıkıştırıcı bir ağrı ile belli eder. % 20 vakada ise ağrı ortaya çıkmaz. Yaşlı ya da şeker hastalığı olan kimselerde, ağrısız ME'ye daha sık rastlanmaktadır [22].



Şekil 3.1. Miyokardiyal Enfarktüs

EKG kaydı, akut koroner sendrom tanısı için en önemli tanı araçlarından biridir. İdeali, hasta semptomatik iken alınan kaydın, yakınması geçtikten sonra alınanlar ile karşılaştırılmasıdır. ST-segment çökmesi mortalite ve ME için ciddi bir risk göstergesidir.

Klinik bulguları akut koroner sendrom ile uyumlu olan hastalarda, ilgili iki veya daha fazla derivasyonda > 1 mm (0.1 mV) ST çökmesi ve R dalgalarının belirgin olduğu derivasyonlarda > 1 mm T dalgası negatifliği bulunması akut koroner sendrom tanısını kuvvetle destekler. Tek başına T dalgası negatifliği daha az duyarlıdır.

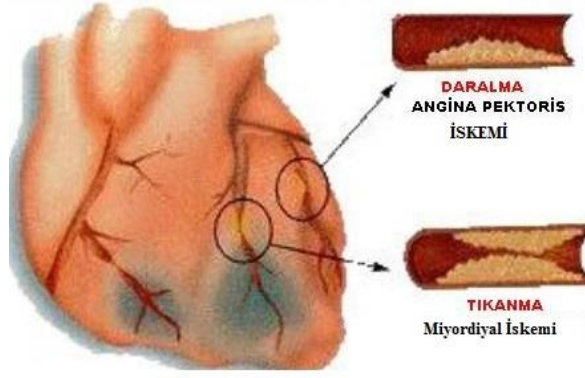
Yeni gelişen ya da yeni olduğu düşünölen dal bloęu da bir yüksek risk belirtcidir. Eęer klinik uyumlu ise yeni gelişen dal bloęu, ST yükselmeli ME bulgusu olarak kabul edilmelidir.

3.2. Miyokardiyal İskemi

Miyokardiyal İskemi (MİS), doku perfüzyonunun azalması sonucunda gelişen oksijen yoksunluğu olarak tanımlanabilir [23]. Halsizlik, terleme, kusma ve nefes alma güçlüğü, gibi belirtileri taşıyan MIS, diyabet hastaları ve yaşlılarda daha sık görölmektedir.

Subendokardiyal iskemiye baęlı ST çökmesi deęişken derivasyonlarda deęişken morfolojilerde görölebilmektedir. Genelde en çok B4-6 derivasyonlarında göze çarpar. aVR'de ST yükselmesi ile birlikte yaygın ST çökmesi sol ana koroner arter (LMCA) tıkanıklığında görölmektedir.

Hasan Akkoç "Miyokardiyal İskemi Reperfüzyon Hasarı" adlı çalışmasında şunları ifadelere ver vermiştir: "Kalpte, İskemi- reperfüzyon hasarına baęlı olarak miyokardiyal sersemleme, reperfüzyon aritmileri, miyositlerde nekroz, koroner endotelial ve mikrovasküler disfonksiyon gözlenebilir. Miyokardiyal sersemleme, iskemi reperfüzyona baęlı olarak geri dönüşsüz hasar olmamasına ve reperfüzyonun tam veya tama yakın bir şekilde sürmesine rağmen kalpte oluşan uzamış mekanik fonksiyon bozukluğu olarak tanımlanır [24]. İlk olarak 1975 yılında Heyndrickx ve ark. tarafından tanımlanmıştır [25]. Miyokardiyal sersemleme genellikle global iskemik ataklardan sonra gözlenir [24]. Fakat kısa süreli iskemiye takip eden dönemlerde bile/dahi miyokardiyal sersemleme beklenmedik derecede uzun sürebilir. Örneęin, köpek kalbinde oluşturulan 15 dakikalık iskeminin, 24 saatlik miyokardiyal sersemleme oluşturduğu gözlenmiştir" [26].



Şekil 3.2. Miyokardiyal İskemi

Miyokarddaki iskemik süreç eğer 20 dakikadan kısa sürer ve dokunun perfüzyonu bu süre içinde tekrar sağlanabilirse herhangi bir yapısal ve biyokimyasal hasar oluşmadan normal hale tam dönüş sağlanır. Eğer bu süreç uzarsa (> 45 dakika), doku kan ihtiyacının sağlanması ile beraber reperfüzyon hasarı dediğimiz olaylar zinciri tetiklenir.

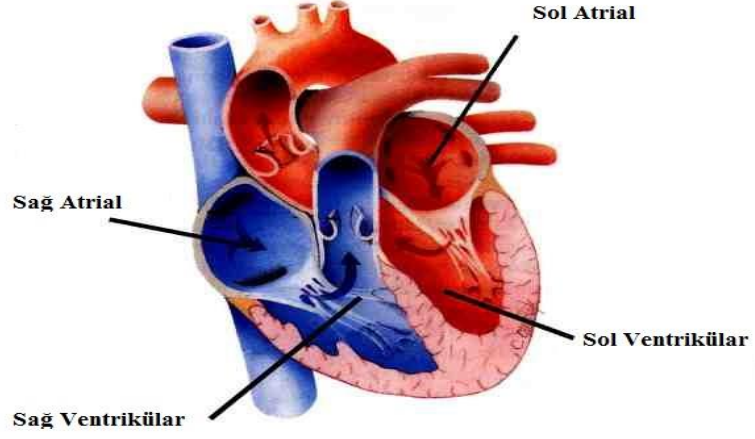
3.3. Sol Ventriküler Büyümesi

Hipertansiyonu olan hastalarda kalp sürekli yüksek basınçta kan pompalar. Bu artmış iş yükünün sonucu olarak, akciğerlerden gelen temiz (oksijenlenmiş) kanı vücuda pompalayan kalbin sol karıncığının duvarı (kas tabakası) kalınlaşır. Buna sol ventrikül büyümesi denilir. Zamanla bu karıncığın iç boşluğu genişler, fonksiyonları bozular, temiz kanı vücuda yeterince pompalayamamaya başlar. Sol kalbin pompalama fonksiyonunu yeterince gerçekleştiremediği bu durum sol ventrikül disfonksiyonu olarak adlandırılır [27].

3.4. Atrial Büyümesi

Kalbin odacıklarındaki büyüme olarak bilinir. Atriyal genişleme (büyüme) atriyumda basınç veya volüm artmasına bağlı bir problemdir. Derivasyon II'de P dalgasında çift tepe ya da çentikli P dalgası görünümü vardır. Çentiğin sebebi sol atriyum aktivasyonunun gecikmesidir. P dalgasının süresi genellikle > 120 ms'dir (normalde P süresi < 120 ms). V1'de tamamen negatif P dalgasının görülmesi de

genellikle sol atriyal anormalliđi destekler. Atriyal prematüre atım varsa, V1'de negatif komponenti olan P dalgası görülmeyebilir.



Şekil 3.3. Atriyal Büyümesi

4. SINIF, NİTELİK VE DEĞER KÜMESİNİN BELİRLENMESİ

4.1. Sınıf Belirleme

EKG ile tanı konulan hastalıklara, sağlık sorunlarına daha önceden değinilmişti. Bu tezde yürütülen araştırmalar sonucunda sadece aşağıdaki hastalıkların ele alınmasına karar verilmiştir:

- Miyokardiyal Enfarktüs
- Miyokardiyal İskemi
- Sol Ventriküler Büyümesi
- Atrial Büyümesi

Uzmanlardan alınan geri bildirimlerden ve araştırmalar sonucu elde edilen bulgulara göre, bu hastalıkların kalp krizine sebep olan en önemli hastalıklar olduğu saptanmıştır. Bu durum bu çalışmanın önemini ve gerekliliğini artırmakta ve üzerinde çalışılması gerekliliği konusundaki durumunu ortaya çıkarmaktadır. Çalışmada bu hastalıklar aynı zamanda birer sınıf olarak da adlandırılmıştır. Bu hastalıkların veya sınıfların yanı sıra bunlara ek olarak sağlıklı bireylerin EKG sinyallerinden oluşan “Normal Birey” sınıfı da eklenmiştir. Bu sınıflar aynı zamanda tasarlanmış olan modelin de sınıflarını oluşturmaktadır.

4.2. Sınıfların EKG Belirtileri

Belirlenmiş olan hastalıkların her biri bir sınıf olarak tanımlanmıştır. Her bir hastalığın veya sınıfın kendine has nitelikleri mevcuttur. Bu nitelikler sayesinde hastalığın tespiti ve teşhisi yapılmaktadır. Bu nitelikler bazen sadece bir hastalığa ait bir belirti olurken bazen de birden fazla hastalık için ortak olabilir. Örneğin, *R dalga progresyon kaybı* niteliği sadece ME hastalığına ait iken; *ST segment yükselmesi*, MİS, ME ve SVM hastalıkları için ortak bir niteliktir. Hastalıklar ve nitelikleri arasındaki bağıntı, bu ortak niteliklere bağlı olarak karmaşık veya basit olabilmektedir.

Çalışmada her bir sınıf için genel EKG belirtileri Çizelge 4.1’de verilmiştir. Bu çizelgede belirtiler sınıflarına veya hastalıklarına göre gruplandırılmışlardır.

Çizelge 4.1. Sınıflar için EKG belirtileri

Sınıf	EKG Belirtileri
Miyokardiyal İskemi	ST Segment Çökmesi
	ST Segment Yükselmesi
	T Dalgası Negatif Olması
	T Dalgasının Sivri Olması (Pozitif Olması)
	Arritmeler
	Atriyoventriküler (AV) bloklar
	QT Aralığı Uzaması
	Negatif U Dalgası
	Dal Bloğu Olması
	QRS Atternansı
	PR Aralık Anormal Olma
Miyokardiyal Enfarktüs	ST Segment Çökmesi
	ST Segment Yükselmesi
	R Dalga Progresyon Kaybı
	T Dalgasının Sivri Olması (Pozitif Olması)
	Atriyal aritmiler
	PR Aralık Anormal Olma
	Patolojik Q Dalgası Olması
Sol Ventriküler Büyümesi	QRS Yükseklik Artışı
	ST Segment Yükselmesi
	QT Aralık Uzaması
	QRS Genişlik(Süre) Uzaması
	T Dalgası Negatif Olması
	Sol aks sapması
	R Dalga Progresyon Kaybı
Atrial Büyümesi	P Dalgası Uzaması
	P Dalgası Kısa olması
	Negatif P Dalgası Olması
	P Dalgası Çentikli(Çift tepe) olması
	P Dalgasının Düşük olması
	P Dalgası Yüksek Olması

4.3. Nitelik Belirleme

Bir EKG sinyalindeki dalga, aralık ve segmentler EKG parametreleri olarak isimlendirilebilir. Bu parametreler sayesinde EKG yorumlanır. Bu parametreler sayesinde kalp ile ilgili bir rahatsızlığın teşhisi mümkün olmaktadır. Bu parametreler ne kadar doğru okunup yorumlanırsa o kadar sağlıklı bir teşhis koyma işlemi gerçekleşmiştir denilebilir. Dikkat edilmemesi durumunda ise hayati derece önemli durumların gözden kaçmasına neden olabilir ve bu da insan hayatına mal olabilir. Bu noktada insanoğlunun dikkatsizliği, dalgınlığı, tutarsızlığı ve karmaşık yapıları değerlendirmedeki yetenekleri düşünüldüğünde insan yerine bilgisayar gibi araçlar ile bu işlemlerin yapılması daha sağlıklı olduğu açıktır. Ancak bu araçların tek başına bu iş için yeterli olamayacağı, bu konuda uzman kişilere her zaman ihtiyaç duyulacağı da unutulmamalıdır.

EKG parametreleri, asıl ve asıl olmayan parametreler olarak sınıflandırılabilir. hem literatürde hem de doktorlar ile yapılan birebir görüşmelerde saptanmıştır. Çalışmada belirleyici olacak parametreler, nitelik olarak alınmıştır. Örneğin, *Atriyovenriküler bloklar*, *Arritmeler*, *QT aralığı uzaması*, *Negatif U dalgası*, *Dal bloğu olması*, *QRS atternansı*, *Atriyal aritmiler*, *Sol aks sapması* belirtilerin hastalık belirlemede zayıf belirtiler olduğu görülmüştür. Bu sebeple Çizelge 4.1’de verilen bu belirtiler göz ardı edilmişlerdir. Bu belirtiler aynı zamanda asıl olmayan EKG parametreleridir. Çizelge 4.2’deki tüm parametreler asıl EKG parametrelerinden alınmıştır. Mesela, MİS için *ST segment çökmesi* veya *yükselmesi* asıl olarak tanımlanabilecek birer nitelik veya EKG parametresidir.

Açıkçası, bu noktada gerek uzman hekimler ile yapılan görüşmelerden çıkarılan bilgiler ile gerekse literatür çalışmasıyla bir ön işlem (pre-processing) işlemi yapılmıştır. Bu ön işlem ile hem nitelik indirgemesi hem de sınıf ile nitelik arasındaki ilişkide sadelik sağlanmıştır.

Çizelge 4.2. Seçilen nitelikler

Nitelik	Açıklama
Rate	Kalp atış hızı.
ST segment	S dalgası sonu ile T dalgası arasındaki yassı, izoelektrik EKG kısmıdır. Ventriküler depolarizasyonu ve repolarizasyonu gösterir.
T dalgası	T dalgası her QRS kompleksinden sonraki pozitif saptadır. Ventrikül repolarizasyonunu temsil eder [1].
Patolojik Q	Aşağıdakilerin varlığında Q dalgaları patolojik olarak kabul edilir: <ul style="list-style-type: none">• > 40 ms (1 mm) genişlik• > 2 mm derinlik• QRS kompleksinin %25'inden daha fazla derinlik• V1-3'te görülmesi
R dalga progresyon kaybı	V3'te ≤ 3 mm R dalga yüksekliği.
PR aralık	PR aralığı P dalgası ile QRS kompleksi arasında geçen süredir. AV düğümünden geçen iletiyi gösterir [1].
P dalgası yükseklik	P dalgasının yüksekliği normal sınırlardadır ancak süresi 120 ms'den uzundur [1].
P dalgası geniş	Genişliği değişmeyen (< 120 ms) ancak normalden daha yüksek P dalgaları meydana gelir (2,5 mm) [1].
QRS genişlik	(QRS > 100 ms) Hem ventriküler kaynaklı hem de supraventriküler komplekslerin aberran iletimine bağlı olabilmektedir [1].
QRS yükseklik	Voltajın normalinden yüksek olma durumudur.
QTc	<ul style="list-style-type: none">• Düzeltilmiş QT aralığı (QTc), 60/dk kalp hızındaki QT aralığını hesaplar.• QT değerlerinin değişik kalp hızlarında karşılaştırılmasına ve artmış aritmi riski olan hastaların saptanmasına olanak verir.
T-reciprocity	T dalgasındaki karşıtlık.
QT aralık	<ul style="list-style-type: none">• QT aralığı Q dalga başlangıcından T dalga sonuna kadar geçen süredir.• Ventriküler depolarizasyon ve repolarizasyon için geçen süreyi belirtir [1].
Cinsiyet	Hastanın Cinsiyeti

4.4. Niteliklerin Değer Kümesinin Belirlenmesi

Her hastalık için nitelikler belirlendikten sonra, her niteliğin değer kümesini (domain value) belirlemek gerekir. Her niteliğin belli bir değeri veya değer aralığı mevcuttur. Niteliklerin bu değerde ya da değer aralığında olması normal [28], dışında olması da anormal olarak değerlendirilir. Çizelge 4.3’de her nitelik için normal değer veya değer aralığı verilmiştir. Bir niteliğin bir hastalık için nitelik olabilmesi için normal değer veya değer aralığında olmaması gerekir. Niteliğin Çizelge 4.3’deki değerlerin dışında olması bir anormalliği, dolayısıyla bir hastalığın belirtisini ifade eder. Bir niteliğin tek başına anormal olması bir hastalığı tek başına belirmesi için yeterli olmayabilir. Bunun için birden fazla nitelik bir araya gelerek bir hastalığı belirtebilir. Örneğin, *ST segment çökmesi*, *ST segment yükselmesi* ve *T dalgasının negatif* olması üçü birlikte MİS hastalığını işaret ederler.

Yapılan çalışmada her nitelik için değer kümesi belirlemesi Çizelge 4.4’te verilmiştir. Her nitelik için değer kümesi küme halinde aşağıdaki gibi yazılabilir.

$Rate = \{B, N, T\}$, $ST\ segment = \{C, N, Y\}$, $T\ dalgası = \{P, N, Ne\}$,
 $Patolojik\ Q = \{V, Y\}$, $R - progresyon = \{V, Y\}$, $PR\ aralık = \{N, A\}$,
 $P\ dalgası\ yükseklik = \{N, A\}$, $P\ dalgası\ geniş = \{N, A\}$, $QRS\ genişlik =$
 $\{K, N, U\}$, $QRS\ yükseklik = \{D, N, Y\}$, $QTc = \{N, A\}$, $T - reciprocity =$
 $\{V, Y\}$, $QT\ aralık = \{K, N, U\}$, $Cinsiyet = \{C, YK, YE\}$

Değer kümesindeki her bir değer farklı anlamları mevcuttur. Örneğin, bunlardan *ST segmenti* için; C: Çökme, N: Normal (sağlıklı), Y: Yükselmeyi ifade ederken, *R-Progresyonu* için; V: Var olduğunu, Y: Var olmadığını göstermektedir (R-Progresyonun olmadığını). Diğer nitelikler için de bu şekilde örnekler çoğaltılabilir.

Çizelge 4.3. Her nitelik için normal değer veya değer aralığı

Nitelik	Normal Değer, Değer Aralığı
Rate	60-100 beats/min.
ST segment	Çökme ya da Yükselmenin olmaması.
T dalgası	Sivri ya da Negatif olmaması.
Patolojik Q	Patolojinin varlığının olmaması.
R dalga progresyon kaybı	R dalgasının V3'te 3 mm'ten yüksek olmaması.
PR aralık	Normal PR aralığı 120 – 200 ms süreleri arasındadır.
P dalgası yükseklik	Amplitüdü (yüksekliği) < 2,5 mm (0.25mV).
P dalgası geniş	Süresi (Genişliği) < 120 ms (üç küçük kare),
QRS genişlik	<ul style="list-style-type: none">• QRS genişliği 4 yaş altındaki çocuklarda 70 - 90 ms,• 4-16 yaş arası çocuklarda 70 -100 ms,• 16 yaş üzerindekielerde 70 -110 ms olmalıdır.
QRS yükseklik	R dalga yüksekliği > 35 mm olmaması
QTc	<ul style="list-style-type: none">• Erkeklerde > 440 ms ya da kadınlarda > 460 ms ise QTc uzamıştır.• QTc > 500 ms olması artmış torsades de pointes riski ile ilişkilidir.• < 350 ms olması durumunda QTc anormal olarak kısadır.• Kullanışlı bir pratik kural olarak normal QT öncesindeki RR aralığının yarısından azdır.
T-reciprocity	Karşıtlığın olmaması
QT aralık	<ul style="list-style-type: none">• 1-15 yaş arasında<ul style="list-style-type: none">▪ Normal: < 440 ms▪ Sınırdaki: 440 - 460 ms▪ Uzamış: > 460 ms• Yetişkin erkekte<ul style="list-style-type: none">▪ Normal: < 430 ms▪ Sınırdaki: 430 - 450 ms▪ Uzamış: > 450 ms• Yetişkin kadında<ul style="list-style-type: none">▪ Normal: < 450 ms▪ Sınırdaki: 450 - 470 ms▪ Uzamış: > 470 ms
Cinsiyet	Çocuk, Yetişkin Erkek, Yetişkin Kadın

Çizelge 4.4. Niteliklerin değer kümeleri

Nitelik	Koşul	Domain	Simge
Rate	< 60 beats/min	Bradikardi	B
Rate	60-100 beats/min	Normal	N
Rate	100 <	Taşikardi	T
ST segment	Negatif olma	Çökme	C
ST segment	Pozitif olma	Yükselme	Y
ST segment	Normal olma	Normal	N
T dalgası	+	Sivri (Pozitif)	P
T dalgası	-	Negatif	Ne
T dalgası	Normal olma	Normal	N
Patolojik Q	Var	Var	V
Patolojik Q	Yok	Yok	Y
R-progresyon	Var	Var	V
R-progresyon	Yok	Yok	Y
PR aralık	120 – 200 ms	Normal	N
PR aralık	< 120 – 200 ms <	Anormal	A
P dalgası yükseklik	< 2.5 mm	Normal	N
P dalgası yükseklik	< 2.5 mm <	Anormal	A
P dalgası geniş	< 120 ms	Normal	N
P dalgası geniş	< 120 ms <	Anormal	A
QRS genişlik	< 70	Kısa	K
QRS genişlik	70-110 ms	Normal	N
QRS genişlik	110 ms <	Uzun	U
QRS yükseklik	< 10 mm	Düşük	D
QRS yükseklik	10- 35 mm	Normal	N
QRS yükseklik	35 mm <	Yüksek	Y
QTc	440 - 460	Normal	N
QTc	< 440 - 460 <	Anormal	A
T-reciprocity	Var	Var	V
T-reciprocity	Yok	Yok	Y
QT aralık	< 330	Kısa	K
QT aralık	430 - 470	Normal	N
QT aralık	470 <	Uzun	U
Cinsiyet	< 16 yaş	Çocuk	C
Cinsiyet	16 yaş bayan <	Yetişkin Kadın	YK
Cinsiyet	16 yaş bay <	Yetişkin Erkek	YE

4.4.1. Niteliklerin deęer kümesinin sayısallaştırılması

Bu çalışmada uygulama kısmında sayısal veriler üzerinde çalışıldığında, tüm niteliklerin deęer kümeleri sayısal deęerlere dönüştürülmüştür. Aşağıda Çizelge 4.5'te her deęer kümesinin sayısal karşılığı verilmiştir.

