

## Sınıflandırıcıların Kalp Hastalığı Verileri Üzerine Performans Karşılaştırması

Özlem ÖZMEN, Ahmad KHDR, Engin AVCI

Yazılım Mühendisliği Bölümü, Fırat Üniversitesi, 23000 ELAZIĞ  
e-mail: ozlmozmn23@gmail.com

(Geliş/Received: 23.03.2018; Kabul/Accepted: 03.09.2018)

### Özet

Bu çalışmada kalp hastalığı verileri kullanılarak bazı sınıflandırıcıların avantajları, dezavantajları göz önünde bulundurularak performans karşılaştırmaları yapılmıştır. Araştırmada kullanılan algoritmalar şunlardır: Destek Vektör Makinesi (DVM), Naive Bayes, J48, Random Forest, Adaboost, Logistic Regresyon, Tek Katmanlı Perceptron, Çok Katmanlı Perceptron, Bagging karar Ağaçları. Burada sonuçların karşılaştırılması için veri setindeki kayıt sayısı, doğruluk ortalaması, doğru olarak sınıflandırılmış örnekler, yanlış olarak sınıflandırılmış örnekler, kappa istatistiği, ortalama mutlak hata, ortalama kare hata, kök ortalama kare hata, göreceli mutlak hata, kök nispi kare hata gibi ölçütleri kullanıldı. Elde edilen sonuçlara göre en yüksek başarımlı DVM algoritması sonucunda bulunmuştur.

**Anahtar Kelimeler:** Kalp hastalığı, Makine öğrenimi, Algoritmaların karşılaştırmaları, Veri sınıflandırma Veri madenciliği.

## Comparison of Data Mining Classification Algorithms for Heart Disease Prediction

### Abstract

In this study, performance evaluations for classification algorithms been made according to the advantages and disadvantages by using cardiac disease dataset. The used algorithms in this study are Support Vector Machine (SVM), Naive Bayes, J48, Random Forest, Adaboost, Logistic Regression, Single Layer Perceptron, Multi Layer Perceptron, Bagging and Decision Trees. In order to compare the results, criteria such as the number of records in the data set, accuracy averages, correctly classified samples, incorrectly classified samples, kappa statistics, mean absolute error, mean square error, root mean square error, relative absolute error, root mean square error were used. According to the obtained results, the highest performance yield by SVM algorithm from the accuracy, kappa, etc point of view.

**Keywords:** Heart disease, Machine learning, Algorithms Comparison, Data classification, Data mining.

### 1. Giriş

Veri madenciliği, büyük veri tabanından veya veri kümelerinden bilgi edinme veya çıkarma süreci olan bir bilgisayar bilimi alanıdır. Başka bir ifadeyle gerçek zamanlı kullanıcılar ve yazılım çözümleri sağlayıcıları tarafından veri tabanlarında bilgi keşfi olarak ta adlandırılabilir. Günümüzde sağlık ve medikal sektörün veri madenciliğine olan ihtiyacı diğer sektörlerin ihtiyacından daha fazladır. Bu bağlamda veri madenciliği yöntemleri kullanılarak, büyük veri tabanlarından değerli bilgiler çıkarılabilir bu da tıbbi uygulayıcının karar almasına ve sağlık hizmetlerini iyileştirmesine yardımcı olabilir. Kalp hastalığı, kalbi, damarlarını, kaslarını,

valflerini veya kas kasılmasından sorumlu iç elektrik yollarını içeren bir dizi durumu ifade eder. Hastalık Kontrol Merkezlerine (HKM) göre, kalp hastalığı Birleşik Krallık, Amerika Birleşik Devletleri, Kanada ve Avustralya'da önde gelen ölüm nedenlerinden biridir. ABD'deki her dört ölümden biri, kalp hastalığının bir sonucu olarak ortaya çıkar [1].

