Araştırma Makalesi
Zeki Sistemler Teori ve Uygulamaları Dergisi 4(1) (2021) 55-64
DOI: 10.38016/jista.877292

Makine Öğrenmesi Algoritmaları Kullanarak Erken Dönemde Diyabet Hastalığı Riskinin Araştırılması

Gürkan BİLGİN[1]*

[1]Burdur Mehmet Akif Üniversitesi, Elektrik Elektronik Mühendisliği Bölümü, Burdur, Türkiye

Öz

Diyabet insanoğlunun yaşam kalitesini önemli derecede etkileyen, dünyada ve Türkiye'de görülme sıklığı giderek artan önemli bir hastalıktır. Özellikle sinir sistemi, böbrek, kalp, gözler, uzuvlar ve kan damarlarının tahribatına yol açmakta ve önemli kayıplara sebebiyet verebilmektedir. Bu sebeple diyabetin önlenebilmesi veya vereceği tahribatın en aza indirilebilmesi için erken tanısı ve takibi büyük önem kazanmaktadır. Makine öğrenme algoritmaları ile elde edilen sınıflandırma teknikleri, hastalığın risk tahmin modeli için araştırmacılar tarafından önemli olarak kabul görmüştür. Çalışmada, diyabete yakalanma olasılığını tahmin etmek için, 520 denekten alınan bilgiler ile oluşturulmuş olan bir veri tabanı kullanılmıştır. Çalışmada, makine öğrenmesi metotları olarak Çok Katmanlı Algılayıcı Yapay Sinir Ağları (ÇKAYSA), Destek Vektör Makinaları (DVM), Karar Ağaçları (KA), Topluluk Öğrenme Algoritmaları (TÖA), Doğrusal Ayrımcı Analizi (DAA), k-NN Metotları kullanılmıştır. Bu metotlar arasında en yüksek doğruluğu k-NN algoritması sağlamış ve bu algoritma ile %99,81 doğruluk elde edilmiştir. En yüksek doğruluk değeri sağlayan algoritmanın çalışma kapsamında geliştirilmiş olan bir bilgisayar kullanıcı arayüzü içerisine dâhil edilmesiyle bir diyabet erken tanı kiti geliştirilmiştir.

Anahtar kelimeler: Diyabet riski, Erken dönem, Destek Vektör Makinesi, Topluluk Öğrenme Algoritmaları, Karar ağaçları.

Investigation of The Risk of Diabetes in Early Period using Machine Learning Algorithms

Abstract

Diabetes significantly affecting the quality of human life, the world and the incidence of a disease in Turkey is increasingly important. In particular, it causes damage to the nervous system, kidney, heart, eyes, limbs and blood vessels and can cause significant losses. For this reason, early diagnosis and follow-up is of great importance in order to prevent diabetes or to minimize the damage it will cause. Classification techniques obtained by machine learning algorithms have been accepted as important by researchers for the risk prediction model of the disease. In the study, a database created with information from 520 subjects was used to estimate the probability of developing diabetes. In the study, Multilayer Perceptron Artificial Neural Networks (MLPNN), Support Vector Machines (SVM), Decision Trees (DT), Ensemble Learning Algorithms (ELA), Linear Discriminant Analysis (LDA), k-NN Methods were used as machine learning methods. Among these methods, k-NN algorithm provided the highest accuracy and 99,81% accuracy was achieved with this algorithm. A diabetes early diagnosis kit was developed by including the algorithm providing the highest accuracy value into a computer user interface developed within the scope of the study.

Keywords: Diabetes risk, Early stage, Support Vector Machine, Ensembles Learning Algorithms, Decision Trees.

1. Giriş (Introduction)

İnsan vücudundaki insülin hormonunun gereğinden az ya da yetersiz üretilmesi ve üretilen insülinin etkili kullanılamaması gibi durumlarda kronik bir rahatsızlık durumu ortaya çıkar. Bu rahatsızlık Diyabet (Şeker Hastalığı) olarak adlandırılır. Diyabet insanoğlunun yaşam kalitesini önemli derecede etkileyen, dünyada ve Türkiye'de görülme sıklığı giderek artan önemli bir hastalıktır (TEMD, 2013; TSHGM, 2011). Diyabet hastalığı organ kayıplarına ve ölümlere sebep olabilmektedir. Özellikle sinir sistemi, böbrek, kalp, gözler, uzuvlar ve kan damarlarının tahribatına yol