Çizelge 4. 5. Deęer kümesinin sayısal karşılıkları

Nitelik	Simge	Sayısal Karşılığı
Rate	= { B, N, T }	= { 1, 2, 3 }
ST segment	= { C, Y }	= { 1, 2, 3 }
T dalgası	= { P, N, Ne }	= { 1, 2, 3 }
Patolojik Q	= { V, Y }	= { 1, 2 }
R-progresyon	= { V, Y }	= { 1, 2 }
PR aralık	= { N, A }	= { 1, 2 }
Cinsiyet	= { C, YK, YE }	= { 1, 2, 3 }
P dalgası yükseklik	= { N, A }	= { 1, 2, 3 }
P dalgası geniş	= { N, A }	= { 1, 2 }
QRS genişlik	= { D, N, U }	= { 1, 2, 3 }
QRS yükseklik	= { K, N, Y }	= { 1, 2, 3 }
QTc	= { N, A }	= { 1, 2 }
T-reciprocity	= { V, Y }	= { 1, 2 }
QT aralık	= { N, A }	= { 1, 2 }

5. EKG SİNYALLERİNİN SINIFLANDIRMASI

5.1. Kaba Kümeler Teorisi

Pawlak tarafından 1982 yılında ortaya atılan KKT, eksik bilgi sistemleri için yeni bir matematiksel yaklaşımdır [29, 30]. Doğada var olan problemler ve olaylar hakkında belirsizlikler ve eksik bilgi kümeleri sıklıkta vardır. KKT ve bulanık mantık gibi kuramlar eksik ve belirsizlik bilgi tabanından faydalı bilgi çıkarımını sağlayan teoremlerdir [3]. Şüpheli, eksik, tutarsız gizli bilgilerin sık olduğu sağlık alanında KKT kendine geniş bir çalışma alanı bulmuştur.

5.1.1. Bilgi sistemi

KKT'nin ilk aşamasını oluşturan bilgi sistemin, satırları nesnelere (örnekleri, numuneleri (n: nesne sayısı)), sütunları nesnelere nitelikleri (m: nitelik sayısı) göstermek üzere bir bilgi sistemi $n \times m$ boyutunda bir matristir. Bu bilgi sistemi, problem ile ilgili en geniş ve kapsamlı bilgi içeren sistemdir. KKT'de bir bilgi sistemi aşağıdaki gibi tanımlanır:

$$IS = (U, A) \quad (5.1)$$

U = Boş olmayan sonlu nesne kümesi ($U = \{x_1, x_2, \dots, x_n\}$),

A = Boş olmayan sonlu nitelik kümesi ($A = \{a_1, a_2, \dots, a_m\}$) olarak tanımlanır. Her $a \in A$ ve $a: U \rightarrow V_a$ olmak üzere, V_a gibi bir değer kümesine sahiptir.

5.1.2. Ayırt edilemezlik

Bir karar çizelgesi (tablosu), var olan problem veya oluşturulmuş olan model hakkında geniş ve kapsamlı bir şekilde bilgileri barındırmaktadır. Buna bağlı olarak bu bilgi sistemi gereksiz yere çok büyümüş, genişlemiş olabilir. Örneğin, aynı nesne ya da ayırt edilemeyen nesnelere birden fazla tekrarlanmış olabilir ya da bazı nitelikler gereksiz veya fazladan bilgi sisteminde bulunabilirler [31]. (5.1) için, herhangi bir $B \subseteq A$ olduğunda, B özelliklerinin her alt kümesi ayırt edilemezlik bağıntısı adı verilen U evreninin bir denklik bağıntısını tanımlar. (x, y) ikilisi, U 'dan

nesne çiftleri olmak üzere, aşağıda tanımlanan bir $IND_S(B)$ denklik bağıntısına *B-ayırt edilemezlik* bağıntısı denir [32].

$$IND_S(B) = \{(x, y) \in U^2 \mid \forall a \in B \ a(x) = a(y)\} \quad (5.2)$$

(5.2) 'de S alt indisi, hangi bilgi sisteminde bahsedildiği açıksa genellikle ihmal edilir. Eğer $(x, y) \in IND_S(B)$ ise, x ve y nesnelere B 'ye göre ayırt edilemezdirler denilir.

5.1.2.1 Ayırt edilebilirlik matrisi

(5.1) bağıntısında n adet nesnesi olan bir bilgi sistemi için ayırt edilebilirlik (discernibility) matrisi $(M_{i,j})$, aşağıda verilen $Z_{i,j}$ elemanları olan simetrik bir $n \times n$ matristir. Bu matrisin her $Z_{i,j}$ elemanı, x_i ve x_j nesnelere farklı kılın özellikler kümesinden oluşmaktadır [29].

$$Z_{i,j} = \{a \in A \mid a(x_i) \neq a(x_j) \text{ and } d_i \neq d_j\} \quad i, j = 1, 2, \dots, n \quad (5.3)$$

$$M_{i,j} = \begin{bmatrix} Z_{1,1} & \dots & Z_{1,n} \\ \vdots & \ddots & \vdots \\ Z_{n,1} & \dots & Z_{n,n} \end{bmatrix}$$

5.1.2.2 Ayırt edilebilirlik fonksiyonları

Ayırt edilebilirlik fonksiyonları, ayırt edilebilirlik matrisindeki elemanlar kendi aralarında *boolean* cebire göre "VEYA" işlemine, birbirleri arasında ise "VE" işleme tabi tutulmasıyla elde edilen lojik fonksiyonlardır. Ayırt edilebilirlik fonksiyonu indirgenerek indirgenmiş nitelik kümeleri bulunur. Bir bilgi sistemi için ayırt edilebilirlik fonksiyonu f_S , a_1, \dots, a_m özelliklerine karşılık gelen m adet a_1^*, \dots, a_m^* *boolean* değişkeninin bir *boolean* fonksiyonudur [29].

$$f_S(a_1^*, \dots, a_m^*) = \bigwedge \{ \bigvee Z_{i,j}^* \mid 1 \leq j \leq i \leq n, Z_{i,j}^* \neq \emptyset \}$$

$$\text{Burada } Z_{i,j}^* = \{a \mid a^* \in Z_{i,j}\} \quad (5.4)$$

5.1.3. Karar sınıfları

d karar özelliği U nesnelere evreninin bir bölüntüsünü tüme varır. Genellemeyi bozmadan V_d 'nin $\{1, \dots, r(d)\}$ tamsayılar kümesi olduğu kabul edilebilir. Burada $r(d)$, d 'nin rankını belirtmektedir.

Dolayısıyla tüme varılan bölüntüler; $\{X_1, \dots, X_{r(d)}\}$ "karar sınıfları" denilen denklik sınıflarının bütünüdür. Burada iki nesne karar özelliği için aynı değere sahipse, aynı karar sınıfına ait oldukları söylenir.

$$X_i = \{x \in U \mid d(x) = i\} \quad (5.5)$$

Bir S karar sistemi ve $B \subseteq A$ ise, S 'nin B -genelleştirilmiş kararı adı verilen

$\delta_B: U \rightarrow P(V_d)$ fonksiyonu şöyle tanımlanır:

$$\delta_B(x) = \{v \in V_d: \exists X' \in U (X' \text{ IND}_S(B)_x \text{ ve } d_x = v)\} \quad (5.6)$$

δ_S : S -genelleştirilmiş kararına S 'nin genelleştirilmiş kararı denir.

Bir S karar sistemi, ancak herhangi bir $x \in U$ için $|\delta_S(x)| = 1$ ise tutarlıdır, aksi takdirde tutarlı değildir. Buradan bir A karar tablosu ancak ve ancak $\text{POSA}(d)=U$ ise tutarlı olduğu kolaylıkla görülebilir [32].

5.1.4. Yaklaşım kümeleri

KKT'nin temel amacı, $\text{IND}(B)$ ikili bağıntısını kullanarak kümelerin yaklaşımlarını belirlemektir. Eğer X nesne kümesi, A nitelik kümesinin özelliklerini kullanarak kesin bir biçimde tanımlanamıyorsa alt ve üst yaklaşımlar ile gösterilebilir.

5.1.4.1 Alt yaklaşım kümesi

Alt yaklaşım kümesindeki nesnelere, kümeye kesin olarak ait olan nesnelere.

Alt yaklaşım kümesi aşağıdaki gibi tanımlanır:

$$\underline{BX} = \{X_i \in U \mid [X_i] \text{ IND}(B) \subset X\} \quad (5.7)$$

5.1.4.2 Üst yaklaşım kümesi

Üst yaklaşım kümesindeki nesnelere, kümeye ait olma olasılığı olan nesnelere.

Üst yaklaşım kümesi aşağıdaki gibi tanımlanır:

$$\overline{BX} = \{X_i \in U \mid [X_i] \text{ IND}(B) \cap X \neq \emptyset\} \quad (5.8)$$

Sınır Bölgesi: $BN_B(X) = \overline{BX} - \underline{BX}$ olmak üzere, $BN_B(X)$ sınır bölgesidir. Bu küme X 'in elemanı olup olmadığına karar verilemeyen birimlerden oluşmaktadır. Eğer sınır bölgesi boş küme değilse kaba bir kümenin varlığı söz konusudur.

- Eğer Alt ve Üst küme yaklaşımları belli ise $X \subset U$ seti belirlidir. Aksi durumda da belirsizdir.
- $X \subset U$ olmak koşuluyla X seti için dört belirsiz durum vardır:
 1. Eğer $\underline{BX} \neq \emptyset$ ve $BX \neq U$ ise X kabaca belirlenir.
 2. Eğer $\underline{BX} \neq \emptyset$ ve $BX = U$ ise X dıştan belirsizdir.
 3. Eğer $\underline{BX} = \emptyset$ ve $BX \neq U$ ise X içten belirsizdir.
 4. Eğer $\underline{BX} = \emptyset$ ve $BX = U$ ise X tamamıyla belirsizdir.

Tamlık Fonksiyonu (Accuracy of approximation): Karar kurallarının doğruluğunun bir ölçüsü olan tamlık fonksiyonu:

$$M_B(X) = \text{card}(\underline{BX}) / \text{card}(BX) \quad (5.9)$$

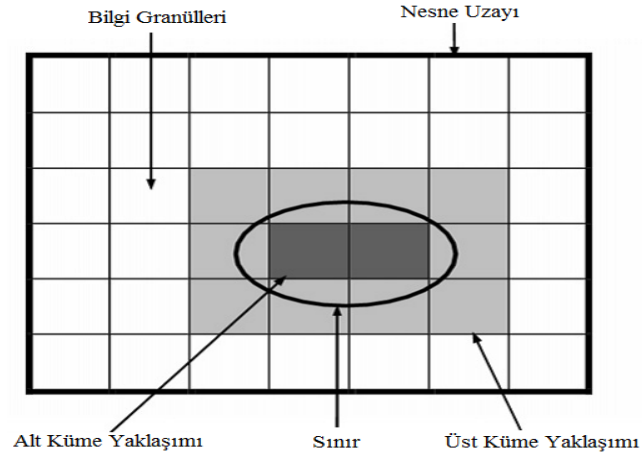
Tamlık Fonksiyonu (Accuracy)'nin değeri;

$0 \leq M_B(X) \leq 1$ arasında bir değerdir. Bu değer 1'e ne kadar yakınsa ayırt edilebilirlik o derece iyidir.

Bağımsız Nesnelere: $a_i \in A$ olmak üzere;

Eğer $\text{IND}(A) = \text{IND}(A - a_i)$ ise a_i niteliği gereksiz (superfluous) olarak nitelendirilir.

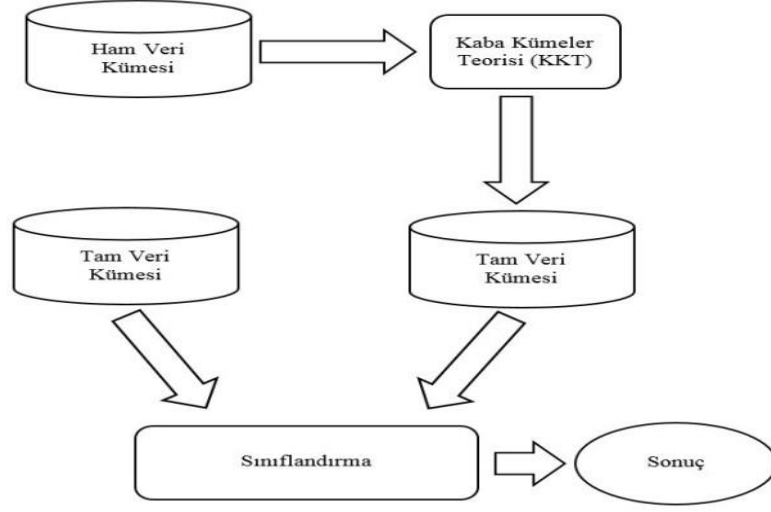
Aksi durumda ise a_i niteliği zorunlu niteliktir.



Şekil 5.1. Alt ve Üst küme yaklaşımları

5.2. Önerilen Model

Sınıflandırma yöntemleri tam veri kümeleri üzerinde daha iyi performans gösterirler. Bunun nedeni tam veri kümelerinde eksik, tutarsız ve tekrarlı verinin olmamasıdır. Başka bir deyişle eksik veri kümesinde eksik ve belirsiz verilerin olması sınıflandırma metotlarının performanslarını düşürmektedir. Kaba kümeler gibi araçlar; eksik, belirsiz ve tutarsız veri kümelerini tam veri kümelerine dönüştürürler. Eğer bir veri kümesinde sınıflandırma işlemi yapılacaksa önce veri kümesinin kaba kümeler gibi yöntemlerden geçirilmesi sınıflandırma işleminin performansını artırabilir. Kısacası, bu çalışmada öngörülen ve önerilen model olarak, eksik veri kümesinin sınıflandırma yöntemine tabi tutulmadan önce kaba kümeler ile tam veri kümesine dönüştürülmesidir. Aşağıda önerilen modelin şeması verilmiştir.



Şekil 5.2. Önerilen model

Sınıflandırma yapılırken, bir nesnenin hangi sınıfa ait olduğunu bulmak, nesnelere ait niteliklerin ayırt ediciliğine bağlıdır. Bir niteliğin değer kümesindeki değeri, birden fazla sınıf için ortak olabilir. Örneğin, ST segment niteliğinin ST segment çökmesi değeri, ME ve MİS sınıfları için ortak bir değerdir. Bu ortak değerler karmaşıklık yaratmakta ve ayırt ediciliği azaltmaktadır. Ayrıca bir bilgi sistemindeki niteliklerden bazıları gereksiz olabilir. Bu gereksiz nitelikler gizli sınıflandırmaya, fazladan iş yüküne neden olabilmektedir. Bu çalışmada bu problemlerden gereksiz nitelik tespiti yapılarak fazladan yapılan iş yükü azaltılması amaçlanmıştır. Dolayısıyla sınıflandırmaların performanslarında iyileşmeler sağlanması hedeflenmiştir. Çözüm olarak KKT ile çözüm sunulmuştur. Gereksiz, tekrarlı ve eksik bilgi içeren ham veri kümesi KKT için bilgi sistemidir. KKT ayırt edilebilirlik matrisi ve ayırt edilebilirlik fonksiyonları kullanarak bu bilgi sistemini, özgünlüğünü bozmadan gereksiz ve tekrarlı bilgiden arındırarak tam veri kümesine dönüştürür. Aynı zamanda nesne ile nitelik arasındaki bağıntı karmaşıklığını sadeleştirir [2].

Ayırt edilebilirlik matrisi ve ayırt edilebilirlik fonksiyonları işlemleri için ROSE (Rough Set Data Explorer) programı kullanılmıştır. ROSE programı KKT'nin temel işlevlerini yerine getiren ve kural çıkarımı yapan bir modül yazılımıdır [33].

Sınıflandırma işlemi için WEKA (Waikato Environment for Knowledge Analysis) adlı program kullanılmıştır. WEKA bugün birçok insan tarafından kullanılan Veri Madenciliği uygulaması geliştirme programıdır. Bu program JAVA platformu üzerinde geliştirilmiş açık kaynak kodlu bir programdır [34]. WEKA’da çeşitli sınıflandırma yöntemleri mevcuttur. Önerilen modelde yüksek doğruluk oranına sahip sınıflandırma yöntemleri seçilmiştir.

Aşağıda Örnek 5.1 üzerinde sınıflandırma işlemleri yapılmadan KKT ile hem nitelik hem de nitelik değerinde indirgemenin nasıl yapıldığını göstermektedir. Burada amaç, gereksiz nitelik tespitinin nasıl yapıldığını göstermek ve bu nitelikler olmadan da ilk baştaki bilgi sisteminin, işlemlerden sonra elde edilen alt bilgi sistem ile tüm özellikleri taşıdığı bilgisini vermektir. Yani gerçek bir uygulama için tekrarlı ve gereksiz bütün bilgileri içeren bir bilgi sistemi yerine, bilgi sistemine denk tekrarsız ve gereksiz veriden arındırılmış bir alt bilgi sistemi sunmak.

Örnek 5.1:

IS = (U, A, D) bilgi sistemi verilsin.

Çizelge 5.1. Bilgi sistemi

U	a ₁	a ₂	a ₃	d ₁
x ₁	2	1	3	1
x ₂	3	2	1	2
x ₃	2	1	3	1
x ₄	2	2	3	2
x ₅	1	1	4	3
x ₆	1	1	2	3
x ₇	3	2	1	2
x ₈	1	1	4	3
x ₉	2	1	4	1
x ₁₀	3	2	1	2

$$U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$$

$$A = \{a_1, a_2, a_3\}$$

$$D = \{d_1\}$$

$$Va_1 = \{1, 2, 3\}$$

$$Va_2 = \{1, 2\}$$

$$Va_3 = \{1, 2, 3, 4\}$$

Çizelge 5.1 aynı zamanda bir karar tablosudur. Bu karar tablosu üç karar niteliği {a₁, a₂, a₃} ve bir tane karar niteliği {d₁} içermektedir.

❖ *D-çekirdek (D-core), D-indirgenen (D-reducts) ve D-gereksiz (D-superfluous) nitelik tespiti:*

Çekirdek ve gereksiz nitelikleri tespit etmek için ayırt edilemezlik matrisi (5.3) kullanılarak elde edilir.

Çizelge 5.2. Ayırt edilemezlik matrisi

	1	2	3	4	5	6	7	8	9	10
1	-									
2	a ₁ , a ₂ , a ₃	-								
3	-	a ₁ , a ₂ , a ₃	-							
4	a ₂	-	a ₂	-						
5	a ₁ , a ₃	a ₁ , a ₂ , a ₃	a ₁ , a ₃	a ₁ , a ₂ , a ₃	-					
6	a ₁ , a ₃	a ₁ , a ₂ , a ₃	a ₁ , a ₃	a ₁ , a ₂ , a ₃	-	-				
7	a ₁ , a ₂ , a ₃	-	a ₁ , a ₂ , a ₃	-	a ₁ , a ₂ , a ₃	a ₁ , a ₂ , a ₃	-			
8	a ₁ , a ₃	a ₁ , a ₂ , a ₃	a ₁ , a ₃	a ₁ , a ₂ , a ₃	-	-	a ₁ , a ₂ , a ₃	-		
9	-	a ₁ , a ₂ , a ₃	-	a ₂	a ₁	a ₁ , a ₃	a ₁ , a ₂ , a ₃	a ₁	-	
10	a ₁ , a ₂ , a ₃	-	a ₁ , a ₂ , a ₃	-	a ₁ , a ₂ , a ₃	a ₁ , a ₂ , a ₃	-	a ₁ , a ₂ , a ₃	a ₁ , a ₂ , a ₃	-

Denklem 5.4'den ayırt edilemezlik fonksiyonu ($f_A(D)$);

$$\begin{aligned}
 f_A(D) &= (a_1 \vee a_2 \vee a_3) \wedge a_2 \wedge (a_1 \vee a_3) \wedge (a_1 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_3) \wedge \\
 &\quad (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \wedge \\
 &\quad (a_1 \vee a_2 \vee a_3) \wedge a_2 \wedge (a_1 \vee a_3) \wedge (a_1 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \\
 &\quad \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \wedge a_2 \\
 &\quad \wedge (a_1 \vee a_2 \vee a_3) \wedge a_1 \wedge (a_1 \vee a_2 \vee a_3) \\
 &\quad \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \\
 &\quad \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \\
 &\quad \wedge a_1 \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) = a_1 \wedge a_2
 \end{aligned}$$

Burada, D-indirgeme veya D-çekirdek $\{a_1, a_2\}$ ve D-gereksiz nitelik $\{a_3\}$ küme tespiti var. Bunun anlamı, Çizelge 5.1'in indirgenebileceği ve buna alternatifi olarak Çizelge 5.3 gibi bir karar tablosunun oluşturulabileceğidir.

Bu örnekte D-indirgeme ve D-çekirdek nitelikler aynı çıktı. Örneğin, Eğer $f_A(D) = a_1 \wedge a_2 \vee a_2 \wedge a_3$ çıksaydı o zaman D-indirgeme $\{a_1, a_2\}$ ve $\{a_2, a_3\}$ iken D-çekirdek $\{a_2\}$ olurdu.

Çizelge 5.3. D-çekirdek nitelikler için bilgi sistemi

U	a ₁	a ₂	d ₁
x ₁	2	1	1
x ₂	3	2	2
x ₃	2	1	1
x ₄	2	2	2
x ₅	1	1	3
x ₆	1	1	3
x ₇	3	2	2
x ₈	1	1	3
x ₉	2	1	1
x ₁₀	3	2	2

❖ **Nitelik değer indirgeme:**

Çizelge 5.3 üzerinde ayırt edilemezlik matrisi oluşturulur ve her nesne için ayırt edilemezlik fonksiyonları hesaplanır. Buna bağlı olarak nitelik değerlerinden indirgenen değerler bulunur.