Kalp hastalığı dünyadaki çoğu ülkede hem erkekler hem de kadınlar arasında yaygındır. Bu nedenle, insanlar kalp hastalığı risk faktörlerini göz önünde bulundurmalıdır. Genetik bir rol oynasa da, bazı yaşam tarzı faktörleri, kalp hastalığını önemli derecede etkilemektedir. Kalp hastalığı için bilinen risk faktörleri; yaş, cinsiyet, aile öyküsü, sigara içmek, bazı kemoterapi ilaçları ve kanser için radyasyon

tedavisi, kötü beslenme, yüksek tansiyon, yüksek kan kolesterol seviyeleri, diyabet, obezite, fiziksel hareketlilik, stress ve kötü hijyendir [2].

### 1.1. Literatür Çalışması

Literatürde kalp hastalığı sınıflandırmasına ait bazı çalışmalar bulunmaktadır. Bunlardan bir tanesi de Yuehjen ve arkadaşlarının kalp hastalığı sınıflandırması için hibrit akıllı modelleme şemaları üzerine olan araştırmadır. Geleneksel yaklaşımlar, kalp hastalığını sınıflandırmak için on üç risk faktörünü veya açıklayıcı değişkeni kullanır. Mevcut yaklaşımlardan ayrılan bu çalışma, farklı açıklayıcı değişken kümeleri elde etmek için yeni bir hibrit akıllı modelleme şeması önermektedir. Önerilen hibrid modelleri, kalp hastalığını etkili bir şekilde sınıflandırmaktadır.

Önerilen hibrid modeller, Lojistik Regresyon (LR), Çok Değişkenli Adaptif Regresyon (ÇDAR) ve Yapay Sinir Ağı (YSA) tekniklerinden oluşmaktadır. Önerilen sürecin başlangıç aşaması, açıklayıcı değişkenler kümesini azaltmak için LR, ÇDAR ve RS tekniklerinin kullanımını içerir. Kalan değişkenler daha sonra ikinci aşamada kullanılan YSA yöntemi için girdi olarak kullanılır. Önerilen hibrit modellerin gelişimini göstermek için gerçek bir kalp hastalığı veri seti kullanılmıştır. Modelleme sonuçları, önerilen hibrid programların kalp hastalığını etkili bir şekilde sınıflandırdığını ve tipik, tek aşamalı YSA'yı geride bıraktığını ortaya koymuştur [3].

Priyanka ve arkadaşlarının yaptığı kalp hastalıkları tahmininde veri madenciliği tekniklerinin kullanımına ilişkin çalışmada Naïve Bayes ve Karar ağacı algoritmalarının performansları karşılaştırılmış olup karar ağacı algoritması %98.03 ile %82.35 doğruluk oranı veren Naïve Bayes e göre çok daha başarılı sonuçlar vermiştir [4].

Yekkala ve diğerleri, sonuçları daha kesin olarak tahmin etmek için parçacık yöntemleri (Random Forest, AdaBoost and Bagged Tree) ile birlikte Parçacık Sürü Optimizasyonu (PSO) kullanmışlardır. Kalp Stalog veri kümesi 270 örnek ve 14 özneliğe sahip olup UCI veritabanından alınmıştır [5].

Veriler önceden işlenmiş ve içerisindeki gereksiz ve eksik verileri silmek için bir özellik

seçim yöntemi olarak PSO'yu kullanılmıştır. Ayrıca, etkili özellikler çeşitli performans ölçüleri ve aşağıdaki gibi adımlar için topluluk sınıflandırıcı üzerinde test edilmiştir.

Öncelikle, veri seti yüklendikten sonra, veri üzerine PSO tarafından özellik seçimi uygulandıktan sonra, yetersiz özelliklerin ortadan kaldırılması ve etkin özelliklerin devam ettirilmesinin ardından veriler üzerine uygulanan temizleme tekniği, daha sonra AdaBoost, Bagging ve Random Forest sınıflandırıcıları etkin özelliklere uygulandı. Sonunda her algoritmanın performansı ölçüldü. Sonuç olarak Bagged Tree %100, Random Forest %90.37, AdaBoost ise %88.89 performans göstermiştir.

Test sonuçlarına göre, Yekkal ve arkadaşları PSO üzerinde Bagging Trees kullanmanın kalp hastalığını tahmin etmede öğrenme doğruluğunu artıracığını kanıtladı.