* Sorumlu yazar. Gürkan Bilgin
E-posta adresi: gbilgin@mehmetakif.edu.tr

Alındı : 9 Şubat 2021
Kabul : 7 Mart 2021

açmakta ve önemli kayıplara sebebiyet verebilmektedir (TEMD, 2013; Harding vd., 2019; Islam vd., 2020). Bu sebeple diyabetin önlenebilmesi veya vereceği tahribatın en aza indirilebilmesi için erken tanısı ve takibi büyük önem kazanmaktadır.

Genellikle tip 1 ve tip 2 diyabet olarak adlandırılan iki tür diyabet biliyoruz. Tip 1 diyabet, bağışıklık sistemi yanlışlıkla pankreas beta hücrelerine saldırdığında ve vücuda çok az insülin salındığında veya bazen vücuda hiç insülin salınmadığında ortaya çıkar. Öte yandan tip 2 diyabet, vücudumuz uygun insülin üretmediğinde veya vücut insüline dirençli hale geldiğinde ortaya çıkar. Bazı araştırmacılar diyabeti tip 1, tip 2 ve gestasyonel diyabet olarak ikiye ayırdı (The 6 Different Types of Diabetes, 2018). Diyabetin yaygın semptomları poliüri, polidipsi, polifaji, ani kilo kaybı (genellikle tip 1), halsizlik, obezite (genellikle tip 2), gecikmiş iyileşme, görsel bulanıklık, kaşıntı, sinirlilik, genital pamukçuk, kısmi parezi, kas sertliği ve alopesidir. vb. (The 6 Different Types of Diabetes, 2018; Statistics About Diabetes, 2018).

Diğer taraftan, Diabetes Australia, Harris vd. tarafından daha önce 12 yıla kadar belirtildiği gibi, diyabetin klinik tanıdan 7 yıl öncesine kadar var olabileceğini yayınlamışlardır (Diabetes Australia, 2018; Harris vd., 2018). Bununla beraber, diabetes Australia'ya göre, Tip 2'nin erken teşhisinde başarısız olunması Avustralya sağlık sistemine her yıl 700 milyon dolardan fazlaya mal olabilmektedir (Diabetes Australia, 2018). Düşük ve orta gelirli ülkeler, yaygınlığı endişe verici bir oranda artan diyabet gibi bu kadar maliyetli bir hastalığı yönetmenin yükünü kaldıramaz. Bu nedenle, erken teşhis ve uygun terapötik tedavinin başlatılması, hasta sonuçlarında çok önemli bir rol oynayabilir ve gayri safi ulusal harcamaları ve üretim kaybını azaltabilir.

Dolayısıyla, diyabetin erken tanısının konulabilmesi ve gerekli önlemlerin alınabilmesini, laboratuvar ortamlarında alınan kan örneklerinin, hekimlerin kontrol ve takibinde değerlendirilmesi ile gerçekleştirilmektedir. Fakat kan örneklerinin yanı sıra diyabet hastalığının belirtisi olan farklı semptomlar da mevcuttur. Hastalardan elde edilen bu semptomların dijital ortamlarda işlenerek değerlendirilmesi için veri setleri oluşturulmaktadır. Oluşturulan veri setlerinin erken tanı için kullanılabilmesinde dijital ortamdaki sınıflandırma yaklaşımları bir hayli önem kazanmaktadır. Veri madenciliği, tahmin için kullanılan önemli bir bilgisayar bilimi alanıdır. Veri analizi yoluyla, önceden bilinen verilerden yeni verilerin keşfedilmesi sürecidir (George, vd., 2015). Günümüzde birçok gelişmiş ve karmaşık veri seti analizlerine olanak sağlayan sistemler, algoritmalar ve metodolojiler makine öğrenmesinin tıbbi alandaki uygulamalarına olanak sağlamaktadır (Islam vd., 2020; Goldenberg ve Punthakee, 2013; Kononenko, 2001). Teknolojinin sağladığı depolanmış büyük miktardaki akıllı verilerin işlenmesinde makine öğrenmesi tekniklerinin kısa zaman içinde yaygın olarak kullanılacağı düşünülmektedir (Ogurtsova vd., 2017).