Çizelge 5.4. Ayırt edilemezlik matrisi

	1	2	3	4	5	6	7	8	9	10
1	-	a ₁ , a ₂	-	a ₂	a ₁	a ₁	a ₁ , a ₂	a ₁	-	a ₁ , a ₂
2	a ₁ , a ₂	-	a ₁ , a ₂	-	a ₁ , a ₂	a ₁ , a ₂	-	a ₁ , a ₂	a ₁ , a ₂	-
3	-	a ₁ , a ₂	-	a ₂	a ₁	a ₁	a ₁ , a ₂	a ₁	-	a ₁ , a ₂
4	a ₂	-	a ₂	-	a ₁ , a ₂	a ₁ , a ₂	-	a ₁ , a ₂	a ₂	-
5	a ₁	a ₁ , a ₂	a ₁	a ₁ , a ₂	-	-	a ₁ , a ₂	-	a ₁	a ₁ , a ₂
6	a ₁	a ₁ , a ₂	a ₁	a ₁ , a ₂	-	-	a ₁ , a ₂	-	a ₁	a ₁ , a ₂
7	a ₁ , a ₂	-	a ₁ , a ₂	-	a ₁ , a ₂	a ₁ , a ₂	-	a ₁ , a ₂	a ₁ , a ₂	-
8	a ₁	a ₁ , a ₂	a ₁	a ₁ , a ₂	-	-	a ₁ , a ₂	-	a ₁	a ₁ , a ₂
9	-	a ₁ , a ₂	-	a ₂	a ₁	a ₁	a ₁ , a ₂	a ₁	-	a ₁ , a ₂
10	a ₁ , a ₂	-	a ₁ , a ₂	-	a ₁ , a ₂	a ₁ , a ₂	-	a ₁ , a ₂	a ₁ , a ₂	-

$$f_1(A) = (a_1 \vee a_2) \wedge a_2 \wedge a_1 \wedge a_1 \wedge (a_1 \vee a_2) \wedge a_1 \wedge (a_1 \vee a_2) = a_1 \wedge a_2$$

$$f_2(A) = (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) = a_1 \vee a_2$$

$$f_3(A) = (a_1 \vee a_2) \wedge a_2 \wedge a_1 \wedge a_1 \wedge (a_1 \vee a_2) \wedge a_1 \wedge (a_1 \vee a_2) = a_1 \wedge a_2$$

$$f_4(A) = a_2 \wedge a_2 \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge a_2 = a_2$$

$$f_5(A) = a_1 \wedge (a_1 \vee a_2) \wedge a_1 \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge a_1 \wedge (a_1 \vee a_2) = a_1$$

$$f_6(A) = a_1 \wedge (a_1 \vee a_2) \wedge a_1 \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge a_1 \wedge (a_1 \vee a_2) = a_1$$

$$f_7(A) = (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) = a_1 \vee a_2$$

$$f_8(A) = a_1 \wedge (a_1 \vee a_2) \wedge a_1 \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge a_1 \wedge (a_1 \vee a_2) = a_1$$

$$f_9(A) = (a_1 \vee a_2) \wedge a_2 \wedge a_1 \wedge a_1 \wedge (a_1 \vee a_2) \wedge a_1 \wedge (a_1 \vee a_2) = a_1 \wedge a_2$$

$$f_{10}(A) = (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2) = a_1 \vee a_2$$

Bu örnekte $\{a_1, a_2\}$ D-çekirdek nitelikler olduğundan yeni karar tablomuz Çizelge 5.5'tir.

Çizelge 5.5. D-çekirdek nitelik değer kümesi indirgenmiş bilgi sistemi

U	a ₁	a ₂	d ₁
x ₁	2	1	1
x ₂	3	2	2
x ₃	2	1	1
x ₄	*	2	2
x ₅	1	*	3
x ₆	1	*	3
x ₇	3	2	2
x ₈	1	*	3
x ₉	2	1	1
x ₁₀	3	2	2

* İhmal edilen değeri gösterir.

❖ **Çizelge 5.5'ten kural türetmek:**

$$a_{1_2}a_{2_1} \rightarrow d_1 \qquad a_{2_2} \rightarrow d_2$$

$$a_{1_1} \rightarrow d_3 \qquad a_{1_3}a_{2_2} \rightarrow d_2$$

Burada, $a_{1_2}a_{2_1} \rightarrow d_1$ 'nın anlamı; $a_1 = 2$ ve $a_2 = 1$ iken $d_1 = 1$ olmaktadır. Yani niteliklerden birincisi 2, ikincisi 1 değerini aldığıında nesne '1' sınıfına aittir anlamına gelir.

5.3. Uygulama

5.3.1. Veri kümesi hakkında

Uygulamada kullanılan veri kümesi ESOGÜ Tıp Fakültesi Kardiyoloji Anabilim Dalı doktorların yardımları ile elde edilmiştir. Veri kümesi, 689 hastanın gerçek EKG'lerinden oluşmaktadır. Bu 689 hastadan 145'i SVB, 95'i AB, 211'i MİS, 118'i ME ve 120'si de normal sağlıklı bireylerinden oluşmaktadır. Veri kümesi; hastalar için 689 satırdan, koşul niteliği için 14 ve karar niteliği için 1 olmak üzere toplam 15 sütundan oluşan 689x15 boyutunda bir matristir. Bu matris doktorlar tarafından her hastanın önceden çekilmiş olan EKG kâğıdının yorumlanmasıyla doldurmuştur. Hastaların gizlilik hakkının saklı tutulması ilkesinin ihlali yaşanmaması için hastaların bilgileri verilmemiştir. Her hasta veri matrisinde bulunduğu satır, hasta numarası (i) olacak şekilde "H_i" gibi bir ifade ile gösterilmiştir. Örneğin, hasta numarası 5 olan (i = 5 için) hasta için tablodaki gösterimi H₅ olur. Veri kümesinden bir kesit Çizelge 5.6'da verilmiştir.

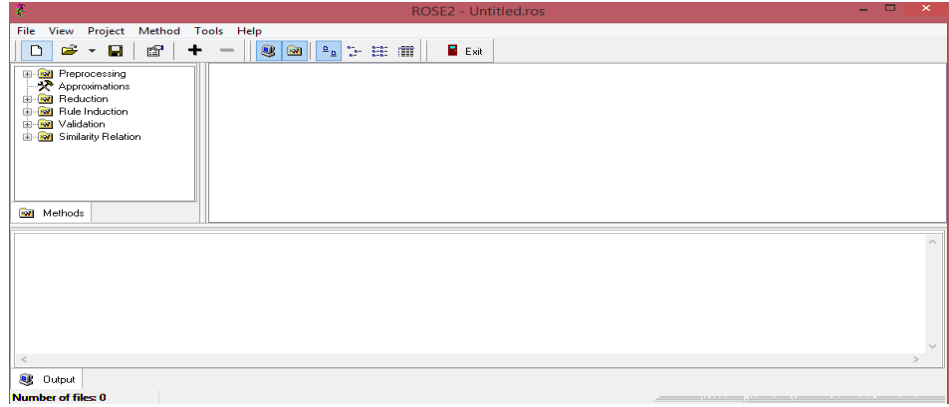
Çizelge 5.6. Veri kümesinden bir kesit

Hastalar	Rate	ST-segment	T-dalgası	Patalojik-Q	R-pregresyon	PR-aralık	P-yükseklik	P-genişlik	QRS-genişlik	QRS-yükseklik	QTc	T-regresyon	QT-aralık	Cinsiyet	Sınıf
H ₁	T	N	N	Y	Y	A	N	A	N	N	N	V	U	Yk	N
H ₂	N	C	N	Y	Y	N	A	N	N	N	N	V	N	Ye	MI
H ₃	N	N	N	Y	Y	N	N	N	N	N	N	Y	N	Ye	AB
H ₄	N	N	Ne	Y	V	N	N	N	N	N	N	V	N	Ye	SVB
H ₅	N	C	N	V	Y	N	N	N	N	N	N	Y	N	Ye	MI
H ₆	N	N	N	Y	V	A	N	N	N	Y	N	V	N	Ye	SVB
H ₇	B	C	N	Y	Y	N	N	N	N	N	N	Y	K	Ye	MIS
H ₈	N	C	P	Y	Y	N	N	N	N	N	N	Y	N	Ye	MIS
H ₉	N	N	P	V	V	N	N	N	N	N	N	V	N	Ye	MI
H ₁₀	N	N	P	V	V	N	N	N	N	N	N	Y	N	Yk	MI
H ₁₁	N	N	N	Y	Y	N	A	N	N	N	N	V	N	Ye	AB
H ₁₂	N	N	N	Y	Y	N	N	N	N	N	N	V	N	Yk	N
H ₁₃	N	N	Ne	Y	V	N	N	N	N	N	A	Y	N	Ye	SVB
H ₁₄	N	N	N	Y	Y	N	A	N	U	N	N	Y	N	Yk	N
H ₁₅	T	Y	Ne	Y	Y	A	N	N	N	N	A	Y	N	Yk	MIS
H ₁₆	N	N	N	Y	Y	N	N	N	N	N	N	V	N	Ye	N
H ₁₇	N	N	N	Y	Y	N	N	A	U	N	N	Y	N	Yk	AB
H ₁₈	N	N	N	Y	Y	N	N	N	N	N	N	Y	N	Ye	N
H ₁₉	N	C	N	V	Y	N	N	N	N	N	N	Y	N	Ye	MI
H ₂₀	N	N	N	Y	Y	A	N	N	U	Y	N	V	N	Ye	SVB
H ₂₁	N	N	N	Y	Y	N	A	N	N	N	N	Y	K	Ye	AB
H ₂₁	B	C	P	V	V	N	N	N	N	N	N	Y	U	Ye	MI
H ₂₂	T	Y	N	Y	Y	A	N	N	K	N	A	Y	N	Ye	MIS
H ₂₃	B	N	N	Y	Y	N	N	N	N	N	N	Y	N	Ye	N
H ₂₄	N	Y	P	Y	Y	N	N	N	N	N	A	V	N	Ye	MI
H ₂₅	N	Y	Ne	Y	Y	N	N	N	N	N	N	V	N	Ye	SVB
H ₂₆	N	C	N	Y	Y	N	N	N	N	N	N	Y	N	Ye	MIS
H ₂₇	N	N	N	Y	Y	N	N	N	N	N	N	Y	N	Ye	N
H ₂₈	N	Y	P	V	V	N	N	N	K	N	N	V	N	Yk	MI
H ₂₉	N	N	N	Y	Y	N	N	N	N	N	N	Y	N	Ye	N
H ₃₀	T	C	Ne	Y	V	N	N	N	N	Y	N	Y	N	Ye	SVB
H ₃₁	N	Y	N	V	Y	N	N	N	N	N	N	V	N	Ye	MI
H ₃₂	N	N	N	Y	Y	N	N	N	N	N	N	Y	N	Ye	N
H ₃₃	T	N	N	Y	Y	A	N	N	U	N	N	Y	K	Ye	N
H ₃₄	N	Y	Ne	Y	Y	N	N	N	N	N	N	V	N	Ye	SVB
H ₃₅	N	N	N	Y	Y	N	N	N	N	N	N	Y	N	Ye	N
H ₃₆	N	Y	Ne	Y	V	A	N	N	N	Y	A	V	N	Ye	SVB
H ₃₇	T	N	N	Y	Y	N	N	N	N	Y	N	Y	N	Ye	N
H ₃₈	N	N	N	Y	Y	N	A	A	N	N	N	Y	N	Yk	AB
H ₃₉	N	N	N	Y	Y	N	N	N	N	N	N	Y	N	Ye	N
H ₄₀	N	C	P	Y	Y	N	N	N	N	N	A	Y	N	Ye	MIS

5.3.2.ROSE yardımı ile veri kümesinde gereksiz nitelik tespiti

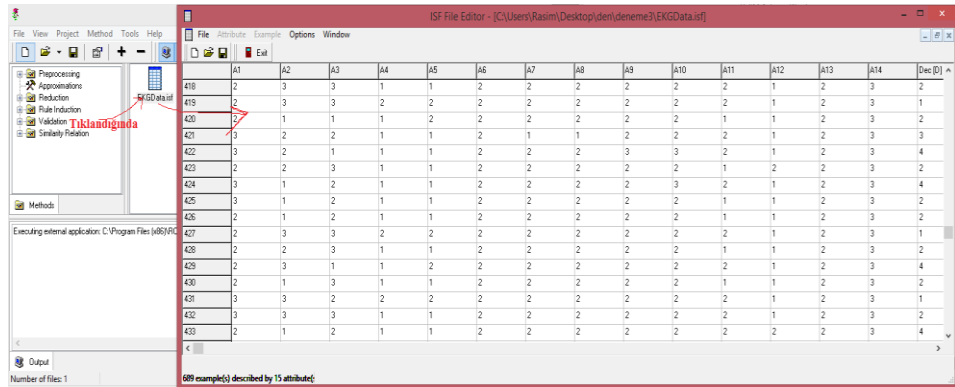
ROSE, KKT'nin temel işlemlerini gerçekleştiren ve kural çıkarımını yapan bir yazılımdır. ROSE yazılımında tüm hesaplamalar Pawlak tarafında ortaya çıkartılan temel KKT'ye dayanmaktadır. Yazılım RoughDAS ve RoughClass sistemlerinin geliştirilmiş versiyonudur. ROSE programı 32-bit GUI işletim sistemi ihtiyacı duyar [35]. Bu çalışmada ROSE programı, çekirdek ve indirgenen niteliklerin tespiti için kullanılmıştır. Bunun için yapılması gerekli adımlar aşağıda verilmiştir.

- 1- ROSE programında yeni bir proje açmak için: File → New Project işlemleri yapılır ve aşağıdaki pencere açılır.



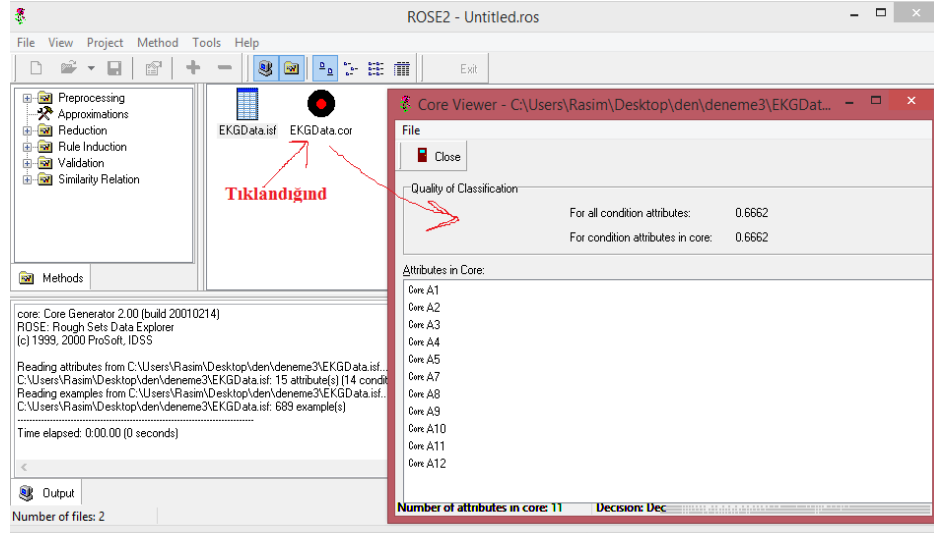
Şekil 5.3. ROSE programında yeni bir proje açmak

- 2- Açılan pencereden Project → Add File(s) to Project 'den projeye “.isf” uzantılı veri kümesi eklenir.



Şekil 5.4. ROSE programında dosya eklemek

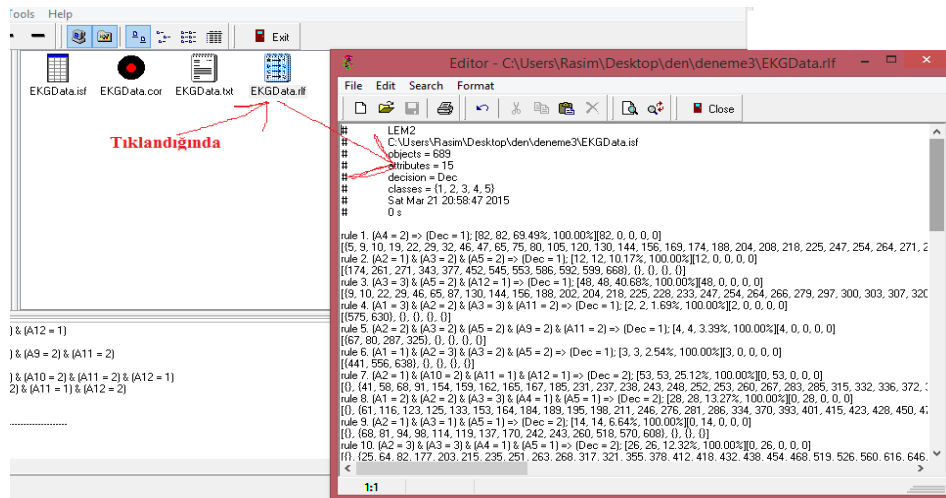
- 3- Çekirdek ve indirgenen nitelikleri tespit etmek için: Veri kümesi simgesine sağ tıklanır ve Reduction → Core seçilir.



Şekil 5.5. ROSE programında nitelik indirgeme

Core viewer penceresinde çekirdek veya indirgenemeyen nitelikler görülmektedir. Bu pencerede olmayan nitelikler gereksiz nitelik olarak tanımlanır. Görüldüğü gibi core viewer penceresinde A_6 , A_{13} ve A_{14} nitelikleri yok. Dolayısıyla bu nitelikler gereksiz nitelikleridir. Bu niteliklerden $A_6 = PR\text{-}Aralık$, $A_{13} = QT\text{-}Aralık$ ve $A_{14} = Cinsiyet$ niteliklerini göstermekteydi. Bunun anlamı, bu üç niteliğin atılması bilgi sisteminin özgünlüğünü bozmayacaktır.

- 4- Kural çıkarımını yapmak için: Veri kümesi simgesine sağ tıklanır ve Rule Induction → Basic Minimal Covering... seçilir.



Şekil 5.6. ROSE programında kural çıkarımı

Bu çalışmada ROSE’de kural çıkarımının yapılmasının amacı, elde edilen kurallar ile uzmanların görüşlerinin kıyaslanma işlemini gerçekleştirmektir. Böylece ortaya atılan modelin ya da mimarinin bir EKG sinyallini ne kadar doğru yorumladığını daha açık bir şekilde görme imkânı olur. Ek-2’de ROSE yardımı ile elde edilen tüm kurallar verilmiştir.

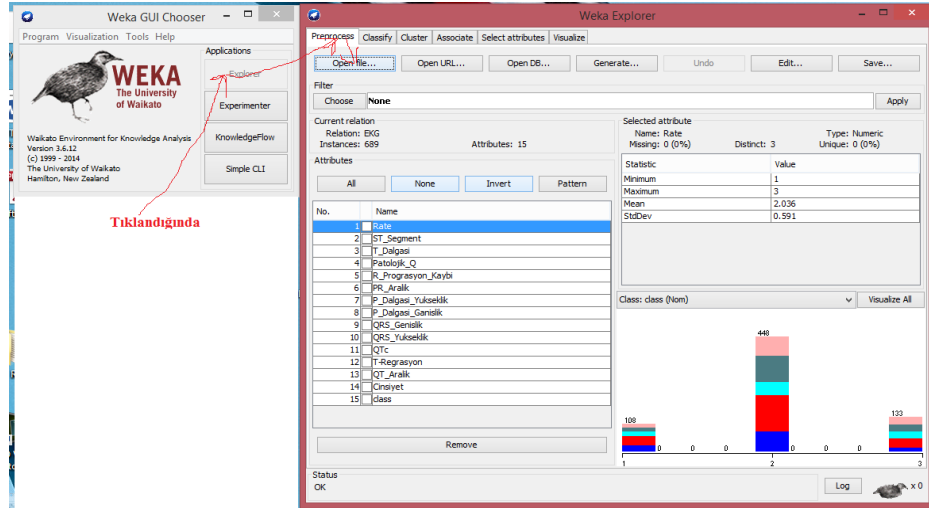
5.3.3.WEKA’da verilerin sınıflandırılması

WEKA, veri madenciliğinde işlemleri gerçekleştirmek için makine öğrenme algoritmalarının bir topluluğudur. Algoritmalar direk veri kümelerine uygulandığı gibi yazılmış olan bir JAVA kodu içerisinde de çağırılabilir. WEKA önışlem, sınıflandırma, kümeleme, regresyon vb. işlemler için araçlar içerir. Ayrıca yeni algoritmalar geliştirmek için de uygun bir ortamdır. WEKA GNU General Public License altında açık kodlu bir yazılımdır [36].

Çalışmada sınıflandırma işlemleri için WEKA’daki sınıflandırma yöntemlerinden faydalanılmıştır. Veri kümesi üzerinde en yüksek sınıflandırma işlemi yapan 9 sınıflandırma yöntemi seçilmiştir.

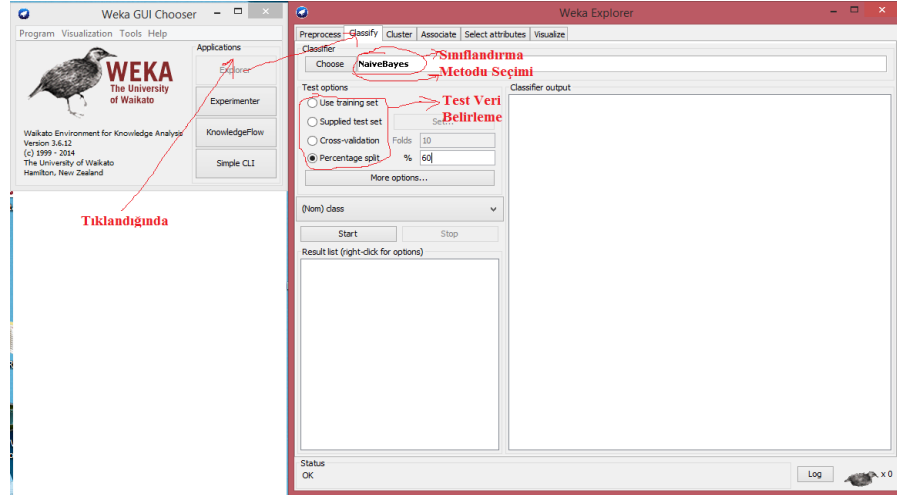
WEKA’da bir veri kümesi açmak ve sınıflandırma yapmak için:

- 1- WEKA açıldıktan sonra Explorer’den Preprocess → Open File’dan .arff veya WEKA’nın çalıştığı başka dosya uzantılı veri kümeleri buradan açılır.