Amin ve arkadaşları, yaş, aile öyküsü, diyabet, hipertansiyon, yüksek kolesterol, tütün kullanımı, alkol alımı, obezite veya fiziksel inaktivite gibi başlıca risk faktörlerini temel alan kalp hastalıklarını öngörmektedir. Teknikleri en başarılı iki veri madenciliği aracını, sinir ağları ve genetik algoritmaları içeriyordu. Bu nedenle kalp hastalıklarının öngörülmesi ve tanısı için Naïve Bayes, Yapay Sinir Ağları, Karar Ağacı, Bagging Trees, Çekirdek Yoğunluğu ve DVM uygulamışlardır. Öğrenme, geri yayılma ile karşılaştırıldığında hızlı, daha istikrarlı ve doğru olarak bulunmuştur. Sistem kalp hastalığı riskini tahmin etmek üzere Matlab'da uygulanmıştır [6].

50 hastanın risk faktörleri verileri toplanmış ve hibrid model %96, 2 eğitim doğruluğu ve %89 test doğruluk ile sonuçlanmıştır. Daha sonra Amin ve arkadaşları, kalp hastalıklarının öngörülmesi için hibrit bulanık ve k-en yakın komşu yaklaşımını kullanarak sistemi geliştirdiler; diğer bir sistemde ise nöral ağ topluluğu, kalp hastalığının tanısında %89.01'lik bir doğrulukla kullanılmıştır. Bu hibrit sistemin avantajı, hastaların kalp hastalıkları ve yan etkilerinden önceki tıbbi muayeneler için maliyet ve zamanlarını azaltmalarına ve kendilerini kontrol etmelerine yardımcı olmaktır. Yazar yukarıdaki algoritmaları Güney Afrikadaki pratisyen hekimlerden topladığı hasta verilerine uygulamıştır. Veri tabanından sadece 11 öznelik, kalp hastalığını öngörmek için

ayrıydı. Ek olarak, kullanılan özellikler aşağıdaki gibidir; hasta kimlik numarası, cinsiyet, kardiyogram, yaş, göğüs ağrısı, kan basıncı düzeyi, kalp atım hızı, kolesterol, sigara kullanımı, alkol tüketimi ve kan şekeri seviyesi.

Araştırmacılar kullanılan algoritmalar arasında karışıklık matrisine göre karşılaştırma yaptılar. Sonunda, J48'in %99 ile en yüksek doğruluğu kaydettiği anlaşıldı.

## 2. Malzemeler ve Yöntemler

Karşılaştırma çalışmamız için UCI' den almış olduğumuz "Heart Disease Data Set" isimli veri setini kullandık [9]. Verisetimiz 303 kayıt ve 75 özellik içermekte olup ancak sadece 13 özellik kullanıldı ve bunlar: yaş, cinsiyet, göğüs ağrısı tipi, dinlenme tansiyonu, kolesterol, açlık kan şekeri, istirahat elektrokardiyografik sonuçları, elde edilen maksimum kalp atım hızı, egzersize bağlı anjina, istirahate göre egzersizle indüklenen depresyon, pik egzersizin eğimi, majör damar sayısı ve "num" olarak adlandırılan kalp hastalığı tanısı veri kümesidir. Her bir algoritma için önce iki farklı test seçeneğini denedik; ilki 10 kat çapraz geçerlilik, ikincisi ise; veri setini eğitim seti için %75, test seti için %25 paylaştırarak kullanmak oldu. Bunun arkasındaki sebep ise hangi algoritmanın en yüksek sonucu kaydettiğinden emin olmak için aralarında karşılaştırma yapmaktır. Bu çalışmada açık kaynaklı bir yazılım olan WEKA yardımıyla on algoritmanın karşılaştırmasını yaptık. Verilerin analizi için kullandığımız WEKA yazılımı sınıflandırma, kümeleme, özellik seçimi, veri görselleştirme vb. denetimli ve denetimsiz öğrenme yöntemlerini içerdiği için güçlü bir araçtır. WEKA kullanılmasının temel nedeni, bizim gibi araştırmacıların bu yazılım sayesinde veri madenciliği tekniklerini gerçek veya sentetik veriler üzerinde çok kolay bir şekilde karşılaştırabilme imkanını bulmasıdır [7].