Gerçekleştirilen bu çalışmada, hekimler tarafından erken tanısı konulmaya çalışılan söz konusu diyabet durumu erken dönemde tespit edilmeye çalışılmış, elde edilen veri setleri farklı makine öğrenme metotlarına uygulanarak doğru sınıflandırmanın en yüksek oranda elde edilmesi hedeflenmiş ve metotların karşılaştırmaları yapılmıştır. Çalışmada, makine öğrenmesi metotları olarak Çok Katmanlı Algılayıcı Yapay Sinir Ağları (ÇKAYSA), Destek Vekör Makinaları (DVM), Karar Ağaçları (KA), Topluluk Öğrenme Algoritmaları (TÖA), Doğrusal Ayrımcı Analizi (DAA), k-NN Metotları kullanılmıştır. En yüksek doğruluk değeri sağlayan algoritmanın çalışma kapsamında geliştirilmiş olan bir bilgisayar kullanıcı arayüzü içerisine dâhil edilmesiyle bir diyabet erken tanı kiti geliştirilmesi planlanmıştır.

Sağlık sektöründe gelişen teknoloji ile birçok çalışmada makine öğrenme algoritmaları kullanılarak farklı sağlık problemlerine çözüm üretmektedir. Herhangi bir sınıflandırma problemine çözüm üretmek için geleneksel algoritmaların zor olduğu veya mümkün olmadığı pek çok durumda makine öğrenmesi belirleyici olmuştur (Özer, 2020). Makine öğrenmesinin en önemli özelliklerinden birisi doğrusal olmayan ve çok karmaşık veri setlerini kullanarak, kararlı ve performansı yüksek tahminler yapabilmesidir (Hastie vd., 2009). Literatürdeki tıbbi tanı ile ilgili yapılan çalışmalardan birisinde, sınıflandırmanın tıp alanındaki tanı pratiğine rehberlik ettiğini ifade edilmiştir (Jutel, 2011). Bir başka çalışmada hastaları risk gruplarına göre ayırabilmek için hasta verileri kullanılmış ve analizlerin gerçekleştirilmesi için sınıflandırma modeli önerilmiştir (Parikh vd., 2016). Kullanılan yüksek boyutlu veri setlerinin güvenirliliğinin ve etkinliğinin artırılarak işlevsel hale getirilmesinde yapay zekâ metodları etkin rol oynamaktadır (Cichosz vd., 2015; Tran vd., 2019; Zhong vd., 2019; Bishop 2006). Örneğin bir akciğer kanseri alt tiplerinin belirlenmesinde nefes örnekleri kullanılmış ve sınıflandırma için k-NN algoritması kullanılmıştır. Sonuçta yüksek başarı oranlarına ulaşılmıştır (Wang vd., 2020).

Diyabet hastalığının erken tanısı için makine öğrenmesi ile sınıflandırma yapmak yüksek doğrulukla sonuçlar vermektedir. Sınıflandırmaların yapılabilmesi için çok farklı makine öğrenmesi algoritmaları kullanılmaktadır. Literatürdeki farklı diyabet hastalığının tanısı için yapılan çalışmalara bakılırsa, 25-78 yaş aralığındaki 250 adet hastadan elde edilen veri seti, Sapon vd. tarafından yapılan bir çalışmada Bayesian Regülasyon algoritmasında kullanılmıştır ve %88,8 oranında bir doğruluk başarısına ulaşılmıştır (Sapon vd., 2011). 2013 yılında gerçekleştirilen bir başka diyabet tahmini çalışmasında farklı metodolojiler kullanılıp karşılaştırmalar yapılmıştır. Karşılaştırmalarda kısmi En Küçük Kareler yöntemi (EKK)- DAA kombinasyonu %74 ile en yüksek doğrulukta tahmin değeri üretmiştir (Karthikeyani ve