Şekil 5.7. WEKA programında proje açmak

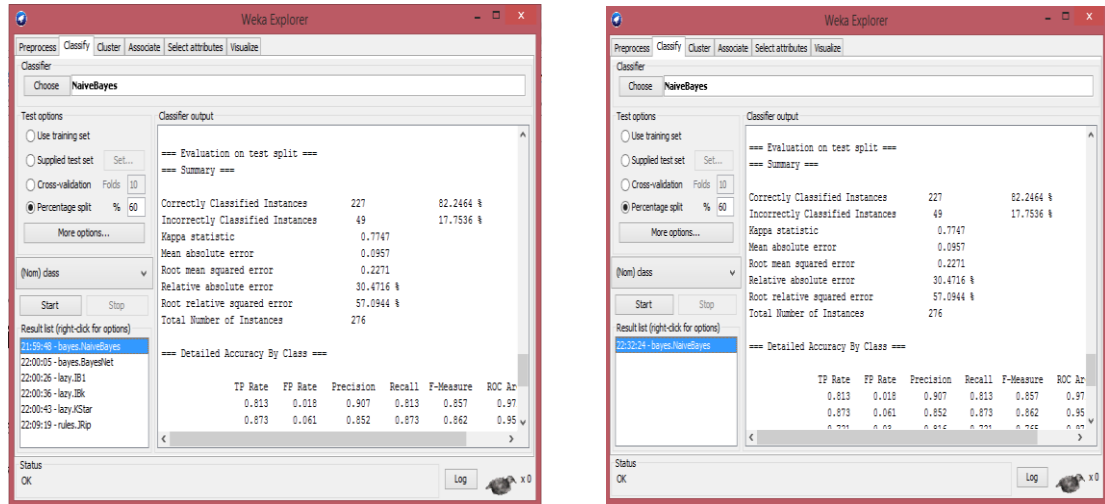
- 2- Açılan veri kümesinin sınıflandırma işlemi *Explorer* → *Classify* 'den yapılabilir. Burada açılan pencereden sınıflandırma metodu seçilebilir ve sınıflandırma yöntemi için veri kümesinin eğitim ve test verileri *Cross-Validation*, *Percentage Split* gibi seçenekler ile ayarlanabilir.



Şekil 5.8. WEKA programında sınıflandırma işlemi

- 3- Adım 2 yapıldıktan sonra sınıflandırma işlemi başlatmak için, *Classify* penceresindeki *Start*'da basmak yeterli olacaktır.

WEKA üzerinde seçilen tüm sınıflandırma metodlarının çalıştırılmasını göstermek yerine bir tane metodun çalıştırılması (*NaiveBayes* metodu için) gösterilmiştir. Çünkü Uygulama Sonuçları başlığı altında tüm sınıflandırma metodlarının performansları gösterilmiştir.



(a)

(b)

Şekil 5.9. WEKA'da (a) 15 nitelik (b) 12 nitelik için sonuçları göstermektedir

Şekil 5.16 incelendiğinde WEKA’da *NaiveBayes* sınıflandırıcısı hem 15 nitelik hem de 12 nitelik için tamamen aynı sonuçları vermiştir. Dolayısıyla tespit edilen $A_6 = Pr\text{-Aralık}$, $A_{13} = QT\text{-Aralık}$ ve $A_{14} = Cinsiyet$ niteliklerin sınıflandırıcının doğruluğunda rol oynamadıkları açıktır. Bu niteliklerin elenmesinde herhangi bir hata olamayacağı ve bu nitelikler anlam ifade etmediği söylenebilir.

5.3.4.Uygulama sonuçları

Bu çalışmada ESOGÜ Kardiyoloji Anabilim dalı doktorlarının katkılarıyla elde edilen hastaların verileri kullanılmıştır. Bu bilgi sistemi üzerinde ilk olarak önerilen model olmadan WEKA ile sınıflandırma işlemleri yapılmıştır. Daha sonra model ile ROSE ve WEKA programı yardımı ile sınıflandırma işlemleri gerçekleştirilmiştir. Her iki işlemin sonuçları çizelgeler üzerinde gösterilmiş ve bu çizelgelerin analizleri yapılmıştır. Aşağıda Çizelge 5.7’de model olmadan sınıflandırma sonuçları gösterilmiştir.

Çizelge 5.7. 15 nitelik için sınıflandırma sonuçları

Algoritma	Nitelik Sayısı	Eğitim Kümesi (%)	Doğruluk (Accuracy) (%)	Zaman (Runing-Time) (sn)	Mutlak Ortalama Hata (MAE)	F-ölçümü (f-score)
NaiveBayes	15	60	82.246	0.080	0.096	0.823
BayesNET	15	60	84.058	0.110	0.090	0.840
IB1	15	60	76.087	0.020	0.096	0.762
IBK	15	60	83.333	0.010	0.081	0.832
KStar	15	60	83.333	0.010	0.115	0.832
LWL	15	60	60.145	0.010	0.199	0.558
JRip	15	60	81.159	0.300	0.102	0.811
NBTree	15	60	84.058	1.500	0.093	0.841
VFI	15	60	67.391	0.010	0.315	0.683

ROSE programında veri kümesine KKT uygulandığında veri kümesindeki 6. , 13. ve 14. koşul niteliklerinin gereksiz oldukları saptanmıştı (Şekil 5.12). Bu 3 koşul nitelikleri sırasıyla; PR Aralık, QT Aralık ve Cinsiyet niteliklerini göstermekteydi. Sınıf belirlemede işleminde etkisiz olduklarından bu niteliklerin göz ardı

edilebileceği daha önceden açıklanmıştı. Bu yargıya varılmasının arka planında hem uzmanların görüşü hem de literatür çalışmasının olduğunu ifade etmekte yarar vardır. Aşağıda Çizelge 5.8’de sınıflandırma yöntemlerinin geriye kalan 12 nitelik üzerindeki sonuçları gösterilmiştir. Başka bir ifade ile önerilen model uygulandığında sınıflandırıcıların davranışları Çizelge 5.8’de verilmiştir.

Çizelge 5.8. 12 nitelik için sınıflandırma sonuçları

Algoritma	Nitelik Sayısı	Eğitim Kümesi (%)	Doğruluk (Accuracy) (%)	Zaman (Runing-Time) (sn)	Mutlak Ortalama Hata (MAE)	F-ölçümü (f-score)
NaiveBayes	12	60	82.246	0.012	0.096	0.823
BayesNET	12	60	84.058	0.020	0.090	0.840
IB1	12	60	76.087	0.010	0.096	0.762
IBK	12	60	83.333	0.001	0.081	0.832
KStar	12	60	83.333	0.001	0.115	0.832
LWL	12	60	60.145	0.001	0.199	0.558
JRip	12	60	83.333	0.110	0.111	0.832
NBTree	12	60	84.058	1.500	0.093	0.841
VFI	12	60	67.391	0.010	0.315	0.683

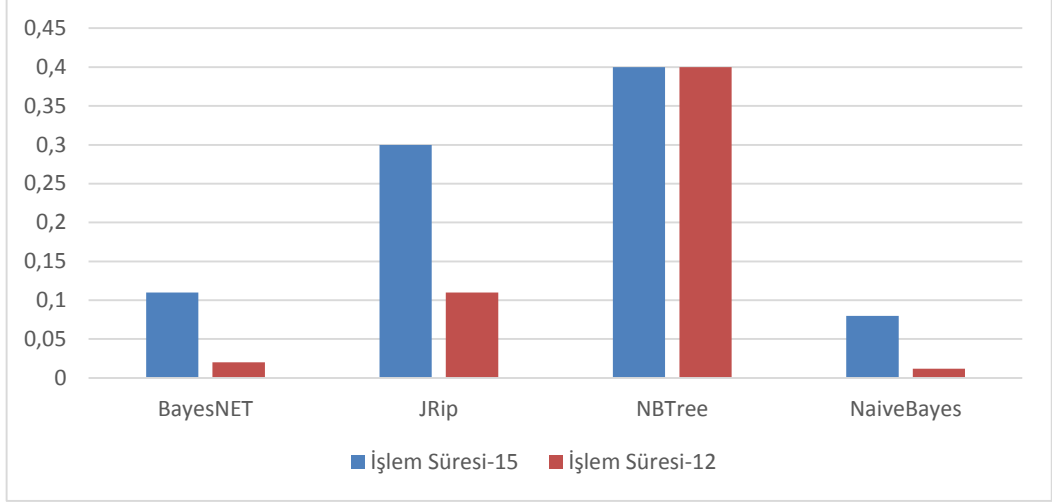
Çizelge 5.7 ve Çizelge 5.8 incelendiğinde JRip algoritması dışında diğer tüm sınıflandırma algoritmalarının doğruluklarında bir değişim söz konusu olmazken, JRip algoritması dâhil tüm algoritmaların performanslarında (yürütme zamanı bakımından) bir iyileşmenin olduğunu görülmek mümkündür. JRip algoritmasının doğruluğundaki iyileşmenin olmasının nedeni JRip [37], J48 gibi algoritmaların kural tabanlı çalışan algoritmalar olmasıdır. Kural tabanlı çalışan bu algoritmaların nitelik sayısında azalma olduğunda kural üretme sayısında da azalma olacağından sınıflandırma işleminde işlem indirgemesi yaşanır. Bu işlem sayesinde nesne ile nitelik arasındaki bağıntı karmaşıklığı azaltılabilir. Önerilen model bunu sağladığında JRip algoritmasındaki iyileşmenin bundan kaynaklandığı düşünülmektedir.

Aşağıda Çizelge 5.9’da model ve model olmadan, sınıflandırma metotlarının doğruluk ve çalıştırma zamanlarındaki değişimleri gösterilmiştir.

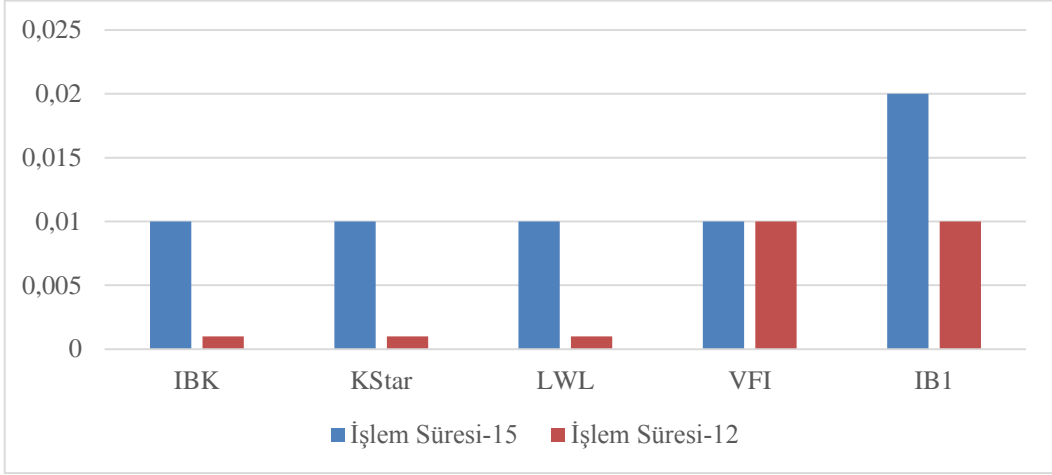
Çizelge 5.9. Nitelik sayısına bağlı değişmeler

Algoritma	15 Nitelik İçin Doğruluk (%)	12 Nitelik İçin Doğruluk (%)	15 Nitelik İçin Zaman (sn)	12 Nitelik İçin Zaman (sn)	Eğitim Kümesi (%)
NaiveBayes	82.246	82.246	0.080	0.012	60
BayesNET	84.058	84.058	0.110	0.020	60
IB1	76.087	76.087	0.020	0.010	60
IBK	83.333	83.333	0.010	0.001	60
KStar	83.333	83.333	0.010	0.001	60
LWL	60.145	60.145	0.010	0.001	60
JRip	81.159	83.333	0.300	0.110	60
NBTree	84.058	84.058	1.500	1.500	60
VFI	67.391	67.391	0.001	0.001	60

Çizelge 5.9 incelendiğinde hemen hemen tüm algoritmaların işlem sürecinde kayda değer bir iyileşmenin sağlandığı görülebiliyor. Örneğin, NaiveBayes sınıflandırma metodu, önerilen model olamadan işlem süreci ya da yürütme zamanı 0.80 sn iken, önerilen model uygulandığında NaiveBayes sınıflandırma metodunun yürütme zamanı 0.20 sn'ye kadar düşmektedir. Bu da NaiveBayes sınıflandırma metodu için yürütme zamanı bakımından %75'lik bir iyileşmeye denk gelmektedir. Büyük veri kümelerinde sınıflandırıcıların işlem yükü düşünüldüğünde sınıflandırıcılar için yürütme zamanında bu denli bir iyileşme sağlanması büyük bir başarıyı olduğu söylenebilir. Bu başarının arkasında KKT'nin etkili bir şekilde çalışması yatmaktadır. Çünkü KKT nesne ve nitelik arasındaki karmaşıklığı, belirsizliği belli bir ölçüde azaltılmaktadır. Bu da sınıflandırma metotlarının doğrulukları sabit iken, yürütme zamanında önemli bir iyileşmeyi beraberinde getirmiştir. İyileşme sonuçları aşağıda Şekil 5.10 üzerinde gösterilmiştir. Şekil 5.10 (a) ve (b) olarak iki parçaya bölünmüştür. Amaç, tek grafik üzerinde yürütme zaman değeri çok küçük olan algoritmaların bu değerlerini görmek güç olduğunda Şekil 5.10 iki ayrı parçaya bölünerek bu değerlerin açık olarak gözlemlenmesini sağlamaktır. Şekil 5.10 üzerinde model olmadan yürütme zamanını göstermek için “İşlem Süresi-15”, model olduğunda yürütme zamanı için “İşlem Süresi-12” ifadeleri kullanılmıştır.



(a)



(b)

Şekil 5.10. Zaman bazında iyileşme sonuçları

Bu bölümün diğer ayağı ve önemli konusu, herhangi bir insana ait EKG sinyal kayıtlarının % 85'e yakın bir doğruluk oranı ile doğru sınıflandırılmasıdır. Sınıflandırmadaki bu % 85'lik oran sağlık alanındaki eksik ve tutarsız veri çokluğu ve farklı hastalıkların ortak belirti gösterme çokluğu düşünüldüğünde önemli bir başarı sağladığı söylenebilir. Aşağıda ROSE programı ortamında kaba kümeler yardımı ile elde edilmiş bazı kural çıkarımları verilmiştir. Bu kurallar, koşul nitelikleri ile karar niteliği arasındaki bağlantıya bağlı olarak nesnenin sınıfını belirten kurallardır. Denklem 5.10 kullanılarak gösterilen kurallar temelini ROSE programında üretilen kurallardan almaktadır. Aşağıdaki kurallar için ROSE kurallarından toplamda 15 tanesi rasgele seçilmiştir.

$$H(r) = \begin{cases} \text{ME} & , r = 1 \\ \text{MİS} & , r = 2 \\ \text{AB} & , r = 3 \\ \text{SVB} & , r = 4 \\ \text{N} & , r = 5 \end{cases} \quad (5.10)$$

Kurallar:

K1.	$a_{2_1}a_{3_2}a_{5_2}$	$\Rightarrow r = 1,$	$H(1) = \text{ME}$
K2.	$a_{3_3}a_{5_2}a_{11_1}$	$\Rightarrow r = 1,$	$H(1) = \text{ME}$
K3.	$a_{1_2}a_{2_2}a_{3_3}a_{4_1}a_{5_1}$	$\Rightarrow r = 2,$	$H(2) = \text{MİS}$
K4.	$a_{1_2}a_{2_1}a_{5_1}a_{8_3}$	$\Rightarrow r = 2,$	$H(2) = \text{MİS}$
K5.	$a_{1_2}a_{2_3}a_{4_1}a_{8_1}$	$\Rightarrow r = 2,$	$H(2) = \text{MİS}$
K6.	$a_{1_3}a_{6_1}a_{7_1}a_{11_2}$	$\Rightarrow r = 3,$	$H(3) = \text{AB}$
K7.	$a_{1_1}a_{3_2}a_{6_2}a_{7_1}$	$\Rightarrow r = 3,$	$H(3) = \text{AB}$
K8.	$a_{3_1}a_{5_2}a_{10_2}$	$\Rightarrow r = 4,$	$H(4) = \text{SVB}$
K9.	$a_{3_1}a_{7_2}a_{10_2}$	$\Rightarrow r = 4,$	$H(4) = \text{SVB}$
K10.	$a_{1_3}a_{2_2}a_{6_2}a_{9_2}a_{10_2}$	$\Rightarrow r = 5,$	$H(5) = \text{N}$
K11.	$a_{2_2}a_{3_2}a_{7_2}a_{9_1}$	$\Rightarrow r = 5,$	$H(5) = \text{N}$
K12.	$a_{1_1}a_{2_2}a_{3_3}a_{7_2}a_{11_1}$	$\Rightarrow r = 1,$	$H(1) = \text{ME}$
K13.	$a_{2_1}a_{10_2}a_{11_1}a_{12_1}$	$\Rightarrow r = 2,$	$H(2) = \text{MİS}$
K14.	$a_{1_2}a_{5_1}a_{6_2}a_{7_2}a_{8_3}a_{10_2}$	$\Rightarrow r = 4,$	$H(4) = \text{SVB}$
K15.	$a_{3_1}a_{9_3}$	$\Rightarrow r = 4,$	$H(4) = \text{SVB}$

Yukarıdaki kurallar çoğaltılabilir. Burada a_{ij} gibi ifadelerin anlamı; örneğin a_{3_3} için 3. koşul niteliği olan *T dalgasının* tanım kümesi değerinin *Pozitif* olduğu anlaşılır. Yine a_{1_2} için 1. koşul niteliği *Rate* olan tanım kümesi değerinin *Normal* olduğu anlaşılır.

Daha açık bir ifade ile bu kurallar aşağıdaki gibi de tanımlanabilir:

K1. = {T-dalgası = Pozitif & R-dalga progresyon kaybı = Var & T-reciprocity = Yok \rightarrow Karar = ME}

K3. = {ST-segment = Çökme & T- dalgası = Negatif & R-dalga progresyon kaybı
= Yok → Karar = MİS }

K8. = {T-dalgası = Negatif & R-dalga progresyon kaybı = Var & QT_c = Var →
Karar = SVB}

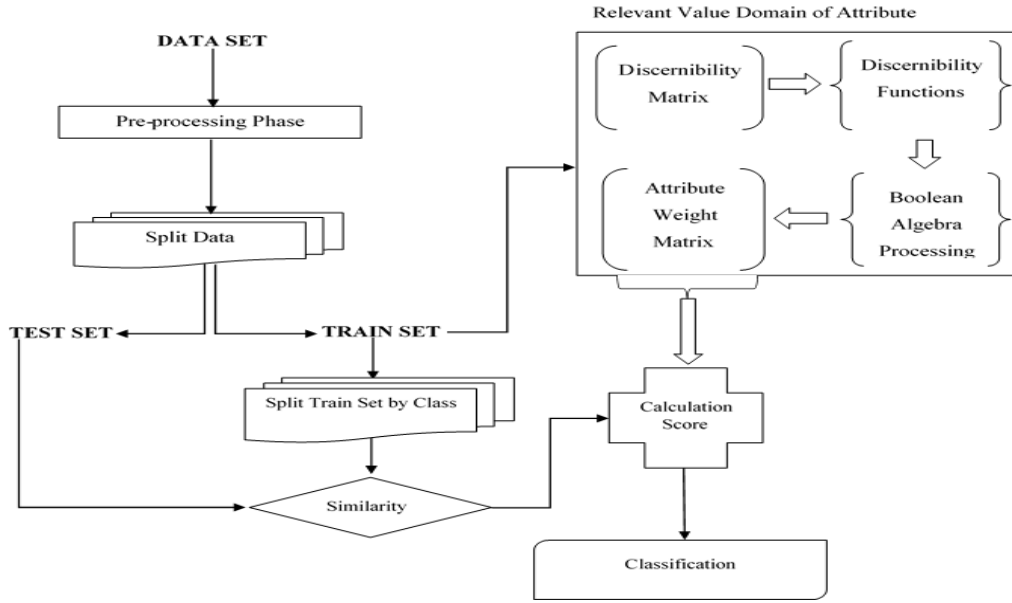
Uzman hekimlerle yapılan görüşmede yukarıdaki kurallardan ilk 11'inin % 100'lük, geriye kalan kuralların da % 80'lik bir doğruluk oranıyla sınıflandırıldığı görülmüştür. Bunun anlamı, oluşturulan modelin EKG yorumlamadaki becerisinin başarılı olduğu ve doktorların daha hızlı ve erken teşhisini kolaylaştıracak ve doktorların hata yapma oranını düşürecek bir yardımcının olmasıdır.

6. YENİ SINIFLANDIRMA YÖNTEMİ

6.1. Nitelik Ağırlıklı Kaba Küme Sınıflandırma

Sınıflandırma işlemi, bir nesneyi niteliklerini kullanarak daha önceden belli olan bir sınıfa koyma işlemidir. Başka bir ifade ile bir sınıflandırma tekniği, giriş veri kümesinden sınıflandırma modelleri inşa eden sistematik bir yaklaşımdır. Her bir teknik, giriş verisinin sınıf etiketi ve nitelik arasındaki ilişkiyi en iyi belirtmeye çalışan bir öğrenme algoritmasına dayanır [38]. Tanımdan da anlaşılacağı üzere sınıflandırma işlemi nitelik ile sınıf arasındaki ilişkiye ve sınıflar arasındaki ortak ve ayrı niteliklere bağlıdır. Bu ilişki ne kadar basit ve ortak nitelik sayısı ne kadar az olursa sınıf belirleme o kadar kolay olur. Ancak maalesef gerçek hayatta, uygulamaların çoğunda bunun tersi durum söz konusudur. Doğada karmaşık yapıları ve farklı sınıflarda ortak nitelik sayısı fazla olan veri kümeleri çoğunluktadır. Ortak nitelikler sınıflandırıcılar için bir problem yaratmaktadır. Bu ve benzeri problemlere çözüm bulmak, sınıflandırma işleminin hem performansını artıracak hem de işini basitleştireceği açıktır. Bu noktada bu çalışmada eğitim kümesindeki her nesnenin nitelikleri ile ait olduğu sınıf arasındaki ilişkisi açıkça ortaya konularak karmaşıklık azaltılmış ve hangi niteliklerin nesnenin sınıfını belirlemede daha etkin olduğu bulunarak bahsedilen problemlerin etkisi minimuma indirgenmiştir.

Yukarıdaki bilgiler ışığında Nitelik Ağırlıklı Kaba Küme Sınıflandırma <FWRSC> adında yeni bir sınıflandırma metodu geliştirilmiştir. Bu sınıflandırma metodu, puanlama sistemi ile oluşturulan bir ağırlık matrisi yardımıyla nitelik değerlerinin etkisinin azaltılması ya da artırılmasına dayalı olarak benzerlik tabanlı çalışmaktadır. Bir nesnenin sınıfı belirlemede etkili olan niteliklerinin değerlerinin etkileri artırılırken etkisiz olan niteliklerin ağırlıkları ise azaltılmıştır. Bu şekilde daha etkili bir ayrıştırma, sınıflandırma sağlanmıştır.



Şekil 6.1. FWRSC modeli

KKT'nin ilk aşamasını oluşturan bilgi sistemi; problem ile ilgili en geniş ve kapsamlı bilgi içeren sistemdir.