## 3. Algoritmaların Karşılaştırılması

**Tablo 1:** Destek vektör makinesinin avantaj ve dezavantajları

DVM [8] [9]	
Avantajlar	Dezavantajlar
Veriler hakkında hiçbir fikrimiz yokken DVM çok iyi sonuç verir.	İyi bir çekirdek fonksiyonu seçmek kolay değildir.
Yarı yapılandırılmış ve yapılandırılmamış verilerle SVM iyi çalışır.	Büyük veri setleri uzun eğitim süreleri alır.
SVM'nin güç noktası onun çekirdekleridir.	Son modeli, bireysel etki ve değişken ağırlıkları anlamak ve yorumlamak zordur.
Kullanışlı çekirdek fonksiyonunun doğru seçimi ile, karmaşık bir sorunu çözebiliriz.	Son modelin görülmesi zordur; bu nedenle, modele küçük standartizasyon yapılamıyor.
Yüksek boyutlu veriler için, nispeten iyi ölçeklendirir.	
SVM'de aşırı takılma riski daha azdır.	
Doğrusal ayırma ile DVM, lojistik regresyon gibi performans gösterir.	
Lineer olmayan sınırla, DVM kullanılmış çekirdeğe göre iyi performans gösterir.	
Çok boyutlu verilerle uğraşmak çok iyi sonuç sağlar.	

**Tablo 2:** J48 Karar Ağacı avantaj ve dezavantajları

J48 Karar Ağacı [11]	
Avantajlar	Dezavantajlar
Karar ağacı ya değişkenleri tarar veya özellikleri seçer.	Birçok dalı olan karar ağaçları hazırlamak karmaşık ve zaman kaybedicidir.
Veri hazırlama için kullanıcılardan çok az çalışma gerektirir.	Dikdörtgen sınıflandırma kutularına yol açan, her seferinde sadece tek bir alanı test eder.
Ağaç performansı, parametreler arasında varolan doğrusal olmayan ilişkilerden etkilenmeyecektir.	Bu, karar alanındaki gerçek kayıt dağılımı ile iyi bir korelasyon gösteremeyebilir.
Analitik için ağaçları kullanmanın en iyi özelliği, yorumlama ve yöneticilere açıklama kolaylığıdır.	Çizimler sunarken pratik olmayabilir.
	Verileri sınıflara ayırmaya ve sadece basit eşik testlerini kullanmasına rağmen, büyük bir karar ağacı gerektirir.
	Büyük ağaçlar anlaşılabilir değildir ve sunumda zorluklarla karşılaşılır.

**Tablo 3:** Naïve Bayes avantaj ve dezavantajları

NAİVE BAYES [10] [11]	
Avantajlar	Avantajlar
Hızlıdır, yüksek sonuç verir ve esnek bir modeldir. Doğrusal olarak, satır ve belirleyicilerin sayısı ile ölçülür. Yapı sürecinin paralelleştirilmesi. Çoklu sınıf ve ikili sınıflandırma problemlerini çözmek için kullanılabilir. Yüksek boyutlarda iyi çalışır	Hızlıdır, yüksek sonuç verir ve esnek bir modeldir. Doğrusal olarak, satır ve belirleyicilerin sayısı ile ölçülür. Yapı sürecinin paralelleştirilmesi. Çoklu sınıf ve ikili sınıflandırma problemlerini çözmek için kullanılabilir. Yüksek boyutlarda iyi çalışır