Begum, 2013). Parashar vd. DDA-DVM kombinasyonu ve İleri Beslemeli Sinir Ağları (İBSA) sınıflandırma teknikleri kullanılarak karşılaştırmalar yapılmıştır. DVM kombinasyonu doğruluğu %75 değerine ulaşmıştır (Parashar vd., 2014). Ahmed tarafından yapılan çalışmada insülin, ilaç tedavisi ve diyet gibi diyabetik tedavi planlarını sınıflandırmak için yeni bir model geliştirilmiştir. 318 medikal kayıt ile gerçekleştirilen çalışmada 9 farklı öznitelik kullanılmıştır. Çalışma sonunda geliştirilen J48 algoritması ile %70,8 başarım değeri elde edilmiştir (Ahmed, 2016). Mercaldo vd. diyabet teşhisine yardımcı olmak ve hızlandırmak için, Dünya Sağlık Örgütü kriterlerine göre seçilmiş bir dizi özellik kullanılarak diyabetten etkilenen hastaları sınıflandırabilen bir yöntem önermiştir (Mercaldo vd., 2017). Mercaldo vd. çalışmalarında diyabet tanısı ve sınıflandırılma için Hoeffding Tree algoritmasını kullanmışlardır. Diyabet tanısı konusunda klinik veriler kullanılarak gerçekleştirilen bir başka çalışmada, diyabetin tahmini, komplikasyonları, genetik arka plan ve çevre açısından sistematik bir inceleme yapılmıştır. Çeşitli makine öğrenme algoritmaları kullanılmış, DVM en başarılı ve yaygın olarak kullanılan algoritma olarak ortaya çıkmıştır (Kavakiotis vd., 2017). 2018 yılında Joshi ve Chawan tarafından yapılan çalışmada glikoz, yaş, kan basıncı ve vücut kitle indeksi gibi 7 farklı öznitelik kullanılarak diyabet tahmini gerçekleştirilmiştir. Analizler için makine öğrenmesi algoritması olarak DVM, lojistik Regresyon ve YSA kullanmışlardır. Sınıflandırma performanslarına bakıldığında en iyi sonuçlar DVM ile elde etmişlerdir (Joshi ve Chawan, 2018). Aynı yıl yapılan bir başka Kaur ve Kumari Ulusal Diyabet, Sindirim ve Böbrek Hastalıkları Enstitüsüne ait Pima Hindistan nüfuslu en az 25 kadın hastanın oluşturduğu veri kümesi ile çalışmıştır. Çalışmada Gözetimli (Supervised) Makine Öğrenmesi algoritmaları ile tahmin çalışmaları yapmışlardır (Kaur ve Kumari, 2018). Al Helal vd. diyabet tanısında farklı sınıflandırma algoritmaları kullanmışlar ve sonuçlarını k-NN ile %66,19, Naive Bayes ile %72,66 ve Rastgele Orman (RO) ile %73,72 olarak tespit etmişlerdir (Al Helal vd., 2019). Farklı bir çalışmada, belirlenmiş sınıflandırma algoritmaları iki farklı veri setine uygulanmış veri setlerinin doğruluk yüzdesine etkisini ortaya koymuşlardır. Ayrıca bahsi geçen çalışmada, Lojistik Regresyon algoritması ile %96 en yüksek doğruluk başarısına ulaşılmıştır, diğer en yüksek skorlara ise sırasıyla %94 DAA ve %93 AdaBoost algoritması ile ulaşılmıştır (Mujumdar ve Vaidehi, 2019). Güncel bir diyabet tanı çalışmasında, sayısal ve metinsel veri kümesini işlemek için İBSA modeli önerilmiştir. Bu yaklaşım ile Pima Indian Diabetes veri setinde %97,27 eğitim ve %96,09 test doğruluğu elde edilmiştir (Frimpong vd., 2021). Bi vd. çalışmalarında hastaların yaşam tarzı müdahalelerinin etkisini ölçmek için yeni bir makine öğrenimi modeli önermiştir, 6405 kadın ve 5913 erkek hastadan elde edilen veri setine uygulanmıştır. Önerilen yöntemin başarısının ÇKAYSA ve Destek Vektör Regresyon (DVR)' a göre daha yüksek olduğu gözlemlenmiştir (Bi vd., 2021). Benzer bir başka çalışmada, 500 diyabet hastası ve 268 sağlıklı insan olmak üzere toplam 768 örnekle yapılan çalışmadan oluşturulan veri setinin otomatik sınıflandırılması için ÇKA Derin Öğrenme Algılayıcısı ve DVM yöntemleri kullanılmıştır. Derin öğrenme algılayıcı sınıflandırıcısı, diyabet veri setiyle iyi performans göstermiş %77,474 başarı değerine ulaşmıştır (Thaiyalnayaki, 2021). Tiwari vd. temel özelliklerle diyabet tahmini yapmak için Rastgele Orman ve Özyinelemeli Eleme yoluyla önemli özellik seçimi yapmıştır. XGBoost ve YSA kıyaslandığında XGBoost YSA' ya göre %78,91 oran ile daha yüksek doğruluğu sağlamıştır (Tiwari ve Singh, 2021). Aynı yıl yapılan farklı bir çalışmada Fan vd. diyabet hastalığı tahmini için hastaların dijital dil görüntüleri ve mide semptomlarını kullanmışlardır. Verilerin sınıflandırılmasında RO ve DVM kullanılmış, sonuç olarak RF sınıflandırma modelinin DVM sınıflandırıcı modelinden önemli ölçüde daha iyi olduğu tespit edilmiştir (Fan vd., 2021).