Çalışmada KKT'nin ayırt edilemez matrisi ve fonksiyonları yardımı ile her nesne için çekirdek ve indirgenen nitelikler elde edilmiştir.

Bu çalışmada, klasik kaba kümeler teorisindeki ayırt edilebilirlik matrisi oluşturma ve ayırt edilebilirlik fonksiyonlarından faydalanılmıştır. Amaç sınıflar arasındaki farklılığı ortaya çıkarmak olduğunda, aynı sınıftaki nesnelerin değil farklı sınıftaki nesnelerin niteliklerine bağlı olarak farklılıkları ortaya çıkartılmıştır. Örneğin; bir nesne için, 1. ve 3. nitelikler ayırt edilebilir nitelik olarak elde edilmişse bu o nesnenin diğer sınıflardaki nesnelere farklılığından ortaya çıkmıştır. Kendi sınıfı içindeki nesnelere ayırt edilebilirliğine bakılmamıştır.

Çizelge 6.1. Bir nesne için farklı sınıftaki nesnelere ayırt edilebilirliği

Object	a ₁	a ₂	a ₃	a _m	Class
x ₁	2	1	1	5	1
x ₂	2	3	5	2	1
x ₃	1	2	3	1	1
x ₄	1	2	2	1	2
.
.
.
x _{n-1}	0	5	4	2	3
x _n	2	2	1	4	1

6.2. Ağırlık Matrisi Hesaplanması

Elde edilen ayırt edilebilirlik fonksiyonlar yardımı ile çekirdek nitelikler elde edilebiliyordu ve hangi niteliklerin bir nesne için önemli olduğu açıkça görülebiliyordu. Bu bilgiye dayalı olarak bir nesne için her niteliğe puanlama sistemi uygulanabilir. Örneğin 5 nitelikli bir bilgi sisteminde bir nesnenin 1. ve 3. niteliği çekirdek nitelikler ise bu iki niteliğin etkisi artırılabilir. Bu şekilde nesneyi sınıflandıran niteliklerin etkisi artırılırken diğer niteliklerin etkisi azaltılmış olur. Buna bağlı olarak da daha doğru bir sınıflandırma işlemi gerçekleştirilebilir. Çalışmada niteliklere puanlama sistemi için oluşturulan matris, ağırlık matrisi ($W_{i,j}$) olarak isimlendirilmiştir. Bu matrisin hesaplanmasında her nesnenin ayırt edilebilirlik fonksiyonları kullanılmıştır. Her nesne için her niteliğin fonksiyondaki frekansı, toplam ‘veya (V)’ sayısının bir fazlasına bölünmüştür. Eğer f_i , i. nesne için ayırt edilebilirlik fonksiyonu ve $a_j \in A$ ve $a_j \subset f_i$ nitelikleri, $Count(f_i, a_j)$ f_i ayırt edilebilirlik fonksiyonundaki a_j niteliğinin frekansını ve ‘V’ boolean cebirdeki “VEYA” işlemini temsil etmek üzere ve $V \in f_i$ iken, $Count(f_i, V)$ f_i ayırt edilebilirlik fonksiyonundaki ‘V’ işlemin frekansını belirtir:

$$N = Count(f_i, V) + 1 \quad (6.1)$$

$$d_{i,j} = \frac{Count(f_i, a_j)}{N} \quad (6.2)$$

olur.

Ağırlık Matrisi ($W_{i,j}$):

$$W_{i,j} = \begin{bmatrix} d_{1,1} & d_{1,2} & d_{1,3} & \dots & d_{1,j} & \dots & d_{1,n} \\ d_{2,1} & d_{2,2} & d_{2,3} & \dots & d_{2,j} & \dots & d_{2,n} \\ d_{3,1} & d_{3,2} & d_{3,3} & \dots & d_{3,j} & \dots & d_{3,n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ d_{i,1} & d_{i,2} & d_{i,3} & \dots & d_{i,j} & \dots & d_{i,n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ d_{m,1} & d_{m,2} & d_{m,3} & \dots & d_{i,j} & \dots & d_{m,n} \end{bmatrix}$$

Örneğin, 10 nesnesi ve 4 niteliği olan bir bilgi sisteminde, 5. nesne için, $f_5 = a_1 \wedge a_2 \vee a_1 \wedge a_3 \wedge a_4$ ayırt edilebilirlik fonksiyonu elde edilsin. Buna göre, 5. nesne için denklem 6.1 ve 6.2’den her niteliğin ağırlığı aşağıdaki gibi olur.

$$\left. \begin{array}{l} Count(f_5, a_1) = 2 \\ Count(f_5, a_2) = 1 \\ Count(f_5, a_3) = 1 \\ Count(f_5, a_4) = 1 \\ N = Count(f_i, V) + 1 = 2 \end{array} \right\} \begin{array}{l} d_{5,1} = 2/2 \quad d_{5,2} = 1/2 \\ d_{5,3} = 1/2 \quad d_{5,4} = 1/2 \\ W_{5,j} = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 \end{bmatrix} \end{array}$$

Yukarıda a_2, f_5 içinde sadece $a_1 \wedge a_2$ içinde geçmekte ve dolayısıyla $Count(f_5, a_2) = 1$ olmaktadır. Ancak a_1 , hem $a_1 \wedge a_2$ hem de $a_1 \wedge a_3 \wedge a_4$ geçtiğinde $Count(f_5, a_1) = 2$ olmaktadır. $W_{5,j}$ vektörü ise 5. nesne için her niteliğin ağırlık değerini göstermektedir. $W_{5,j}$ vektörü bize 5. nesne için a_1 niteliğin diğer niteliklerden daha önemli olduğunu göstermekte ve aynı zamanda ayırt ediciliğini artırmaktadır.

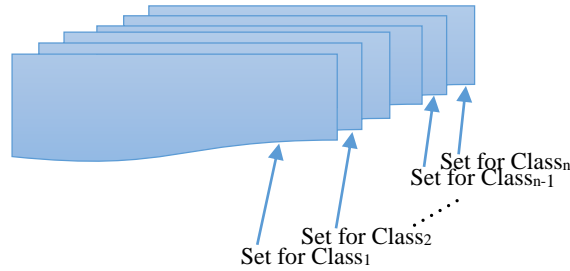
6.3. Sınıf Belirleme

Nitelikler ile sınıflar arasındaki ilişki ortaya çıkarıldıktan ve önem arz eden niteliklere bir puanlama sistemi ile önemi artırıldıktan sonra sınıflandırma işlemine geçilebilir. Çalışmada sınıflandırma işleminde, ilk önce eğitim veri kümesi sınıf bazında bölünür. Her sınıf ait olduğu nesnelere içererek, veri kümesi sınıf sayısı kadar parçalara ayrılır.

C_x , x. sınıfı belirlemekte olsun:

$$C_x \subset D, \quad x, y = 1, 2 \dots S(D)$$

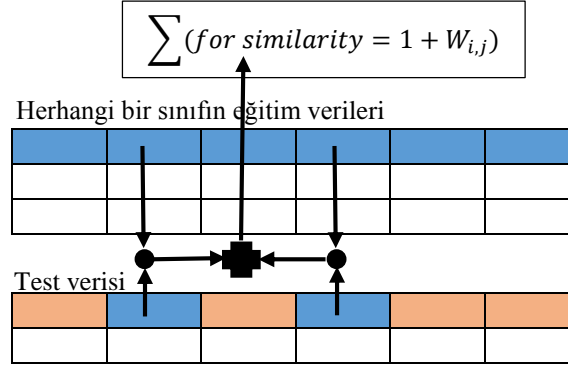
$$C_x = \{x_k \in U \mid x_k; d_x, d_y \in D \text{ ve } d_x = d_y\} \quad (6.3)$$



Şekil 6.2. Sınıflara göre veri kümesinin bölünmesi

Daha sonra test kümesindeki her veri için nitelikleri, her bir sınıf kümesindeki verilerin nitelikleri ile kıyaslanarak benzerliğine bakılır. Niteliklerinin değerlerinin aynı olduğu yerde benzerlik değeri olarak 1 puan ve ağırlık matrisindeki değeri ile bu puan toplanarak tüm sınıf kümesi için bir değer (skor) elde edilir. Bu işlem tüm sınıf kümeleri için yapılır. Son olarak sınıfların skorlarına bakılır; nesne, skoru en büyük olan sınıfa dâhil edilir.

FWRSC, kullanıcı (user) bazlı çalışan bir sınıflandırıcı olarak tasarlanmıştır. Literatürde var olan sınıflandırıcılardan *Naive Bayes* [39, 40] sınıflandırıcı gibi bazıları kullanıcı, bazıları da parça (item) bazlı çalışırlar.



Şekil 6.3. Sınıf hesaplamaları için bir gösterim

Şekil 6.3 test verisinin sınıf eğitim verisinde sadece bir nesne için yapılan işlemi göstermektedir. Gösterimdeki “for similarity = 1”nin anlamı, test veri ile eğitim verisindeki değerler aynı olduğunda benzerlikten dolayı 1 puan verilmiştir demektir. Bu benzeşen niteliğin sınıf belirleme etkisini azaltmak veya artırmak için Ağırlık Matrisindeki değeri bu puana eklenir. Son olarak da bir sınıf kümesindeki nesne sayısı tüm nesne sayısına oranı hesaplanan bu puanla toplanır. Eğer nitelik sınıf belirlemede etkin ise Ağırlık Matrisindeki değeri büyük olacağından sınıflandırmaya etkisi daha fazla olur. Ters durumda ise matristeki değeri küçük olacağından sınıflandırmaya etkisi azaltılmış olur. Bu şekilde daha sağlıklı bir sınıflandırma yapılabilir. Bu işlem sınıf eğitim verisinin tümü için yapılır ve tüm işlemler sonunda bir skor elde edilir. Bu işlem tüm sınıflar için yapıldıktan sonra elde edilen skorlar kıyaslanır, nesne maksimum skora sahip sınıfa eklenir. Buna göre her sınıfın skorunu hesaplama için:

β = test verisi kümesi, γ = eğitim verisi kümesi olmak üzere

$$\hat{R} = S(U) \text{ and } \hat{r} = S(A) - 1$$

$$r = \{d_x, d_y \in D \mid \forall d_x \neq \forall d_y, 0 < x \leq y \leq S(D)\} \quad (6.4)$$

$$Class_{score}^{t=1 \text{ to } S(r)} = \frac{\sum_{i=1}^{\hat{R}} \sum_{\substack{j=1 \\ \beta(a_j) = \gamma_i(a_j)}}^{\hat{r}} (1 + W_{i,j})}{S(C_t)} + \frac{S(C_t)}{S(U)} \quad (6.5)$$

Test nesnesi için tüm sınıfların skoru yukardaki şekilde hesaplanır. Bir sonraki adım sınıf belirleme işlemidir. Bunun için (6.6) kullanılır:

$$\text{Calculated class} = \max \left(\text{Class}_{\text{score}}^{t=1 \text{ to } S(r)} \right) \quad (6.6)$$

Örnek 6.1’de FWRSC ile bir nesnenin sınıf belirleme işlemi yapılmıştır.

Örnek 6.1.

IS = (U, A, D) bilgi sistemi Çizelge 6.2’de verilmiştir.

Çizelge 6.2. Bilgi sistemi

U	a ₁	a ₂	a ₃	a ₄	d ₁
x ₁	1	1	1	1	1
x ₂	1	1	1	2	1
x ₃	2	1	1	1	2
x ₄	3	2	1	1	2
x ₅	3	3	2	1	2
x ₆	3	3	2	2	1
x ₇	2	3	2	2	2
x ₈	1	2	1	1	1
x ₉	1	3	2	1	2
x ₁₀	3	2	2	1	2
x ₁₁	1	2	2	2	2
x ₁₂	2	2	1	2	2
x ₁₃	2	1	2	1	2
x ₁₄	3	2	1	2	1

Çizelge 6.3. Ayırt edilmezlik matrisi hesaplanması

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	-	-	a ₁	a ₁ , a ₂	a ₁ , a ₂ , a ₃	-	a ₁ , a ₂ , a ₃ ,a ₄	-	a ₂ , a ₃	a ₁ , a ₂ , a ₃	a ₂ , a ₃ , a ₄	a ₁ , a ₂ , a ₃	a ₁ , a ₂ , a ₃	-
2	-	-	a ₁ , a ₄	a ₁ , a ₂ , a ₄	a ₁ , a ₂ , a ₃ ,a ₄	-	a ₁ , a ₂ , a ₃	-	a ₂ , a ₃ , a ₄	a ₁ , a ₂ , a ₃ ,a ₄	a ₂ , a ₃ , a ₄	a ₁ , a ₂ , a ₃	a ₁ , a ₂ , a ₃	-
3	a ₁	a ₁ , a ₄	-	-	-	a ₁ , a ₂ , a ₃ ,a ₄	-	a ₁ , a ₂	-	-	-	-	-	a ₁ , a ₂ , a ₄
4	a ₁ , a ₂	a ₁ , a ₂ ,a ₄	-	-	-	a ₂ , a ₃ ,a ₄	-	a ₁	-	-	-	-	-	a ₄
5	a ₁ ,a ₂ , a ₃	a ₁ ,a ₂ , a ₃ ,a ₄	-	-	-	a ₄	-	a ₁ ,a ₂ , a ₃	-	-	-	-	-	a ₂ ,a ₃ , a ₄
6	-	-	a ₁ , a ₂ , a ₃ ,a ₄	a ₂ , a ₃ , a ₄	a ₄	-	a ₁	-	a ₁ , a ₄	a ₂ , a ₄	a ₁ , a ₂	a ₁ , a ₂ , a ₃	a ₁ , a ₂ , a ₃	-
7	a ₁ , a ₂ , a ₃ ,a ₄	a ₁ , a ₂ , a ₃	-	-	-	a ₁	-	a ₁ , a ₂ , a ₃ ,a ₄	-	-	-	-	-	a ₁ , a ₂ , a ₃
8	-	-	a ₁ , a ₂	a ₁	a ₁ , a ₂ , a ₃	-	a ₁ , a ₂ , a ₃ ,a ₄	-	a ₁ , a ₂	a ₁ , a ₃	a ₃ , a ₄	a ₁ , a ₄	a ₁ , a ₂ , a ₃	-
9	a ₂ , a ₃	a ₂ , a ₃ , a ₄	-	-	-	a ₁ , a ₄	-	a ₁ , a ₂	-	-	-	-	-	a ₁ ,a ₂ , a ₃ ,a ₄

Çizelge 6.3. (Devam) Ayırt edilemezlik matrisi hesaplanması

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
10	a ₁ , a ₂ , a ₃	a ₁ , a ₂ , a ₃ , a ₄	-	-	-	a ₂ , a ₄	-	a ₁ , a ₃	-	-				a ₂ , a ₃
11	a ₂ , a ₃ , a ₄	a ₂ , a ₃	-	-	-	a ₁ , a ₂	-	a ₃ , a ₄	-	-	-			a ₁ , a ₃
12	a ₁ , a ₂ , a ₄	a ₁ , a ₂	-	-	-	a ₁ , a ₂ , a ₃	-	a ₁ , a ₄	-	-	-	-		a ₂
13	a ₁ , a ₃	a ₁ , a ₃ , a ₄	-	-	-	a ₁ , a ₂ , a ₄	-	a ₁ , a ₂ , a ₃	-	-	-	-	-	a ₁ , a ₂ , a ₃ , a ₄
14	-	-	a ₁ , a ₂ , a ₄	a ₄	a ₂ , a ₃ , a ₄	-	a ₁ , a ₂ , a ₃	-	a ₁ , a ₂ , a ₃ , a ₄	a ₂ , a ₃	a ₁ , a ₃	a ₂	a ₁ , a ₂ , a ₃ , a ₄	-

Ayırt edilemezlik fonksiyonları ($f_i(A)$) hesaplamaları;

$$f_1(A) = a_1 \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3 \vee a_4) \wedge (a_1 \vee a_3) \wedge (a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_2 \vee a_3 \vee a_4) \wedge (a_1 \vee a_2 \vee a_4) \wedge (a_1 \vee a_3) = \mathbf{a_1 \wedge a_2 \vee a_1 \wedge a_3}$$

$$f_2(A) = (a_1 \vee a_4) \wedge (a_1 \vee a_2 \vee a_4) \wedge (a_1 \vee a_2 \vee a_3 \vee a_4) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3 \vee a_4) \wedge (a_2 \vee a_3) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_3 \vee a_4) = \mathbf{a_1 \wedge a_2 \vee a_2 \wedge a_4 \vee a_1 \wedge a_3}$$

$$f_3(A) = a_1 \wedge (a_1 \vee a_4) \wedge (a_1 \vee a_2 \vee a_3 \vee a_4) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2 \vee a_4) = \mathbf{a_1}$$

$$f_4(A) = (a_1 \vee a_2) \wedge (a_1 \vee a_2 \vee a_4) \wedge (a_2 \vee a_3 \vee a_4) \wedge a_1 \wedge a_4 = \mathbf{a_1 \wedge a_4}$$

$$f_5(A) = (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3 \vee a_4) \wedge a_4 \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_2 \vee a_3 \vee a_4) = \mathbf{a_1 \wedge a_4 \vee a_2 \wedge a_4 \vee a_3 \wedge a_4}$$

$$f_6(A) = (a_1 \vee a_2 \vee a_3 \vee a_4) \wedge (a_2 \vee a_3 \vee a_4) \wedge a_4 \wedge a_1 \wedge (a_1 \vee a_4) \wedge (a_2 \vee a_4) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_4) = \mathbf{a_1 \wedge a_4}$$

$$f_7(A) = (a_1 \vee a_2 \vee a_3 \vee a_4) \wedge (a_1 \vee a_2 \vee a_3) \wedge a_1 \wedge (a_1 \vee a_2 \vee a_3 \vee a_4) \wedge (a_1 \vee a_2 \vee a_3) = \mathbf{a_1}$$

$$f_8(A) = (a_1 \vee a_2) \wedge a_1 \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3 \vee a_4) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_3) \wedge (a_3 \vee a_4) \wedge (a_1 \wedge a_4) \wedge (a_1 \vee a_2 \vee a_3) = \mathbf{a_1 \wedge a_3 \vee a_1 \wedge a_4}$$

$$f_9(A) = (a_2 \vee a_3) \wedge (a_2 \vee a_3 \vee a_4) \wedge (a_1 \vee a_4) \wedge (a_1 \wedge a_2) \wedge (a_1 \vee a_2 \vee a_3 \vee a_4) = \mathbf{a_1 \wedge a_2 \vee a_2 \wedge a_4 \vee a_1 \wedge a_3}$$

$$f_{10}(A) = (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3 \vee a_4) \wedge (a_2 \vee a_4) \wedge (a_1 \vee a_3) \wedge (a_2 \vee a_3) = \mathbf{a_1 \wedge a_2 \vee a_2 \wedge a_3 \vee a_3 \wedge a_4}$$

$$f_{11}(A) = (a_2 \vee a_3 \vee a_4) \wedge (a_2 \vee a_3) \wedge (a_1 \vee a_2) \wedge (a_3 \vee a_4) \wedge (a_1 \vee a_3) = \mathbf{a_1 \wedge a_3 \vee a_2 \wedge a_3 \vee a_1 \wedge a_2 \wedge a_4}$$

$$f_{12}(A) = (a_1 \vee a_2 \vee a_4) \wedge (a_1 \vee a_2) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_4) \wedge a_2 = a_1 \wedge a_2 \vee a_2 \wedge a_4$$

$$f_{13}(A) = (a_1 \vee a_3) \wedge (a_1 \vee a_3 \vee a_4) \wedge (a_1 \vee a_2 \vee a_4) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3 \vee a_4) \\ = a_1 \vee a_2 \wedge a_3 \vee a_3 \wedge a_4$$

$$f_{14}(A) = (a_1 \vee a_2 \vee a_4) \wedge a_4 \wedge (a_2 \vee a_3 \vee a_4) \wedge (a_1 \vee a_2 \vee a_3) \wedge (a_1 \vee a_2 \vee a_3 \vee a_4) \wedge \\ (a_2 \vee a_3) \wedge (a_1 \vee a_3) \wedge a_2 \wedge (a_1 \vee a_2 \vee a_3 \vee a_4) = a_1 \wedge a_2 \wedge a_4 \vee a_2 \wedge a_3 \wedge a_4$$

Denklem 6.1 ve 6.2'den Ağırlık Matrisi hesaplanması

$d_{1,1}$ için,

$$Count(f_1(A), V) = \{a_1 \wedge a_2 \vee a_1 \wedge a_3\} + 1 = 1+1 = 2$$

“v” karakteri $f_1(A)$ 'de sadece 1 defa geçmektedir.
Dolayısıyla Count değeri 1 olmaktadır.

$$Count(f_1(A), a_1) = \{a_1 \wedge a_2 \vee a_1 \wedge a_3\} = 1+1 = 2$$

“a₁” karakteri $f_1(A)$ 'de 2 defa geçmektedir.
Dolayısıyla Count değeri 2 olmaktadır.

Burada, $d_{1,1} = Count(f_1(A), a_1) / N = 2/2 = 1$ olur.

$d_{1,2}$ için,

$$Count(f_1(A), V) = \{a_1 \wedge a_2 \vee a_1 \wedge a_3\} + 1 = 1+1 = 2$$

“v” karakteri $f_1(A)$ 'de sadece 1 defa geçmektedir.
Dolayısıyla Count değeri 1 olmaktadır.

$$Count(f_1(A), a_2) = \{a_1 \wedge a_2 \vee a_1 \wedge a_3\} = 1$$

“a₁” karakteri $f_1(A)$ 'de 1 defa geçmektedir.
Dolayısıyla Count değeri 1 olmaktadır.

Burada, $d_{1,2} = Count(f_1(A), a_2) / N = 1/2 = 0,5$ olur. Bu şekilde tüm $d_{i,j}$ değerleri bulunur ve Ağırlık Matrisi aşağıdaki gibi olur.

Ağırlık Matrisi;

$$W_{ij} = \begin{bmatrix} 1 & 0,5 & 0,5 & 0 \\ 0,66 & 0,66 & 0,33 & 0,33 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0,33 & 0,33 & 0,33 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0,5 & 0,5 \\ 0,66 & 0,66 & 0,33 & 0,33 \\ 0,33 & 0,66 & 0,66 & 0,33 \\ 0,66 & 0,66 & 0,66 & 0,33 \\ 0,5 & 1 & 0 & 0,5 \\ 0,33 & 0,33 & 0,66 & 0,33 \\ 0,5 & 1 & 0,5 & 1 \end{bmatrix} 14 \times 4$$

Verilen bir $\beta = \{a_1 = 3, a_2 = 1, a_3 = 1, a_4 = 1\}$ nesnesinin sınıfını belirlemek için yapılan işlemler aşağıdaki gibi sıralanabilir.