**Tablo 4:** Random Forest avantaj ve dezavantajları

RANDOM FOREST [12] [13]	
Avantajlar	Dezavantajlar
Karar ağacı korelasyonlu olabilecek birden fazla özellik ile uğraşırken önemlidir. Sonuç olarak azalan varyans, neredeyse tüm durumlarda, Random Forest karar ağaçlarından daha doğrudur.	Aşırı takılmanın azalması: Birkaç ağacın ortalaması ile, aşırı takılma riski çok düşüktür. Daha az varyans: Birden fazla ağaç kullanarak, eğitim ve test verileri arasındaki ilişki nedeniyle iyi performans göstermeyen bir sınıflandırıcıya rastlanma şansını azaltırsınız. Görsel olarak yorumlamak kadar kolay değil

**Tablo 5:** AdaBoost avantaj ve dezavantajları

ADABOOST [14] [15]	
Avantajlar	Dezavantajlar
Her ağacın büyüklüğünü tanımlayabilen kullanıcılara göre, AdaBoost'u Bagging Trees ağaçlarından, Random Forest'dan daha fazla açıklanabilir hale getirir. Belirli özellikleri kolayca ele alabilir (kategorik özellikler)	Bagging Trees ve Random Forest'dan farklı olarak, ağaç sayısı çok büyükse aşırıya kaçabilir.

**Tablo 6:** Lojistik Regresyon avantaj ve dezavantajları

LOJİSTİK REGRESYON [16]	
Avantajları	Dezavantajları
Düşük varyans. Sonuçlar için olasılıklar sağlar. Çapraz karar sınırları ile iyi çalışır. Lojistik regresyon, çekirdek yöntemleri ile de kullanılabilir	Yüksek bias: çözüm alanı üzerinde oldukça güçlü kısıtlamalar uygular ve böylece sapma yoluyla varyans azaltma yoluyla daha fazla hata ekleyebilir.

**Tablo 7:** Tek Katmanlı Perceptron Karar Ağacı avantaj ve dezavantajları

TEK KATMANLI PERCEPTRON	
Avantajlar	Dezavantajlar
Hesaplama zamanı çok kısadır.	Tek katman sadece basit problemler için kullanılabilir.

**Tablo 8:** Çok Katmanlı Perceptron avantaj ve dezavantajları

ÇOK KATMANLI PERCEPTRON [17]	
Avantajlar	Dezavantajlar
Çok katmanlı, sinir ağlarının çoğu derin öğrenmeyi eklemektedir. Bir veya iki gizli katman kullanır. Zor, karmaşık problemler için kullanılabilir. Kompaktlık: çok daha küçük modellerle aynı doğruluğu elde etme potansiyeli	Bazen uzun eğitim zamanı gereklidir. Optimizasyon zordur. Pek çok çözüm imkanı vardır. Doğrusal olmayan çıktıda gizli ağırlıkların etkisi.

**Tablo 9:** Bagging Ağaçları avantaj ve dezavantajları

BAGGING AĞAÇLARI [15]	
Avantajlar	Dezavantajlar
Düzenli karar ağaçlarına göre azalan farklılıklar. Değişken öneme sahip tedbirler. Nitel özelliklerin kolayca ele alınması. Torba dışı, model doğrulama için tahminler kullanılabilir.	Görsel olarak açıklamak zordur. Özellikler ilişkiliyse, varyansı azaltamayacaktır.

**Tablo 10:** Karar tablosu avantaj ve dezavantajları

KARAR TABLOSU [13]	
Avantajlar	Dezavantajlar
Ağaçlar sadece birkaç seviye içerdiğinde görsel olarak açıklama kolaylığı. Belirli özellikleri kullanma. Karar sınırları ile iyi çalışma Özelliğ eksenine paralel. Sadece açıklanabilir ve görselleştirilebilir. İç işleyişi gözlemlemek bu nedenle işi yeniden üretmek mümkündür. Hem sayısal hem de kategorik verilerle uğraşmak. Büyük veri kümelerinde iyi performans gösterir ve çok hızlıdır.	Bir ağaç özellikle belirli bir derinlikte olduğunda, aşırı donatmaya maruz kalır. Çapraz karar sınırları ile ilgili sorunlar.

#### 4. Sonuçlar

Her bir algoritma için iki farklı test seçeneğini denedik; birincisi 10-kat çapraz

geçerlilik ve ikincisi veri kümesini %75 eğitim ve %25 test setine bölüyor.