2. Materyal ve Yöntem (Material and Method)

Önerilen sistemin temel yapısı Şekil 1'de gösterilmektedir. Şekilde gösterildiği gibi, hastaların semptomları hakkındaki bilgileri içeren veri seti, çalışma kapsamında tasarlanan bir bilgisayar arayüzüne hekim tarafından girilecektir. Arayüze girilen bu bilgiler program vasıtası ile sayısal değerlere çevrilip makine öğrenme metot girişlerine uygulanacaktır. Daha sonra veriler sırasıyla, DAA, DVM, TÖA, KA, k-NN Metotları, ÇKAYSA algoritmaları ile eğitilecek ve her birinin performansı değerlendirilecektir. Performans değerlendirmesi çapraz doğrulama, duyarlılık ve özgüllük değerleri ile gerçekleştirilecektir. Daha sonra en yüksek performans değerini sağlayan algoritma arayüz yazılımına dahil edilecektir. Burada seçilen algoritma gelen verileri tahmin edecektir. Elde edilen sonuçlar, tekrardan arayüz üzerinden çıkışta hekim ekranında gösterilecektir.

Şekil 1. Çalışmanın Genel Yapısı (The Structure of the Study)

2.1. Veri Tabanı (Database)

Çalışmada kullanılan veri tabanı, 520 denek üzerinde yapılan diyabetle ilgili semptomların sorgulandığı bir rapor kümesinden oluşmaktadır. Veri tabanında, diyabet belirtileri sayılabilecek semptomlara sahip insanlardan elde edilen veriler bulunmaktadır. Bu veri tabanı, yakın zamanda diyabet tanısı konmuş denekler veya henüz diyabet olmamış ve semptom

gösteren deneklere doğrudan yapılan bir anket çalışması ile oluşturulmuştur. Veriler, Bangladeş Sylhet Sylhet Diyabet Hastanesinden doğrudan anket yöntemi kullanılarak hastalardan toplanmıştır (Islam vd., 2020). Çalışmada kullanılan bu veri kümesine, UCI Makine Öğrenmesi veri tabanı üzerinden ulaşım sağlanmıştır (Dua ve Graff, 2019). Veri tabanı Tablo 1'de de gösterildiği gibi 16 özellikten oluşmaktadır. Tablonun son satırında ise diyabet riski pozitif veya diyabet riski negatif olarak gösterilmektedir.

Tablo 1. Özniteliklerin Açıklaması (The Explanation of Attributes)

Özellikler	Değerler
Yaş	1. 20-35, 2. 36-45, 3. 46-55, 4. 56-65, 5. 65 yaşüstü
Cinsiyet	1.Erkek, 2.Kadın
Poliüri	1.Evet, 2.Hayır
Polidipsi	1.Evet, 2.Hayır
Ani kilo kaybı	1.Evet, 2.Hayır
Zayıflık	1.Evet, 2.Hayır
Polifaji	1.Evet, 2.Hayır
Genital pamukçuk	1.Evet, 2.Hayır
Görsel bulanıklık	1.Evet, 2.Hayır
Kaşıntı	1.Evet, 2.Hayır
Sinirlilik	1.Evet, 2.Hayır
Gecikmiş iyileşme	1.Evet, 2.Hayır
Kısmi parezi	1.Evet, 2.Hayır
Kas sertliği	1.Evet, 2.Hayır
Alopesi	1.Evet, 2.Hayır
Obezite	1.Evet, 2.Hayır
Sınıf	1.Pozitif, 2.Negatif.