- Çizelge 6.2'deki bilgi sistemi sınıflarına göre bölünmesi (split data):

$S(D) = 2$ iken ve (6.3) kullanılarak bu örnek için her sınıf için nesne kümesi saptanabilir.

$$C_1 = \{x_1, x_2, x_6, x_8, x_{14}\} \text{ ve } C_2 = \{x_3, x_4, x_5, x_7, x_9, x_{10}, x_{11}, x_{12}, x_{13}\} \text{ olur.}$$

- Her sınıfın skorunu hesaplamak:

(6.4) kullanılarak ilgili hesaplamalar aşağıdaki gibi yapılabilir.

$$\begin{aligned} \hat{R} &= S(U) = 14 \\ \hat{r} &= 5 - 1 = 4, \\ r &= \{1, 2\} \rightarrow S(r) = 2 \text{ olur.} \end{aligned}$$

(6.5) kullanmak koşuluyla:

t = 1 için,

$$Class_{score}^1 = \frac{15}{5} + \frac{5}{14} = 3,357$$

t = 2 için,

$$Class_{score}^2 = \frac{20}{9} + \frac{9}{14} = 2,865$$

- Sınıf belirleme (denklem 6.6'dan) :

Sınıf = $\max (Class_{score}^1\{3,357\}, Class_{score}^2\{2,865\}) = \text{Class 1}$ olur. Yani sınıflandırıcı bu nesneyi 1 sınıfına dahil eder.

6.4. FWRSC Algoritması

Ek-1’de FWRSC sınıflandırma yönteminin algoritması verilmiştir. Burada A bölümde FWRSC’nin ana programı, B bölümünde ana programda çağrılan Ağırlık matrisini hesaplayan metot, C bölümünde sınıfları hesaplayan metot ve D bölümünde ise ayırt edilebilirlik matris ve fonksiyon hesaplamalarını yapan metot verilmiştir.

6.5. Uygulama

6.5.1. Veri kümeleri

Uygulamada kullanılan veri kümeleri UCI Machine Learning Repository [41] web sitesinde temin edilmiştir. 5 farklı veri kümesi üzerinde WEKA [27-29] adlı programdaki sınıflandırma yöntemleri ile FWRSC sınıflandırma yöntemi karşılaştırılmıştır. Seçilen veri kümelerinin farklı alanlardan veri kümeleri olmasına dikkat edilmiştir. Mesela sağlık alanında bir veri kümesi alınmışsa diğer veri kümelerinin sağlık alanı dışında olmasına dikkat edilmiştir. Seçilen veri kümeleri Çizelge 6.4’te verilmiştir. Bu veri kümeleri ilgili sitenin veri tabanından bahsedilen kurala uygun olarak rastgele seçilmiştir.

Çizelge 6.4. Veri kümeleri

<i>Veri Kümesi</i>	<i>Nitelik Sayısı</i>	<i>Sınıf Sayısı</i>	<i>Nesne Sayısı</i>
<i>Balloon</i>	4 + Class = 5	2	56
<i>Balance Scale</i>	4 + Class = 5	3	625
<i>Haberman's Survival</i>	3 + Class = 4	2	306
<i>Abalone</i>	8 + Class = 9	29	4177
<i>Tic-Tac-Toe</i>	9 + Class = 10	2	958

6.5.2. Değerlendirme ölçütleri

Bir sınıflandırma metodunu değerlendirmenin çeşitli ölçütleri vardır. Bu çalışmada sınıflandırma metotları için *Doğruluk (accuracy)*, *f-ölçümü (f-score or f-*

measure) ve Ortalama Mutlak Hata (mean absolute error(MAE)) ölçütleri kullanılmıştır. Bu ölçütler sınıflandırma modelinin ne kadar doğru sınıflandırma yaptığını ölçer.

Karışıklık Matrisi;

Çizelge 6.5. Karışıklık Matrisi

DOĞRU SINIF	ÖNGÖRÜLEN SINIF		
		Sınıf = 1	Sınıf = -1
	Sınıf = 1	a	b
Sınıf = -1	c	d	

a: TP (true positive)
b: FN (false negative)
c: FP (false positive)
d: TN (true negative)

Bilgi çıkarımı konusunda, kesinlik (precision) ve duyarlılık (recall) kavramları üzerinden hesaplanır.

$$Duyarlılık (recall) = \frac{\text{Doğru Sınıflandırılmış Pozitif Örnek Sayısı}}{\text{Pozitif Sınıflandırılmış Örneklerin Sayısı}} = \frac{TP}{TP+FN} \quad (6.7)$$

$$Kesinlik (precision) = \frac{\text{Doğru sınıflandırılmış pozitif Örnek sayısı}}{\text{Pozitif Örneklerin sayısı}} = \frac{TP}{TP+FP} \quad (6.8)$$

F-ölçütü: Kesinlik ve duyarlılığın harmonik ortalamasını alır.

$$F\text{-ölçütü} = \frac{2 * \text{Duyarlılık} * \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}} \quad (6.9)$$

Doğruluk: Sınıflandırma işlemi sonrasında oluşan sınıfların doğruluklarını belirlemek amacı ile yapılan işlemdir.

$$\text{Doğruluk} = \frac{a+d}{a+b+c+d} = \frac{TP+TN}{TP+FN+FP+TN} \quad (6.10)$$

Ortalama Mutlak Hata (MAE): Tüm test durumları için hesaplanan değer ile gerçek değer arasındaki farkın ortalamasıdır.

x_i : gerçek sınıf serisi.

\hat{x}_i : hesaplanan sınıf serisi.

N : nesne sayısı.

$$MAE = \frac{\sum_{i=1}^N |x_i - \hat{x}_i|}{N} \quad (6.11)$$

6.5.3. Uygulama sonuçları

WEKA programındaki sınıflandırma metotlarının ve FWRSC'nin 5 veri kümesi üzerindeki doğruluk, MAE ve f -ölçümü yönünden performansları tablolar üzerinde gösterilmiştir. Çizelge 6.6'da sınıfların doğruluk, Çizelge 6.7'de MAE ve Çizelge 6.8'de ise f -ölçümü ölçütleri verilmiştir. Çizelge 6.6'te doğruluk yüzdesi en yüksek ve en düşük sınıflandırma metodunun oranı koyu renk ile belirtilmiştir. Ayrıca en yüksek doğruluk yüzdelerin yanına “↑” işareti, en düşük değerlerin yanına “↓” işareti konulmuştur. Böylece veri kümeleri için en yüksek ve en düşük sınıflandırma yapan metotlar açıkça gösterilmesi sağlanmıştır. Çizelge 6.7 ve Çizelge 6.8'de veri kümeleri için en iyi sonucu veren sınıflandırma değer oranları ise koyu renkte yazılmıştır. Bu gösterimler ile hangi sınıflandırmanın hangi veri kümeleri için en iyi doğruluk, f -ölçümü ve MAE değerlerini ürettiğini gösterilmesi açısından faydalı olmuştur.

Çizelge 6.6. Sınıflandırma metotları için doğruluk değerleri

<i>Sınıflandırıcı</i>	<i>Sınıflandırıcı Metotları</i>	<i>Balloon (%)</i>	<i>Balance Scale (%)</i>	<i>Haberman's Survival (%)</i>	<i>Abalone (%)</i>	<i>Tic-Tac-Toe (%)</i>
Bayes	<i>BayesNET</i>	40 ↓	65.71	66.27	53.36	25.86
	<i>NaiveBayes</i>	70	87.61	72.29 ↑	53.36	16.20
	<i>NaiveBayesMultinomial</i>	45	89.52	68.67	43.01	49.22
	<i>NaiveBayesSimple</i>	60	88.09	71.08	53.22	16.20
	<i>ComplementNaiveBayes</i>	65	90.95	71.08	50.53	14.33
Functions	<i>Logistic</i>	70	89.05	69.88	58.46 ↑	07.16 ↓
	<i>MultilayerPerceptron</i>	70	93.33 ↑	71.08	58.40	30.53
	<i>SMO</i>	70	89.52	66.27	56.38	13.39
Lazy	<i>IB1</i>	60	81.42	66.27	49.19	50.15
	<i>IBK</i>	60	80.47	65.06	49.19	47.97
	<i>KStar</i>	70	82.38	66.27	55.24	37.38
	<i>LWL</i>	60	68.57	66.27	55.91	51.71
Meta	<i>AdaBoostM1</i>	70	61.90	66.27	55.91	19.93
	<i>Bagging</i>	60	80.47	67.45	57.86	23.68
	<i>Decorate</i>	60	73.81	63.85	55.37	23.68
	<i>LogitBoost</i>	70	82.86	65.06	56.85	15.89
	<i>Vote</i>	40 ↓	32.85 ↓	66.27	36.02	62.30
Misc	<i>HyperPipes</i>	40 ↓	33.33	66.27	36.29	62.30
	<i>VFI</i>	70	51.90	59.04 ↓	33.40 ↓	17.44
Rules	<i>JRip</i>	60	76.66	69.88	55.51	26.48
	<i>OneR</i>	80	58.09	68.88	50.73	63.55 ↑
	<i>PART</i>	60	78.57	69.88	53.02	21.18
	<i>Ridor</i>	60	81.42	66.26	55.10	42.36
	<i>ZeroR</i>	40 ↓	32.85 ↓	66.27	36.02	62.30
	<i>J48</i>	60	69.04	65.06	52.88	23.36
Trees	<i>CimpleCart</i>	60	66.66	68.67	57.39	26.16
	<i>FT</i>	70	87.62	66.27	56.38	14.01
	<i>NBTree</i>	80	71.42	66.27	53.36	42.05
	<i>FWRSC</i>	90 ↑	65.24	66.27	54.91	62.30
Önerilen Yöntem	<i>Genel Ortalama %</i>	62.41	72.80	67.18	51.49	33.41

Çizelge 6.7. Sınıflandırma metotları için MAE değerleri

<i>Sınıflandırıcı</i>	<i>Sınıflandırma Metotları</i>	<i>Balloon</i>	<i>Balance Scale</i>	<i>Haberman's Survival</i>	<i>Abalone</i>	<i>Tic-Tac- Toe</i>
Bayes	BayesNET	0.5054	0.2716	0.3797	0.3141	0.6618
	NaiveBayes	0.3146	0.1866	0.3480	0.3109	0.7185
	NaiveBayesMultinomial	0.4905	0.2649	0.2990	0.4273	0.5393
	NaiveBayesSimple	0.3042	0.1818	0.3490	0.3114	0.7381
	ComplementNaiveBayes	0.3500	0.0603	0.2892	0.3297	0.8567
Functions	Logistic	0.2787	0.0816	0.3727	0.3463	0.8401
	MultilayerPerceptron	0.2810	0.0719	0.3316	0.3316	0.6974
	SMO	0.3000	0.2614	0.3373	0.3435	0.8660
Lazy	IB1	0.4000	0.1238	0.3373	0.3387	0.4984
	IBK	0.2722	0.1314	0.3406	0.3388	0.5811
	KStar	0.2962	0.1865	0.3567	0.3388	0.5976
	LWL	0.3992	0.2940	0.3701	0.3721	0.5574
Meta	AdaBoostM1	0.2692	0.3488	0.3693	0.4370	0.7270
	Bagging	0.3492	0.2082	0.3693	0.3355	0.7399
	Decorate	0.3885	0.2778	0.3887	0.3493	0.7468
	LogitBoost	0.2823	0.1572	0.3782	0.3493	0.7604
	Vote	0.5053	0.3914	0.4158	0.4436	0.4586
Misc	HyperPipes	0.5000	0.4348	0.5000	0.4388	0.5059
	VFI	0.4748	0.4326	0.4998	0.4335	0.5171
Rules	JRip	0.4167	0.1974	0.3802	0.3519	0.7295
	OneR	0.2000	0.2794	0.3133	0.3284	0.3645
	PART	0.3792	0.1504	0.3525	0.3464	0.7829
	Ridor	0.4000	0.1238	0.3373	0.3189	0.5763
	ZeroR	0.5053	0.3914	0.4158	0.4436	0.4586
Trees	J48	0.3792	0.2512	0.3629	0.3375	0.7598
	CimpleCart	0.3792	0.2571	0.3547	0.3402	0.7343
	FT	0.2801	0.0827	0.3291	0.3081	0.8676
	NBTree	0.3717	0.2433	0.3804	0.3141	0.5905
Önerilen Yöntem	FWRSC	0.4000	0.5381	0.2289	0.3010	0.4587
	<i>Genel Ortalama</i>	0.3680	0.2373	0.3616	0.3745	0.6528

Çizelge 6.8. Sınıflandırma metotları için f-ölçümü değerleri

<i>Sınıflandırıcı</i>	<i>Sınıflandırma Metotları</i>	<i>Balloon</i>	<i>Balance Scale</i>	<i>Haberman's Survival</i>	<i>Abalone</i>	<i>Tic-Tac- Toe</i>
Bayes	<i>BayesNET</i>	0.229	0.631	0.528	0.498	0.268
	<i>NaiveBayes</i>	0.691	0.843	0.650	0.498	0.175
	<i>NaiveBayesMultinomial</i>	0.771	0.865	0.644	0.343	0.411
	<i>NaiveBayesSimple</i>	0.567	0.848	0.641	0.498	0.175
	<i>ComplementNaiveBayes</i>	0.631	0.876	0.679	0.411	0.143
Functions	<i>Logistic</i>	0.691	0.858	0.605	0.578	0.069
	<i>MultilayerPerceptron</i>	0.691	0.931	0.641	0.539	0.315
	<i>SMO</i>	0.691	0.865	0.528	0.471	0.144
Lazy	<i>IB1</i>	0.567	0.807	0.646	0.498	0.469
	<i>IBK</i>	0.567	0.804	0.624	0.498	0.482
	<i>KStar</i>	0.691	0.805	0.566	0.551	0.380
	<i>LWL</i>	0.567	0.661	0.566	0.465	0.425
Meta	<i>AdaBoostM1</i>	0.691	0.603	0.528	0.465	0.211
	<i>Bagging</i>	0.567	0.781	0.574	0.581	0.247
	<i>Decorate</i>	0.567	0.708	0.535	0.553	0.248
	<i>LogitBoost</i>	0.691	0.806	0.542	0.569	0.170
	<i>Vote</i>	0.229	0.163	0.528	0.191	0.478
Misc	<i>HyperPipes</i>	0.229	0.345	0.528	0.198	0.478
	<i>VFI</i>	0.691	0.565	0.601	0.205	0.112
Rules	<i>JRip</i>	0.567	0.784	0.653	0.546	0.205
	<i>OneR</i>	0.800	0.568	0.635	0.511	0.338
	<i>PART</i>	0.567	0.773	0.620	0.514	0.216
	<i>Ridor</i>	0.567	0.787	0.566	0.548	0.428
	<i>ZeroR</i>	0.229	0.163	0.528	0.191	0.478
Trees	<i>J48</i>	0.567	0.663	0.542	0.524	0.243
	<i>CimpleCart</i>	0.567	0.645	0.624	0.576	0.273
	<i>FT</i>	0.691	0.850	0.549	0.565	0.145
	<i>NBTree</i>	0.800	0.680	0.528	0.498	0.427
Önerilen Yöntem	<i>FWRSC</i>	0,899	0.331	0.435	0.338	0.500
	<i>Ortalama</i>	0.564	0.690	0.580	0.462	0.308

Çizelge 6.6 incelendiğinde FWRSC'nin bir veri kümesi için en yüksek doğruluk değerine sahip iken diğer veri kümelerinde de en yüksek değere yakın bir sınıflandırma işlemini gerçekleştirmiştir. Örneğin, Tic-Tac-Toe veri kümesi için en yüksek sınıflandırmayı % 63.55 oranla OneR metodu yaparken, % 62.30 oranla FWRSC metodu ikinci en yüksek sınıflandırma metotlarından biridir. Tüm veri kümelerine bakıldığında FWRSC metodunun başarı sağladığı söylenebilir.

Çizelge 6.6 üzerinde bazı analizler yapılarak sınıflandırma metotlarının doğruluk ölçütü daha net ifade edilebilir. Bunun için ilk önce her sınıflandırıcının tüm veri kümeleri için doğruluk oranının ortalaması hesaplanabilir. Daha sonra bu ortalamalara bir eşik değeri belirlenerek eşik değeri geçen sınıflandırıcılar farklı bir tablo üzerinde gösterilebilir. Böylece bir filtreleme işlemi ile çizelge daha sade ve küçültülebilir. Çalışmada eşik değeri olarak % 60 alınmıştır. Çünkü % 60'ın aşasında bir değer alındığında sınıflandırıcı sayısı fazla olurdu ve bu da istenilene uygun olmazdı. Yukarısı bir değerde ise çok az sınıflandırıcı kalırdı. Bunun için % 60 eşik değeri olarak seçilmiştir.

5 veri kümesi için ortalama sınıflandırma doğruluk yüzdesi % 60 üzerinde olan sınıflandırma yöntemleri aşağıda verilmiştir.

Çizelge 6.9. Ortalama doğruluk yüzdesi % 60 üzerinde olan sınıflandırma metotları

<i>Veri Kümeleri</i>	Multi	IB1	IBK	KStar	LWL	OneR	Ridor	NBTree	FWRSC
<i>Balloon</i>	70.00	60.00	60.00	70.00	60.00	80.00	60.00	80.00	90.00
<i>Balance Scale</i>	93.33	81.42	80.47	82.38	68.57	58.09	81.42	71.42	65.24
<i>Haberman's</i>	71.08	66.27	65.06	66.27	66.27	68.88	66.26	66.27	66.27
<i>Abalone</i>	58.40	49.19	49.19	55.24	55.91	50.73	55.10	53.36	54.91
<i>Tic-Tac-Toe</i>	30.53	50.15	47.97	37.38	51.71	63.55	42.36	42.05	62.30
<i>Ortalama</i>	64.67	61.41	60.54	62.25	60.49	64.25	61.03	62.62	67.74

FWRSC metodunun başarısını Çizelge 6.9 üzerinde daha net görmek mümkündür. Bu çizelgede tüm veri kümeleri için sınıflandırma doğruluk başarı ortalaması % 60 üzerinde olan metotlar alınmıştır. Diğer metotlar ise elenmiştir. Bu çizelgede görüldüğü gibi genel ortalaması en yüksek metot % 67.74 oranla FWRSC metodudur. Bu durum, FWRSC metodunun WEKA'da var olan sınıflandırma metotlarından birçoğundan daha iyi ve güçlü olduğunu gösterir. Başka bir deyişle, FWRSC algoritması WEKA'daki metotların performans açısından en iyileri, en

güçlüleri arasında kendisine yer bulduğu söylenilebilir. Ayrıca Çizelge 6.7 ve Çizelge 6.8 bakıldığında *MAE* tablosunda, *Haberman's Survival* ve *Abalone* veri kümeleri için en iyi sonucu FWRSC elde ederken, diğer veri kümelerinde *Balance Scale* veri kümesi dışında kötü sonuç üretmediği açıktır. *f-ölçümü* tablosunda, *Balloon* ve *Tic-Tac-Toe* veri kümelerinde en iyi sonucu FWRSC üretmişken, diğer veri kümelerinde *Balance Scale* ve *Abalone* kümelerinde istenilen başarıya ulaşamamıştır. Bu noktada FWRSC *f-ölçümü* ölçütü yönünde istenilen başarıyı verememiştir denilebilir.

7. TARTIŞMA, SONUÇ VE ÖNERİLER

Bu çalışmada iki önemli konu işlenmiştir. Bunlardan birincisi hayati önem taşıyan kalp sağlığı ile ilgili 4 hastalığın erken tanısına yardımcı olacak bir model sunulmasıdır. Bu modele göre 689 hastanın gerçek EKG sinyalleri sınıflandırılmıştır. Modelin çatısını KKT ve sınıflandırma yöntemleri oluşturmaktadır. KKT'nin bu modele en önemli katkısı gereksiz nitelikler elenmesini sağlayarak yürütme zamanında iyileştirmeler getirmesidir. Bu durum tablolar üzerinde kullanılan her bir sınıflandırma metodu için açık bir şekilde gösterilmiştir. Yapılan sınıflandırmada % 85'e yakın bir başarı elde edilmiştir. Buna ek olarak deney sonuçları, ilgili alanda uzmanların kontrolüne sunulmuş ve bilgisine başvurulmuştur. Uzmanlardan alınan geri bildirimlerle sonuçların yüksek oranda uyduğu görülmüştür. Uyuşmazlıkların daha çok aynı ana başlık altındaki hastalıkların sınıflandırılmasından kaynaklandığı anlaşılmıştır. Bunun nedeni aynı ana başlık altındaki hastalıkların teşhisinde benzer nitelik sayısının fazla olmasıdır. Buna çözüm olarak, her niteliğin sınıf üzerindeki ağırlığının hesaplanmasıyla niteliğin sınıf belirlemedeki değeri azaltılarak veya artırılarak yeni bir sınıflandırma metodu öngörülebilir. Ayrıca KKT'nin eksik veriden çıkarım yapma yeteneğinden faydalanarak sınıflandırma yöntemleri ile KKT'nin hibrid algoritmaları geliştirilebilir. Özellikle kural tabanlı sınıflandırma yöntemleri ve KKT karışımı algoritmalar geliştirilebilir. Bu çalışmaların gelişimine paralel olarak, EKG sinyallerinin bilgisayar yazılımları yardımıyla yorumlanması, ileride "Kalp Krizini"

önceden haber veren sistemlerin geliştirilmesinde temel unsur olacağı söylenebilir. Günümüzde hızla gelişmekte olan mobil teknolojisine bağlı olarak, mobil cihazlarda EKG yorumu yapan bir sistem olarak geliştirilebilir.