**Tablo 11:** Tahminlerin Karşılaştırması

Değerlendirme Kriterler	Kappa istatistiği	Ortalama mutlak hata	Karekök ortalama hata	Göreceli mutlak hata	Kök göreceli karesi hatası
DVM	0.7883	0.105	0.324	21.1645	64.904
Naïve Bayes	0.6306	0.179	0.351	35.927	69.997
J48	0.6069	0.239	0.409	48.047	81.812
Random Forest	0.6842	0.251	0.329	50.505	65.763
Adaboost	0.7608	0.202	0.315	40.619	63.069
Lojistik	0.7093	0.201	0.345	40.324	68.726
Tek katmanlı Perceptron	0.2184	0.380	0.617	76.594	123.37
Çok katmanlı perceptron	0.7093	0.141	0.348	28.468	69.602
Bagging Ağaçları	0.7361	0.255	0.323	51.385	64.617
Karar tablosu	0.5483	0.315	0.399	63.453	79.939

Tablo 1'de Kappa, ortalama mutlak hata, kök ortalama karesi hatası, göreceli mutlak hata, kök göreceli karesel hataya göre sonucu görebiliriz.

$$K = \frac{P_0 - P_c}{1 - P_c} \quad (1)$$

P<sub>0</sub>, toplam anlaşma olasılığı ve P<sub>c</sub>, şansa bağlı “anlaşma” olasılığıdır. 2 ve 3 numaralı denklemleri kullanarak bulabilecekleri yerler.

$$P_0 = \frac{T_p}{100} + \frac{T_n}{100} \quad (2)$$

DVM en yüksek Kappa'yı kaydetti, kappa, gözlemciler arasında [18] denklem 1 kullanılarak bulunabilecek anlaşma ve anlaşmazlıktır:

$$P_c = \frac{F_p}{100} \times \frac{F_n}{100} \quad (3)$$

Buna ek olarak, tablo 2 duyarlılık, özgünlük ve doğruluğa göre kullanılan her algoritmanın sonuçlarını göstermektedir.

**Tablo 11:** Algoritmaların Performansları

Algoritmalar	Duyarlılık	Özgünlük	Pozitif öngörme değeri	Olumsuz Tahmin Değeri	Doğruluk
DVM	0.925	0.861	0.881	0.912	%89.4737
Naïve Bayes	0.825	0.806	0.825	0.806	%81.5789
J48	0.750	0.861	0.857	0.756	%80.2632
Random Forest	0.825	0.861	0.868	0.816	%84.2105
Adaboost	0.950	0.806	0.844	0.935	%88.1579
Lojistik	0.875	0.833	0.854	0.857	%85.5263
Tek katmanlı Perceptron	0.825	0.389	0.600	0.667	%61.8421
Çok katmanlı perceptron	0.875	0.833	0.854	0.857	%85.5263
Bagging Ağaçları	0.875	0.861	0.875	0.861	%86.8421
Karar Tablosu	0.850	0.694	0.756	0.807	%77.6316

Aşağıda gösterilen formüller duyarlılığı, özgünlüğü ve doğruluğu hesaplamak için kullanılır.

$$\text{Duyarlılık} = \frac{T_p}{T_p + F_n} \quad (4)$$

$$\text{Özgünlük} = \frac{T_n}{T_n + F_p} \quad (5)$$

$$\text{Doğruluk} = \frac{T_p + T_n}{T_p + F_p + T_n + F_n} \quad (6)$$

Tablo 1 ve 2'de DVM, diğer algoritmalar arasında en yüksek sonuç olan %89.4737 doğruluğa sahiptir ve Kappa'da da 0.7883'ü kaydetmiştir, bu da diğer algoritmalarla göre en yüksek sonuç olmuştur. Açıkçası, sonuçlar, veri madenciliği algoritmalarının, teşhisler için bir sınıfı öngörebildiğini göstermiştir.