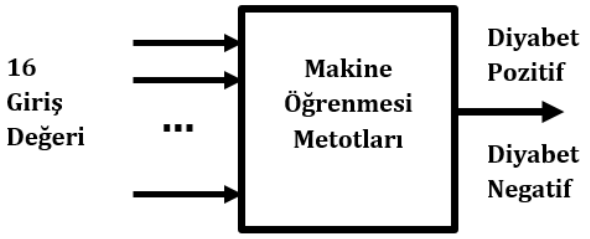
Şekil 2. Diyabet Tahmini İçin Oluşturulan Veri Giriş Arayüzü (Data Input Interface Created For Diabetes Prediction)

Pozitif olması hastanın diyabet olduğu, negatif olması ise deneğin diyabet olmadığını ifade etmektedir. Veri tabanında 320 pozitif bulunurken 200 denekte negatif olarak veri sağlanmıştır. Veri tabanın oluşturulurken yaş grupları da 5 ayrı şekilde ifade edilmiştir. Cinsiyet erkek ve kadın olarak yer alırken, diğer tüm özellikler ise evet veya hayır olarak değerlendirilmiştir. Bu özelliklerde ise "Evet" ifadesi semptomun bulunduğunu ifade ederken, "Hayır" seçeneği ise semptomun bulunmadığı anlamına gelmektedir. Bu özellikler ve seçenekler kullanılarak, veri girişi için bir bilgisayar arayüzü oluşturulmuştur. Bu arayüz detayları şekil 2'de gösterilmektedir.

Şekil 2'de görüldüğü üzere seçenekler sırayla yerleştirilmiş ve seçme işlemi için hekim denetimine sunulmuştur.

2.2. Önerilen Yöntemler (Proposed Methods)

Çalışmanın bundan sonraki kısmında ise, hastalardan elde edilen veriler, farklı makina öğrenmesi metotlarına sırasıyla uygulanmıştır. Bu metotlar DAA DVM, TÖA, KA, k-NN Metotları, ÇKAYSA algoritmaları şeklinde sıralanmaktadır. Bu metotların sisteme nasıl uyarlandığı Şekil 3'te gösterilmektedir.



Şekil 3. Makine Öğrenmesi Metotlarının Sisteme Uygulanması (Application of Machine Learning Methods to the System)

2.2.1. Doğrusal Ayrışım Analizleri (Linear Discriminant Analysis)

DAA algoritmaları genellikle modelleri iki sınıf arasında sınıflandırmak için kullanılır. Ancak, birden çok kategoride sınıflandırmak için genişletilebilir algoritmalardır. Çalışmadaki kategori sayısı iki olduğu DAA algoritması yeterli olabilmektedir. DAA, tüm sınıfların doğrusal olarak ayrılabilir olduğunu varsaymaktadır. Bu varsayımdan yola çıkarak, çoklu doğrusal ayrım işlevine göre, özellik uzayında birkaç hiper düzlemi temsil eden sınıfları ayırt etmek için oluşturulur. Çalışmada sadece iki sınıf olduğu için, DAA bir alt düzlem çizer ve iki kategorinin ayrılmasını optimum seviyeye ulaştıracak şekilde verileri bu alt düzleme yansıtır. Bu alt düzlem, aynı anda ele alınan iki kritere göre oluşturulur: bunlardan birincisi, iki sınıfın ortalamaları arasındaki mesafeyi maksimize etmek; ikincisi ise, her kategori arasındaki varyasyonu en aza indirmektir (Vaibhaw vd., 2020). Çalışmadaki DAA modeli üzerinde, Doğrusal, quadratik, diyagonal doğrusal, diyagonal quadratik, psödo-doğrusal ve psödo-quadratik ayrışım tipleri uygulanmış ve en başarılı sonuç quadratik tipi ile elde edilmiştir.