İkinci olarak FWRSC adında yeni bir sınıflandırma metodu geliştirilmiştir. Bu metot sayesinde gizli sınıflandırmaların önüne geçilmiş, karmaşıklık (komplekslik) ve tutarsızlık azaltılmıştır. Ayrıca bir nesne için nitelik ve sınıf arasındaki ilişkisi açık, anlaşılır bir şekilde ortaya çıkartılmıştır. Buna bağlı olarak, bir nesnenin sınıfını belirleme etkisi fazla olan çekirdek ve etkisi daha az olan indirgenen nitelikler belirlenmiştir. Sınıflandırma işleminin bağlı olduğu nesne, nitelik ve sınıf üçlemesinin çekirdek ve indirgenen nitelikler yardımı ile basitleştirilmesi, sınıflandırma işleminde bir nitelik değerinin birden fazla nesne için ortak olma probleminin etkisinin azalmasına neden olmuştur. Buna bağlı olarak da daha hassas ve doğru bir sınıflandırma işlemi yapılmıştır. Çalışmada 5 farklı alandaki veri kümeleri kullanılmıştır. Geliştirilmiş olan sınıflandırma yöntemi, WEKA'daki sınıflandırma yöntemleriyle karşılaştırılmıştır. Sonuçlar çizelgeler üzerinde açık bir şekilde gösterilmiştir. Tüm veri kümeleri için doğruluk açısından, % 67,74 ortalama ile en yüksek doğruluk oranını FWRSC göstermiştir. MAE açısından, iki veri kümesi için en iyi sonucu vermesinin yanında 0,385 oranla da genel ortalamanın altında performans göstererek iyi bir hata oranına sahip olduğunu ispatlamıştır. Ancak f-ölçümü açısından, iki veri kümesi için en iyi sonucu vermesine rağmen 0,5006'lık bir oranla genel ortalamanın altında bir performans göstermesi, dengeli bir sınıflandırma yapmada biraz zayıf kaldığı söylenebilir. Bir sonraki araştırma ve çalışmalarda bu durumun iyileştirmesi için gerekli çalışmalar yürütülecektir. Ayrıca FWRSC'nin sadece f-ölçümü kriteri bazında değil diğer kriterler yönünden de performansını iyileştirici çalışmalar yapılabilir. Örneğin, çekirdek niteliklere indirgenme safhasındaki ayırt edilebilir fonksiyonların hesaplanması gibi işlemler daha hassas ve detaylı yapılarak FWRSC'nin performansı daha da artırılabilir. Bu işlemler tıpkı kesişen iki kümenin kesişen bölgedeki elemanlarını birbirinden ayırmak için çekilen ince ve hassas bir çizgi gibidir. Bu çizgi ne kadar ince ve hassas olursa iki kümenin ayırt ediciliği o kadar fazla olur.

Kısacası, deney sonuçlarındaki tüm çizelgeler değerlendirildiğinde FWRSC metodu, WEKA programındaki sınıflandırma metotlarının birçoğundan daha etkili ve yüksek performans gösterdiği açıktır. FWRSC metodunun WEKA programındaki sınıflandırma metotları ile yarışabileceği ve mücadele edebileceğini demek mümkündür. Yani FWRSC'nin bu sınıflandırma metotları arasında kendine yer bulduğunu ve hatta listenin üst sıralarında olduğunu demek yanlış olmaz. FWRSC'nin bu başarısının arkasında kaba kümelerin ayırt edilebilirlik matrisi ve fonksiyonlardaki başarısı olduğunu tekrar söylemekte ve altını çizmekte fayda vardır. Son olarak FWRSC'nin hangi tip veri üzerinde daha etkili ve verimli çalıştığını belirleme işlemleri sonraki çalışmalarda yapılacağını belirlemede yarar vardır.

KAYNAKLAR

- [1] Efeoğlu, Acil Tıp Eğitimi İçin EKG, EKG Kütüphanesi Sayfası. Ana sayfa. < <http://www.acilci.net/> > Accessed at 02.10.2014
- [2] Çekik, R., ve Telçeken, S., “EKG Sinyallerinin Kaba Kümeler Teorisi Kullanılarak Sınıflandırılması”, *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi - A - Uygulamalı Bilimler ve Mühendislik*, 15 (2), 2014.
- [3] Avşar, A. ve Kaya M., “Kaba Kümeler Yardımıyla Eksik Verilerden Kesin ve Olası Bulanık Kuralların Çıkarılması”, *Elektrik-Elektronik-Bilgisayar Mühendisliği 12. Ulusal Kongresi ve Fuarı*, 2007.
- [4] Chmielewski, C., Michal R., ve Grzymala-Busse, J. W., "Global Discretization of Continuous Attributes as Preprocessing for Machine Learning" *International Journal of Approximate Reasoning*, 15 (4): 319-331, 1996.
- [5] Lingras, L. P. J., ve Yao, Y. Y., "Data Mining using Extensions of The Rough Set Model" *Journal of The American Society for Information Science*, 49 (5): 415-422, 1998.
- [6] Telçeken, T. ve Doğan, M., “Kaba Kümeler Teorisi Yardımı ile Büyük Veri Topluluklarının Analizi”, *ELECO'04*, Cilt1, 414-418, 2004.
- [7] Pawlak, Z., “Rough sets”, *International Journal of Computer & Information Sciences*, 11 (5):341-356, 1982.
- [8] Witten, I. H., ve Frank, E., *Data Mining, Practical Machine Learning Tools and techniques*, San Francisco, CA: Morgan Kaufmann Publishers, Elsevier, 2005.
- [9] Chang, L. Y., Wang, G. Y., ve Wu, Y., “An Approach for Attribute Reduction and Rule Generation Based on Rough Set Theory”, *Journal of software*, 11, 015, 1999.
- [10] Aysel, Dr. Süleyman Aysel Kardiyoloji Uzmanı, Elektrokardiyografi Sayfası. <<http://drsuleymanaysel.com/17/19/kardiyolojide-tani-yontemleri/elektrokardiyografi-ekg/>> Accessed at 05.08.2014

- [11] Malmivuo, J., ve Plonsey, “R., Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields , The Basis of ECG Diagnosis”, Oxford University Press, England , 1995.
- [12] Thygesen, K., Alpert, J. S., ve White, H. D., “Universal definition of myocardial infarction”, *Journal of the American College of Cardiology*, 50 (22): 2173-2195, 2007.
- [13] Simoons, M., Topol, E., Califf, R., van de Werf, F., Armstrong, P., Aylward, P. E., ve Hampton, J., “An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction”, *New England Journal of Medicine*, 329 (10): 673-682, 1993.
- [14] Kleiger, R. E., Miller, J. P., Bigger, J. T., ve Moss, A. J., “Decreased heart rate variability and its association with increased mortality after acute myocardial infarction”, *The American journal of cardiology*, 59 (4): 256-262, 1987.
- [15] Kawamoto, A., Gwon, H. C., Iwaguro, H., Yamaguchi, J. I., Uchida, S., Masuda, H., ve Asahara, T., “Therapeutic potential of ex vivo expanded endothelial progenitor cells for myocardial ischemia”, *Circulation*, 103 (5): 634-637, 2001.
- [16] Wackers, F. J., Young, L. H., Inzucchi, S. E., Chyun, D. A., Davey, J. A., Barrett, E. J., ve Iskandrian, A. E., “Detection of silent myocardial ischemia in asymptomatic diabetic subjects: the DIAD study”, *Diabetes care*, 27 (8): 1954-1961, 2004.
- [17] Devereux, R. B., Alonso, D. R., Lutas, E. M., Gottlieb, G. J., Campo, E., Sachs, I., ve Reichek, N., “Echocardiographic assessment of left ventricular hypertrophy: comparison to necropsy findings”, *The American journal of cardiology*, 57 (6): 450-458, 1986.
- [18] Nordestgaard, B. G., Benn, M., Schnohr, P., ve Tybjaerg-Hansen, A., “Nonfasting triglycerides and risk of myocardial infarction, ischemic heart disease, and death in men and women”, *Jama*, 298 (3): 299-308, 2007.
- [19] Haffner, S. M., Lehto, S., Rönnemaa, T., Pyörälä, K., ve Laakso, M., “Mortality from coronary heart disease in subjects with type 2 diabetes and

- in nondiabetic subjects with and without prior myocardial infarction”, *New England journal of medicine*, 339 (4), 229-234, 1998.
- [20] Milli Eğitim Bakanlığı, Acil Sağlık Hizmetleri, EKG (Elektrokardiyografi), Ankara, 723H00048, (2011).
- [21] Yanowitz, F. G. MD., “Introduction to ECG Interpretation V8. 0”, *IHC ECG Services LDS Hospital & Intermountain Medical Center Salt Lake City, Utah*, 2012.
- [22] Sağlık.im, Miyokard İnfarktüsü sayfası. <<http://www.saglik.im/miyokard-infarktusu/>> Accessed at 02.10.2014.
- [23] Bilal, M. S., ve Sarioğlu, T., “İskemik Miyokard İnjurisi ve İntraoperatif Miyokard Korunmasına Genel Bir Bakış”, *Türk Göğüs Kalp Damar Cerrahisi Dergisi*, 1 (2): 118-126, 1992.
- [24] Kloner, R. A., Arimie, R. B., Kay, G. L., Cannom, D., Matthews, R., Bhandari, A., ve Burstein, S., “Evidence for stunned myocardium in humans: a 2001 update”, *Coronary artery disease*, 12 (5): 349-356, 2001.
- [25] Heyndrickx, G. R., Millard, R. W., McRitchie, R. J., Maroko, P. R., ve Vatner, S. F., “Regional myocardial functional and electrophysiological alterations after brief coronary artery occlusion in conscious dogs”, *Journal of Clinical Investigation*, 56 (4): 978, 1975.
- [26] Bolli, R., “Oxygen-derived free radicals and postischemic myocardial dysfunction (“stunned myocardium”)”, *Journal of the American College of Cardiology*, 12 (1): 239-249, 1988.
- [27] Abbott Türkiye, Kalp Hastalıkları sayfası. <https://www.abbott.com.tr/tedavi_alanlari_ve_ilac/kalp_hastaliklari.aspx> Accessed at 15.09.2014.
- [28] Gertsch, M., “The Normal ECG and its (Normal) Variants. In The ECG (pp. 19-43)”, *Springer Berlin Heidelberg*, 2004.
- [29] Komorowski, J., Pawlak, Z., Polkowski, L., ve Skowron, A., “Rough sets: A tutorial. Rough fuzzy hybridization: A new trend in decision-making”, *Springer-Verlag, Singapore*, 3-98, 1999.
- [30] Pawlak, Z., ve Skowron, A., “Rudiments of rough sets”, *Information sciences*, 177 (1): 3-27, 2007.

- [31] Aydoğan, A., Kızılkaya, E. ve Gencer, C., "Kaba Küme Yaklaşımı Kullanılarak Veri Madenciliği Problemlerinde Sınıflandırma Amaçlı Yapılmış Olan Çalışmalar", *Savunma Bilimleri Dergisi*, 6 (2): 17-32, 2007.
- [32] Aybar, A., *Kaba Kümeler Teorisi Üzerine Algoritmalar*, Yüksek Lisans Tezi, Ankara Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, 2011.
- [33] Predki, B., Słowiński, R., Stefanowski, J., Susmaga, R., ve Wilk, S., "ROSE-software implementation of the rough set theory: In Rough Sets and Current Trends in Computing", *Springer Berlin Heidelberg*, 1998.
- [34] Dener, M., Dörterler, M., ve Orman, A. "Açık Kaynak Kodlu Veri Madenciliği Programları: Weka'da Örnek Uygulama", *Akademik Bilişim*, 9, 11-13, 2009.
- [35] ProSoft, 1999. ROSE (Rough Set Data Explorer) Version 2.0 – User's Guide. Ana Sayfa.
<<http://www-idss.cs.put.poznan.pl/software/rose/main.html>> Accessed at 02.5.2014.
- [36] WEKA, Weka 3: Data Mining Software in Java Sayfası,
<<http://www.cs.waikato.ac.nz/~ml/weka/index.html>> Accessed at 2.01.2014
- [37] Cohen, W. W., "Fast effective rule induction", *In Proceedings of the twelfth international conference on machine learning*, 115-123, 1995.
- [38] Tan, P. N., Steinbach, M., & Kumar, V., "Classification: basic concepts, decision trees, and model evaluation", *Introduction to Data Mining*, 1, 145-205, 2006.
- [39] Rish, I., "An empirical study of the naive Bayes classifier", *In IJCAI 2001 workshop on empirical methods in artificial intelligence, IBM New York*, 3 (22), 41-46, , 2001.
- [40] Lewis, D. D., "Naive (Bayes) at forty: The independence assumption in information retrieval", *In Machine learning: ECML-98, 4-15, Springer Berlin Heidelberg*, 1998.
- [41] Machine Learning Repository Sayfası, Ana Sayfası.
< <https://archive.ics.uci.edu/ml/datasets.html> > Accessed at 15.09.2014.

- [42] Zhu, W., ve Wang, F. Y., “Reduction and axiomization of covering generalized rough sets”, *Information sciences*, 152, 217-230, 2003.
- [43] Kusiak, A., “Rough set theory: a data mining tool for semiconductor manufacturing”, *Electronics Packaging Manufacturing, IEEE Transactions on*, 24(1): 44-50, 2001.
- [44] Thangavel, K., ve Pethalakshmi, A., “Dimensionality reduction based on rough set theory: A review”, *Applied Soft Computing*, 9(1): 1-12, 2009.
- [45] Han, J., Kamber, M., ve Pei, J., “Data mining, southeast asia edition: Concepts and techniques”, *Morgan kaufmann*, 2006.
- [46] Zhong, N., ve Zhou, L. (Eds.), “Methodologies for Knowledge Discovery and Data Mining:”, *Third Pacific-Asia Conference, PAKDD'99, Beijing, China, April 26-28, 1999, Proceedings (Vol. 1574). Springer Science & Business Media.*, 1999.

EK-1 FWRSC Algoritması

A. FWRSC Algoritması (Ana program)

```
% Algoritma prosedürü
% Giriş verisi: Eğitim kümesi (T = (U, A, D) bir karar tablosu)
begin
for i = 1 to ClassNumber
    for j=1 to size(T(:, 1))
        if (Classes(j) == ClassVektor)
            SeparateClassIndex();
        end {if}
    end{for j}
end{for i}
[ DiscernibilityFunctions ] = CalculateDiscernibilityFunctions(T)
for i = 1 to length(TestData)
    for j = 1 to ClassNumber
        WhichClass = SeparateClassIndex(:, j);
    [WeightMatrix] = CalculateWeightMatrix(Discernibility, WhichClass)
        TestVector = TestData(i, :);
    CalculateClass(T, TestVector, WhichClass, WeightMatrix)
    end {for j}
    end {for i}
end {for begin}
```

B. CalculateClass Metodu

```
% Giriş verisi: {T, TestVector, WhichClass, WeightMatrix}
begin
for i = 1 to length(WhichClass)
    for j = 1 to size( T(:,1))
if (T(WhichClass(i), j) == TestVector(j))
    Similarity++;
    Weight = WeightMatrix(i, j) + Weight;
end{for if}
    end {for j}
    end {for i}
ClassWeight = (Similarity + Weight) /size(WhichClass);
end {for begin}
```

C. CalculateWeightMatrix Metodu

```
% Giriş Verisi: {Discernibility, WhichClass}
begin
for i = 1 to length(WhichClass)
vector = WhichClass(i);
    for j = 1 to size(T(:,2))
    for k = 1 to lenght(vector)
    if (T(:,2).(j) == vector(j))
Senbol_j++;
    end{for if}
    if (vector(j) == 'V')
Senbol_V++;
    end{for if} end {for k}
    Weight = Senbol_j / ( Senbol_V +1);
WeightMatrisTotal(i,j) = Weight;
    end {for j}
    end {for i}
end {for begin}
```

D. CalculateDiscernibilityFunctions

```
% Giriş Verisi: {T: Bilgi sistemi}
Begin
while denge < size(T,1)
  for i = denge to size(T,1)
    if (T(i, size(T, 2)) ~= T(denge, size(T, 2)))
      for j = 1 to size(T, size(T, 2))
        if (T(i, j) ~= T(denge, j))
          Nit(j) = j, durum= 1;
        else
          Nit(j) = 0, durum = 1;
        end {if-else}
      end {for j}
    else
      DisMatrix{k} = 0;
    end {if else}
  if (durum == 1)
    DisMatrix{k} = Nit;
  end {for if}
  durum = 0;
  k++;
end {for i}
  for h =1 to denge
    DisMatris{k}=0;
    k=k+1;
  end {for h}
  denge ++;
end {while denge}
for i =1 to size(T, 1)
  for j = 1 to size(T, 1)
    if (i == j)
      DisMatris{i,j}= 0;
    else
      DisMatris{i,j} = DisMatris{j,i};
    end {if-else}
  end {for j}
end {for i}
while r < size (T, 2)
A = simplify( A | Nitelik)
C= simplify( A & C)
r++;
end {while r}
```

Ek-2 ROSE programında üretilen tüm kurallar

ROSE Programında Üretilen Kurallar...	
#	LEM2
#	C:\Users\Rasim\Desktop\Uygulama\Rose2\Uygulama1\DataEKG -12.isf
#	objects = 689
#	attributes = 12
#	decision = Dec
#	classes = {1, 2, 3, 4, 5}
#	Mon Mar 23 14:25:08 2015
#	0 s
rule 1. (A4 = 2) => (Dec = 1); [82, 82, 69.49%, 100.00%][82, 0, 0, 0, 0]	
[5, 9, 10, 19, 22, 29, 32, 46, 47, 65, 75, 80, 105, 120, 130, 144, 156, 169, 174, 188, 204, 208, 218, 225, 247, 254, 264, 271, 292, 297, 300, 303, 307, 325, 335, 349, 364, 367, 374, 377, 413, 419, 427, 431, 441, 445, 447, 452, 463, 469, 476, 481, 483, 486, 492, 495, 514, 521, 530, 537, 545, 546, 547, 549, 552, 554, 556, 561, 567, 568, 592, 597, 605, 620, 650, 651, 656, 661, 668, 672, 681, 684], {}, {}, {}, {}]	
rule 2. (A3 = 3) & (A5 = 2) & (A11 = 1) => (Dec = 1); [41, 41, 34.75%, 100.00%][41, 0, 0, 0, 0]	
[10, 22, 46, 65, 87, 130, 144, 202, 204, 218, 225, 233, 247, 254, 264, 266, 279, 297, 300, 303, 307, 320, 333, 349, 364, 367, 427, 447, 481, 495, 507, 514, 521, 546, 568, 605, 620, 651, 656, 681, 684], {}, {}, {}, {}]	
rule 3. (A2 = 1) & (A3 = 2) & (A5 = 2) => (Dec = 1); [12, 12, 10.17%, 100.00%][12, 0, 0, 0, 0]	
[174, 261, 271, 343, 377, 452, 545, 553, 586, 592, 599, 668], {}, {}, {}, {}]	
rule 4. (A1 = 1) & (A3 = 3) & (A5 = 2) => (Dec = 1); [17, 17, 14.41%, 100.00%][17, 0, 0, 0, 0]	
[22, 65, 130, 228, 264, 266, 297, 300, 307, 320, 364, 546, 568, 651, 656, 672, 684], {}, {}, {}, {}]	
rule 5. (A1 = 3) & (A2 = 2) & (A3 = 3) & (A10 = 2) => (Dec = 1); [2, 2, 1.69%, 100.00%][2, 0, 0, 0, 0]	
[575, 630], {}, {}, {}, {}]	
rule 6. (A3 = 2) & (A5 = 2) & (A6 = 2) & (A8 = 2) & (A10 = 2) & (A11 = 2) => (Dec = 1); [7, 7, 5.93%, 100.00%][7, 0, 0, 0, 0]	
[43, 67, 169, 452, 549, 592, 599], {}, {}, {}, {}]	
rule 7. (A2 = 1) & (A9 = 1) & (A10 = 2) => (Dec = 1); [2, 2, 1.69%, 100.00%][2, 0, 0, 0, 0]	
[614, 628], {}, {}, {}, {}]	
rule 8. (A1 = 1) & (A3 = 3) & (A7 = 1) => (Dec = 1); [1, 1, 0.85%, 100.00%][1, 0, 0, 0, 0]	
[394], {}, {}, {}, {}]	
rule 9. (A1 = 1) & (A6 = 2) & (A8 = 1) => (Dec = 1); [2, 2, 1.69%, 100.00%][2, 0, 0, 0, 0]	
[638, 651], {}, {}, {}, {}]	
rule 10. (A1 = 1) & (A2 = 2) & (A3 = 3) & (A7 = 2) & (A11 = 1) => (Dec = 1); [5, 5, 4.24%, 100.00%][5, 0, 0, 0, 0]	
[214, 266, 546, 568, 684], {}, {}, {}, {}]	
rule 11. (A2 = 1) & (A6 = 1) & (A8 = 2) & (A10 = 2) => (Dec = 1); [4, 4, 3.39%, 100.00%][4, 0, 0, 0, 0]	
[2, 349, 553, 614], {}, {}, {}, {}]	
rule 12. (A2 = 2) & (A3 = 2) & (A5 = 2) & (A8 = 2) & (A10 = 2) => (Dec = 1); [4, 4, 3.39%, 100.00%][4, 0, 0, 0, 0]	
[67, 80, 287, 325], {}, {}, {}, {}]	
rule 13. (A2 = 1) & (A9 = 2) & (A10 = 1) & (A11 = 1) => (Dec = 2); [42, 42, 19.91%, 100.00%][0, 42, 0, 0, 0]	
[41, 58, 68, 91, 159, 162, 167, 237, 238, 243, 248, 253, 260, 267, 283, 285, 332, 336, 372, 392, 402, 420, 425, 426, 430, 491, 500, 532, 544, 550, 574, 591, 593, 600, 607, 625, 627, 631, 639, 653, 662, 680], {}, {}]	
rule 14. (A2 = 1) & (A3 = 1) & (A11 = 1) => (Dec = 2); [15, 15, 7.11%, 100.00%][0, 15, 0, 0, 0]	
[68, 81, 94, 98, 114, 119, 137, 170, 242, 243, 260, 420, 518, 570, 608], {}, {}, {}]	
rule 15. (A2 = 3) & (A3 = 3) & (A4 = 1) & (A5 = 1) => (Dec = 2); [26, 26, 12.32%, 100.00%][0, 26, 0, 0, 0]	
[25, 64, 82, 177, 203, 215, 235, 251, 263, 268, 317, 321, 355, 378, 412, 418, 432, 438, 454, 468, 519, 526, 560, 616, 646, 655], {}, {}, {}]	
rule 16. (A1 = 3) & (A2 = 1) & (A5 = 1) & (A9 = 2) => (Dec = 2); [25, 25, 11.85%, 100.00%][0, 25, 0, 0, 0]	
[76, 129, 159, 171, 213, 243, 252, 260, 283, 315, 332, 357, 365, 380, 392, 425, 440, 482, 511, 527, 544, 563, 593, 631, 639], {}, {}, {}]	
rule 17. (A2 = 2) & (A5 = 1) & (A9 = 2) & (A10 = 1) & (A11 = 1) => (Dec = 2); [26, 26, 12.32%, 100.00%][0, 26, 0, 0, 0]	
[61, 123, 125, 133, 164, 181, 187, 189, 198, 199, 211, 276, 277, 281, 286, 334, 351, 370, 401, 409, 428, 450, 475, 484, 540, 583], {}, {}, {}]	
rule 18. (A1 = 1) & (A5 = 1) & (A10 = 1) => (Dec = 2); [30, 30, 14.22%, 100.00%][0, 30, 0, 0, 0]	
[58, 68, 91, 167, 185, 199, 237, 238, 248, 277, 336, 351, 372, 402, 491, 500, 506, 532, 535, 550, 560, 581, 591, 600, 607, 616, 625, 627, 674, 680], {}, {}, {}]	
rule 19. (A2 = 1) & (A7 = 1) & (A11 = 1) => (Dec = 2); [2, 2, 0.95%, 100.00%][0, 2, 0, 0, 0]	
[319, 501], {}, {}, {}]	
rule 20. (A9 = 1) & (A10 = 1) => (Dec = 2); [4, 4, 1.90%, 100.00%][0, 4, 0, 0, 0]	
[535, 624, 632, 689], {}, {}, {}]	
rule 21. (A1 = 2) & (A2 = 2) & (A3 = 3) & (A4 = 1) & (A5 = 1) => (Dec = 2); [28, 28, 13.27%, 100.00%][0, 28, 0, 0, 0]	
[61, 116, 123, 125, 133, 153, 164, 184, 189, 195, 198, 211, 246, 276, 281, 286, 334, 370, 393, 401, 415, 423, 428, 450, 475, 484, 531, 583], {}, {}, {}]	
rule 22. (A1 = 2) & (A3 = 2) & (A5 = 1) & (A9 = 3) & (A10 = 1) & (A11 = 2) => (Dec = 2); [2, 2, 0.95%, 100.00%][0, 2, 0, 0, 0]	