## 5. Tartışma

Makine öğrenmesi metotları genellikle en iyi performans için optimize edilmesi gereken birkaç parametreyi içerdiğinden ve verileri manuel olarak optimize etmek te yorucu olmasının yanısıra aşırı yüklemeye karşılaşılabileceğinden tehlikelidir. Bu nedenle, üç adet weka-güçlendiriciden birini kullanabiliriz, ki bunlardan ilki iç çapraz doğrulama kullanarak en iyi performans için parametreleri optimize eden "metalearners" dir. İkincisi, olan "CVParameterSelection" bir parametre için en iyi değeri seçer; "GridSearch" iki boyutlu ızgarayı arayarak iki parametreyi optimize eder; ve sonuncusu olan "ThresholdSelector" ise bir olasılık eşiği seçer.

## 6. Referanslar

1. Felman, A., Everything you need to know about heart disease. Medical News Today, 2018.
2. Staff, M.C., Heart disease. Mayo Clinic.
3. Shao, Y.E., C.-D. Hou, and C.-C. Chiu, Hybrid intelligent modeling schemes for heart disease classification. Applied Soft Computing, 2014. **14**: p. 47-52.
4. Priyanka, N. and P. RaviKumar. Usage of data mining techniques in predicting the heart diseases—Naïve Bayes & decision tree. in Circuit, Power and Computing Technologies (ICCPCT), 2017 International Conference on. 2017. IEEE.
5. Yekkala, I., S. Dixit, and M. Jabbar. Prediction of heart disease using ensemble learning and Particle Swarm Optimization. in Smart Technologies For Smart Nation (SmartTechCon), 2017 International Conference On. 2017. IEEE.
6. Amin, S.U., K. Agarwal, and R. Beg. Genetic neural network based data mining in prediction of heart disease using risk factors. in Information & Communication Technologies (ICT), 2013 IEEE Conference on. 2013. IEEE.
7. Shah, C. and A.G. Jivani. Comparison of data mining classification algorithms for breast cancer prediction. in Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on. 2013. IEEE.
8. Fedorovici, L.-O. and F. Dragan. A comparison between a neural network and a SVM and Zernike moments based blob recognition modules. in Applied Computational Intelligence and Informatics (SACI), 2011 6th IEEE International Symposium on. 2011. IEEE.
9. Yu, W.M., T. Du, and K.B. Lim. Comparison of the support vector machine and relevant vector machine in regression and classification problems. in Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004 8th. 2004. IEEE.
10. McCallum, A. and K. Nigam. A comparison of event models for naive bayes text classification. in AAAI-98 workshop on learning for text categorization. 1998. Citeseer.
11. Patil, T.R. and S. Sherekar, Performance analysis of Naive Bayes and J48 classification algorithm for data classification. International journal of computer science and applications, 2013. **6**(2): p. 256-261.
12. Revathi, S. and A. Malathi, Comparison between J48 And Random Forest Decision Tree Algorithms To Detect Probe Attack.
13. Esmaily, H., et al., A Comparison between Decision Tree and Random Forest in Determining the Risk Factors Associated with Type 2 Diabetes. Journal of research in health sciences, 2018. **18**(2).
14. Randhawa, K., et al., Credit card fraud detection using AdaBoost and majority voting. IEEE ACCESS, 2018. **6**: p. 14277-14284.
15. Budhani, S.K., C. Jha, and A. Ahmad, Comparative Study of Meta Classification Algorithm: Bagging, AdaboostM1 and Stacking with Concept Drift based Synthetic Dataset Hyperplane1 and Hyperplane2. International Journal of Engineering Science, 2018. **15927**.
16. Colombet, I., et al. Models to predict cardiovascular risk: comparison of CART, multilayer perceptron and logistic regression. in Proceedings of

the AMIA Symposium. 2000. American Medical Informatics Association.

- 17.** Mirjalili, S., Evolutionary Multi-layer Perceptron, in Evolutionary Algorithms and Neural Networks. 2019, Springer. p. 87-104.

**18.** Ben-David, A., Comparison of classification accuracy using Cohen's Weighted Kappa. Expert Systems with Applications, 2008. **34**(2): p. 825-832.