[{}], {529, 685}, {}, {}, {}
rule 23. (A6 = 2) & (A7 = 3) & (A8 = 2) & (A11 = 1) => (Dec = 2); [5, 5, 2.37%, 100.00%][0, 5, 0, 0, 0]
[{}], {286, 329, 455, 548, 591}, {}, {}, {}
rule 24. (A4 = 1) & (A5 = 1) & (A6 = 3) & (A8 = 2) => (Dec = 2); [3, 3, 1.42%, 100.00%][0, 3, 0, 0, 0]
[{}], {542, 576, 637}, {}, {}, {}
rule 25. (A2 = 3) & (A3 = 2) & (A9 = 1) => (Dec = 2); [1, 1, 0.47%, 100.00%][0, 1, 0, 0, 0]
[{}], {453}, {}, {}, {}
rule 26. (A1 = 1) & (A2 = 1) & (A5 = 1) => (Dec = 2); [24, 24, 11.37%, 100.00%][0, 24, 0, 0, 0]
[{}], {7, 58, 68, 91, 136, 167, 185, 237, 238, 248, 336, 372, 402, 491, 500, 506, 532, 550, 591, 600, 607, 625, 627, 680}, {}, {}, {}
rule 27. (A1 = 3) & (A2 = 3) & (A4 = 1) & (A9 = 2) & (A11 = 1) => (Dec = 2); [9, 9, 4.27%, 100.00%][0, 9, 0, 0, 0]
[{}], {15, 23, 251, 268, 432, 579, 637, 655, 660}, {}, {}, {}
rule 28. (A2 = 3) & (A7 = 1) & (A11 = 2) => (Dec = 2); [1, 1, 0.47%, 100.00%][0, 1, 0, 0, 0]
[{}], {407}, {}, {}, {}
rule 29. (A1 = 2) & (A2 = 3) & (A4 = 1) & (A8 = 1) => (Dec = 2); [1, 1, 0.47%, 100.00%][0, 1, 0, 0, 0]
[{}], {299}, {}, {}, {}
rule 30. (A2 = 2) & (A3 = 2) & (A5 = 1) & (A9 = 2) & (A10 = 1) => (Dec = 2); [2, 2, 0.95%, 100.00%][0, 2, 0, 0, 0]
[{}], {95, 351}, {}, {}, {}
rule 31. (A1 = 2) & (A2 = 1) & (A5 = 1) & (A8 = 3) => (Dec = 2); [2, 2, 0.95%, 100.00%][0, 2, 0, 0, 0]
[{}], {165, 311}, {}, {}, {}
rule 32. (A3 = 3) & (A4 = 1) & (A5 = 1) & (A11 = 2) => (Dec = 2); [29, 29, 13.74%, 100.00%][0, 29, 0, 0, 0]
[{}], {25, 64, 76, 82, 116, 165, 177, 184, 193, 195, 203, 215, 231, 235, 246, 317, 355, 378, 393, 412, 423, 440, 454, 468, 519, 525, 531, 560, 646}, {}, {}, {}
rule 33. (A1 = 2) & (A3 = 3) & (A4 = 1) & (A11 = 2) => (Dec = 2); [25, 25, 11.85%, 100.00%][0, 25, 0, 0, 0]
[{}], {25, 64, 82, 116, 165, 177, 184, 195, 203, 215, 231, 235, 246, 298, 317, 355, 378, 393, 412, 423, 454, 468, 519, 531, 646}, {}, {}, {}
rule 34. (A1 = 2) & (A2 = 1) & (A9 = 3) & (A10 = 2) => (Dec = 2); [1, 1, 0.47%, 100.00%][0, 1, 0, 0, 0]
[{}], {470}, {}, {}, {}
rule 35. (A2 = 3) & (A3 = 2) & (A4 = 1) & (A5 = 1) & (A8 = 2) & (A11 = 2) => (Dec = 2); [1, 1, 0.47%, 100.00%][0, 1, 0, 0, 0]
[{}], {391}, {}, {}, {}
rule 36. (A1 = 3) & (A6 = 1) & (A7 = 1) & (A11 = 2) => (Dec = 3); [4, 4, 4.21%, 100.00%][0, 0, 4, 0, 0]
[{}], {326, 421, 621, 683}, {}, {}
rule 37. (A1 = 1) & (A3 = 2) & (A6 = 2) & (A7 = 1) => (Dec = 3); [6, 6, 6.32%, 100.00%][0, 0, 6, 0, 0]
[{}], {173, 186, 389, 434, 502, 601}, {}, {}
rule 38. (A1 = 2) & (A3 = 2) & (A4 = 1) & (A6 = 2) & (A7 = 1) & (A11 = 2) => (Dec = 3); [3, 3, 3.16%, 100.00%][0, 0, 3, 0, 0]
[{}], {44, 59, 273}, {}, {}
rule 39. (A6 = 1) & (A8 = 1) => (Dec = 3); [2, 2, 2.11%, 100.00%][0, 0, 2, 0, 0]
[{}], {294, 676}, {}, {}
rule 40. (A1 = 3) & (A6 = 1) & (A9 = 1) => (Dec = 3); [1, 1, 1.05%, 100.00%][0, 0, 1, 0, 0]
[{}], {633}, {}, {}
rule 41. (A1 = 3) & (A4 = 1) & (A6 = 1) & (A7 = 2) & (A8 = 2) & (A10 = 2) => (Dec = 3); [1, 1, 1.05%, 100.00%][0, 0, 1, 0, 0]
[{}], {598}, {}, {}
rule 42. (A6 = 2) & (A7 = 1) & (A8 = 3) & (A11 = 1) => (Dec = 3); [1, 1, 1.05%, 100.00%][0, 0, 1, 0, 0]
[{}], {17}, {}, {}
rule 43. (A1 = 2) & (A2 = 2) & (A6 = 1) & (A7 = 2) & (A11 = 2) => (Dec = 3); [1, 1, 1.05%, 100.00%][0, 0, 1, 0, 0]
[{}], {667}, {}, {}
rule 44. (A3 = 1) & (A7 = 2) & (A10 = 2) & (A11 = 2) => (Dec = 4); [36, 36, 24.83%, 100.00%][0, 0, 0, 36, 0]
[{}], {4, 26, 35, 54, 108, 112, 140, 152, 163, 172, 197, 209, 230, 240, 274, 313, 341, 347, 353, 384, 387, 408, 411, 422, 466, 474, 488, 497, 503, 528, 541, 558, 572, 611, 613, 687}, {}
rule 45. (A8 = 3) & (A9 = 3) => (Dec = 4); [26, 26, 17.93%, 100.00%][0, 0, 0, 26, 0]
[{}], {20, 239, 241, 259, 290, 295, 312, 328, 338, 340, 358, 362, 388, 395, 422, 437, 457, 467, 479, 489, 498, 524, 622, 626, 664, 686}, {}
rule 46. (A3 = 1) & (A5 = 2) & (A10 = 2) => (Dec = 4); [38, 38, 26.21%, 100.00%][0, 0, 0, 38, 0]
[{}], {4, 31, 51, 69, 77, 88, 132, 140, 152, 163, 178, 182, 197, 230, 244, 289, 290, 353, 384, 408, 411, 429, 443, 474, 477, 480, 488, 494, 505, 508, 551, 558, 584, 590, 611, 640, 649, 677}, {}
rule 47. (A1 = 2) & (A5 = 1) & (A6 = 2) & (A7 = 2) & (A8 = 3) & (A10 = 2) => (Dec = 4); [14, 14, 9.66%, 100.00%][0, 0, 0, 14, 0]
[{}], {20, 102, 217, 229, 395, 397, 406, 457, 489, 520, 523, 524, 541, 634}, {}
rule 48. (A3 = 1) & (A9 = 3) => (Dec = 4); [31, 31, 21.38%, 100.00%][0, 0, 0, 31, 0]
[{}], {31, 37, 42, 51, 54, 69, 108, 112, 140, 172, 219, 259, 290, 384, 395, 422, 437, 443, 479, 480, 494, 505, 524, 558, 584, 590, 602, 636, 640, 647, 670}, {}
rule 49. (A9 = 3) & (A10 = 2) & (A11 = 2) => (Dec = 4); [26, 26, 17.93%, 100.00%][0, 0, 0, 26, 0]
[{}], {20, 54, 108, 112, 140, 172, 239, 241, 295, 338, 340, 362, 384, 388, 422, 457, 467, 489, 498, 555, 558, 609, 622, 626, 682, 686}, {}
rule 50. (A1 = 3) & (A2 = 1) & (A9 = 3) => (Dec = 4); [5, 5, 3.45%, 100.00%][0, 0, 0, 5, 0]
[{}], {124, 358, 424, 642, 682}, {}

rule 51. (A2 = 2) & (A5 = 2) & (A10 = 1) => (Dec = 4); [8, 8, 5.52%, 100.00%][0, 0, 0, 8, 0]
[{} , {} , {} , {6, 13, 42, 89, 479, 577, 636, 659}, {}]
rule 52. (A9 = 3) & (A10 = 1) & (A11 = 1) => (Dec = 4); [7, 7, 4.83%, 100.00%][0, 0, 0, 7, 0]
[{} , {} , {} , {42, 330, 437, 479, 513, 636, 670}, {}]
rule 53. (A3 = 1) & (A7 = 3) => (Dec = 4); [1, 1, 0.69%, 100.00%][0, 0, 0, 1, 0]
[{} , {} , {} , {339}, {}]
rule 54. (A6 = 3) & (A11 = 2) => (Dec = 4); [2, 2, 1.38%, 100.00%][0, 0, 0, 2, 0]
[{} , {} , {} , {278, 622}, {}]
rule 55. (A3 = 1) & (A7 = 1) & (A11 = 1) => (Dec = 4); [3, 3, 2.07%, 100.00%][0, 0, 0, 3, 0]
[{} , {} , {} , {477, 508, 589}, {}]
rule 56. (A1 = 2) & (A3 = 2) & (A6 = 2) & (A7 = 2) & (A8 = 3) => (Dec = 4); [15, 15, 10.34%, 100.00%][0, 0, 0, 15, 0]
[{} , {} , {} , {20, 101, 217, 239, 258, 295, 312, 397, 457, 467, 489, 520, 523, 565, 686}, {}]
rule 57. (A1 = 1) & (A2 = 3) & (A3 = 1) => (Dec = 4); [10, 10, 6.90%, 100.00%][0, 0, 0, 10, 0]
[{} , {} , {} , {88, 152, 182, 230, 240, 408, 497, 649, 665, 687}, {}]
rule 58. (A2 = 3) & (A9 = 3) & (A11 = 1) => (Dec = 4); [8, 8, 5.52%, 100.00%][0, 0, 0, 8, 0]
[{} , {} , {} , {31, 51, 158, 437, 480, 494, 590, 640}, {}]
rule 59. (A1 = 2) & (A2 = 3) & (A3 = 1) & (A10 = 1) => (Dec = 4); [7, 7, 4.83%, 100.00%][0, 0, 0, 7, 0]
[{} , {} , {} , {37, 74, 369, 437, 569, 657, 673}, {}]
rule 60. (A1 = 3) & (A2 = 2) & (A3 = 2) & (A5 = 1) & (A6 = 2) & (A7 = 2) => (Dec = 5); [26, 26, 21.67%, 100.00%][0, 0, 0, 0, 26]
[{} , {} , {} , {} , {34, 38, 45, 56, 60, 72, 86, 104, 109, 151, 155, 190, 210, 255, 257, 270, 284, 324, 348, 386, 436, 509, 538, 606, 617, 666}]
rule 61. (A2 = 2) & (A3 = 2) & (A5 = 1) & (A6 = 2) & (A7 = 2) & (A9 = 2) & (A10 = 2) & (A11 = 2) => (Dec = 5); [18, 18, 15.00%, 100.00%][0, 0, 0, 0, 18]
[{} , {} , {} , {} , {12, 16, 45, 50, 73, 83, 86, 97, 109, 115, 121, 143, 161, 205, 255, 396, 512, 658}]
rule 62. (A1 = 1) & (A2 = 2) & (A3 = 2) & (A6 = 2) & (A7 = 2) & (A9 = 2) & (A10 = 2) => (Dec = 5); [13, 13, 10.83%, 100.00%][0, 0, 0, 0, 13]
[{} , {} , {} , {} , {24, 62, 118, 126, 134, 148, 224, 280, 356, 400, 543, 658, 675}]
rule 63. (A2 = 2) & (A3 = 2) & (A7 = 2) & (A9 = 1) => (Dec = 5); [5, 5, 4.17%, 100.00%][0, 0, 0, 0, 5]
[{} , {} , {} , {} , {461, 509, 516, 635, 679}]
rule 64. (A1 = 3) & (A2 = 2) & (A6 = 2) & (A9 = 2) & (A10 = 2) & (A11 = 2) => (Dec = 5); [6, 6, 5.00%, 100.00%][0, 0, 0, 0, 6]
[{} , {} , {} , {} , {1, 45, 86, 109, 255, 414}]
rule 65. (A1 = 2) & (A2 = 2) & (A7 = 3) & (A10 = 2) => (Dec = 5); [1, 1, 0.83%, 100.00%][0, 0, 0, 0, 1]
[{} , {} , {} , {} , {337}]
rule 66. (A2 = 2) & (A5 = 1) & (A6 = 2) & (A8 = 1) & (A11 = 1) => (Dec = 5); [2, 2, 1.67%, 100.00%][0, 0, 0, 0, 2]
[{} , {} , {} , {} , {104, 645}]
Approximate rules
rule 67. (A1 = 2) & (A2 = 2) & (A3 = 2) & (A5 = 1) & (A7 = 2) & (A8 = 2) & (A9 = 2) & (A11 = 1) => (Dec = 3)
OR (Dec = 5); [60, 60, 44.44%, 100.00%][0, 0, 15, 0, 45]
[{} , {} , {3, 21, 99, 139, 149, 275, 360, 399, 439, 448, 517, 571, 619, 644, 678}, {} , {18, 28, 30, 33, 36, 40, 48, 53, 63, 71, 84, 90, 103, 106, 107, 111, 128, 141, 166, 176, 180, 196, 227, 250, 256, 265, 272, 305, 306, 316, 327, 331, 350, 381, 383, 435, 456, 487, 496, 499, 533, 566, 612, 643, 669}]
rule 68. (A1 = 2) & (A2 = 2) & (A3 = 2) & (A5 = 1) & (A6 = 2) & (A8 = 2) & (A9 = 2) & (A11 = 1) => (Dec = 3)
OR (Dec = 5); [54, 54, 40.00%, 100.00%][0, 0, 9, 0, 45]
[{} , {} , {3, 282, 288, 296, 304, 375, 446, 478, 618}, {} , {18, 28, 30, 33, 36, 40, 48, 53, 63, 71, 84, 90, 103, 106, 107, 111, 128, 141, 166, 176, 180, 196, 227, 250, 256, 265, 272, 305, 306, 316, 327, 331, 350, 381, 383, 416, 435, 487, 496, 499, 533, 585, 643, 669, 688}]
rule 69. (A1 = 3) & (A7 = 1) & (A9 = 2) & (A11 = 1) => (Dec = 3) OR (Dec = 5); [23, 23, 17.04%, 100.00%][0, 0, 21, 0, 2]
[{} , {} , {66, 122, 168, 212, 220, 226, 245, 309, 323, 345, 352, 403, 444, 522, 536, 564, 573, 587, 629, 648, 671}, {} , {596, 603}]
rule 70. (A1 = 2) & (A6 = 1) & (A7 = 1) => (Dec = 3) OR (Dec = 5); [26, 26, 19.26%, 100.00%][0, 0, 21, 0, 5]
[{} , {} , {11, 39, 52, 57, 70, 113, 160, 183, 232, 234, 269, 302, 366, 368, 385, 410, 462, 464, 490, 604, 615}, {} , {291, 293, 578, 594, 652}]
rule 71. (A3 = 2) & (A6 = 1) & (A8 = 3) => (Dec = 3) OR (Dec = 5); [6, 6, 4.44%, 100.00%][0, 0, 3, 0, 3]
[{} , {} , {354, 410, 490}, {} , {14, 652, 663}]
rule 72. (A1 = 1) & (A2 = 2) & (A3 = 2) & (A6 = 1) & (A11 = 1) => (Dec = 3) OR (Dec = 5); [12, 12, 8.89%, 100.00%][0, 0, 10, 0, 2]
[{} , {} , {78, 138, 191, 200, 236, 301, 371, 417, 539, 623}, {} , {472, 493}]
rule 73. (A1 = 2) & (A2 = 1) & (A3 = 3) & (A4 = 1) & (A5 = 1) & (A8 = 2) & (A10 = 2) => (Dec = 1) OR (Dec = 2); [7, 7, 50.00%, 100.00%][1, 6, 0, 0, 0]
[{} , {} , {580}, {8, 147, 361, 390, 442, 654}, {} , {} , {}]
rule 74. (A2 = 3) & (A3 = 2) & (A4 = 1) & (A5 = 1) & (A7 = 2) & (A8 = 2) & (A9 = 2) & (A10 = 2) & (A11 = 1) => (Dec = 1) OR (Dec = 2); [5, 5, 35.71%, 100.00%][1, 4, 0, 0, 0]
[{} , {} , {314, 322, 363, 398}, {} , {} , {}]
rule 75. (A1 = 2) & (A2 = 1) & (A3 = 2) & (A8 = 1) => (Dec = 1) OR (Dec = 2); [2, 2, 14.29%, 100.00%][1, 1, 0, 0, 0]
[{} , {} , {} , {} , {342}, {55}, {} , {} , {}]

```

rule 76. (A1 = 2) & (A2 = 1) & (A3 = 2) & (A4 = 1) & (A5 = 1) & (A6 = 2) & (A7 = 2) & (A8 = 2) & (A9 = 2) &
(A10 = 2) & (A11 = 1) => (Dec = 1) OR (Dec = 2) OR (Dec = 4); [27, 27, 100.00%, 100.00%][3, 22, 0, 2, 0]
[{}100, 405, 641}, {27, 85, 146, 175, 179, 192, 201, 206, 207, 216, 221, 223, 262, 310, 318, 346, 379, 404,
465, 562, 595, 610}, {}, {79, 117}, {}]
rule 77. (A1 = 2) & (A2 = 2) & (A3 = 2) & (A5 = 1) & (A9 = 3) & (A10 = 2) & (A11 = 1) => (Dec = 4) OR (Dec =
5); [6, 6, 100.00%, 100.00%][0, 0, 0, 5, 1]
[{}}, {}, {}, {49, 135, 142, 473, 485}, {449}]
rule 78. (A1 = 2) & (A2 = 1) & (A3 = 2) & (A4 = 1) & (A5 = 1) & (A6 = 2) & (A9 = 2) & (A11 = 2) => (Dec = 2)
OR (Dec = 4); [13, 13, 56.52%, 100.00%][0, 10, 0, 3, 0]
[{}}, {92, 154, 249, 344, 359, 373, 451, 460, 471, 559}, {}, {127, 376, 433}, {}]
rule 79. (A1 = 2) & (A2 = 3) & (A3 = 1) & (A5 = 1) & (A7 = 2) & (A8 = 2) & (A10 = 2) & (A11 = 1) => (Dec = 2)
OR (Dec = 4); [5, 5, 21.74%, 100.00%][0, 1, 0, 4, 0]
[{}}, {157}, {}, {150, 194, 458, 557}, {}]
rule 80. (A3 = 1) & (A5 = 1) & (A9 = 2) & (A10 = 1) & (A11 = 2) => (Dec = 2) OR (Dec = 4); [3, 3, 13.04%,
100.00%][0, 2, 0, 1, 0]
[{}}, {131, 308}, {}, {93}, {}]
rule 81. (A2 = 2) & (A3 = 1) & (A5 = 1) & (A6 = 2) & (A8 = 2) & (A9 = 2) & (A10 = 2) & (A11 = 1) => (Dec = 2)
OR (Dec = 4); [2, 2, 8.70%, 100.00%][0, 1, 0, 1, 0]
[{}}, {534}, {}, {582}, {}]
rule 82. (A2 = 3) & (A3 = 2) & (A4 = 1) & (A5 = 2) & (A8 = 2) & (A11 = 1) => (Dec = 1) OR (Dec = 4); [9, 9,
100.00%, 100.00%][5, 0, 0, 4, 0]
[{}110, 382, 459, 504, 515}, {}, {}, {96, 145, 222, 510}, {}]
**END

```