2.2.2. Destek Vektör Makinaları (Support Vector Machines)

DVM, bir sınıfın tüm veri noktalarını başka bir sınıfın tüm noktalarından ayıran en iyi hiper düzlemi algılayarak verileri sınıflandırmaktadır. Optimum seviyedeki hiper düzlem seçimi, iki kategori arasındaki en büyük marjın olduğu anlamına gelmektedir. Marjın, iç kısmında bir veri noktası olmayan bir hiper düzleme paralel olan maksimum genişlik anlamına gelmektedir. Destek vektörleri ise, hiper düzlem ayrımına en yakın verilerden geçen noktalardır. DVM'da, sınıflandırma fonksiyonu denklem 1'deki şekilde ifade edilebilmektedir (Mohammed vd., 2016; Huang vd., 2018).

$$f(z) = sign\left[\sum_{j=1}^{n} a_j y_j x_j z + b\right] \quad (1)$$

Denklem (1) 'de, x_j parametreleri destek vektörlerini ifade ederken, aj parametreleri optimum Lagrange çarpanlarını, y_j parametreleri ise sınıf etiketleri olarak (-1) veya (+1)'e eşittir. B değişkeni ise fonksiyonun eğilim (bias) parametresi olarak ifade edilmektedir. Metot üzerinde, optimal bir hiper düzlem tanımlanması için, marjın genişliğinin maksimum seviyeye çekilmesi gerekmektedir. Sonuçlar, doğrusal, Gauss fonksiyonu, Radyal Tabanlı fonksiyon ve farklı derecelerde polinom fonksiyonları da dâhil olmak üzere dört farklı çekirdek fonksiyonu kullanılarak test edilmiştir. Çalışma kapsamında, en iyi sonuç Radyal Tabanlı fonksiyon ile elde edilmiştir. Marjı ihlal eden gözlemlere uygulanan maksimum cezayı kontrol eden yumuşak-marj sabiti (YMS), 10-5'ten 105'e arasında seçilerek Bayes optimizasyon algoritması olarak seçilmiştir. Çalışmada kullanılan YMS parametre değeri 22,956 olarak seçilmiştir. Diğer taraftan, çekirdek (Kernel) ölçeği (ÇÖ) 10-5 ile 105 arasında ayarlanır ve aynı şekilde Bayesian optimizasyon algoritması kullanılarak seçilmiştir. Optimizasyon sonucunda ÇÖ ise 0,53067 olarak belirlenmiştir.

2.2.3. Topluluk Öğrenme Algoritmaları (Ensemble Learning Algorithms)

Topluluk öğrenme algoritmaları, tahmine dayalı analitik çalışmalarda en başarılı makine öğrenme algoritmalarından biridir. Bu tür algoritmalar somut bir problemin çözümü için geliştirilmiş modellerin bir araya gelmesi ile oluşmaktadır. Bu tür modellerin bir araya gelmesindeki temel amaç doğruluk değerlerinin artırılmasıdır. TÖA temel olarak, torbalama (Bagging), artırma (Boosting), istif (Stacking) şeklinde ortaya çıkmaktadır. Çalışmada farklı yöntemler test edilmiştir. Bu metotlar sırasıyla, Bag, Subspace, AdaBoostM1, AdaBoostM2, GentleBoost, LogitBoost, LPBoost, RobustBoost, RUSBoost, TotalBoost yöntemleridir. Bu yöntemler arasında en başarılı sonuç üreten metot Uyarlanabilir lojistik regresyon (LogitBoost) yöntemi olarak gözlemlenmiştir. LogitBoost yöntemi, Jerome Friedman, Trevor Hastie ve Robert Tibshirani tarafından formüle edilmiş bir yükseltme algoritmasıdır (Friedman vd., 2000). Logitboost algoritması aşağıda ifade edilen üç adımda gerçekleştirilir.

Adım 1: Ağırlık vektörü ve F(x) sınıflandırma fonksiyonu $w_i = \frac{1}{N}\ i = 1,2,\dots,N\ , F(x) = 0$ ve olasılık ise $p(x_i) = \frac{1}{2}$ olsun.

Adım 2: Bu işlem m=1, 2, …, M için tekrar yazılırsa;

- Çalışma cevabı ve ağırlıklarının hesaplanması;

$$z_i = \frac{y_i^* - p(x_i)}{p(x_i)(1-p(x_i))}, \quad (2)$$

$$w_i = p(x_i)(1 - p(x_i)) \quad (3)$$

Şeklinde olsun.

- $f_m(x)$ fonksiyonu, w_i ağırlıklarını kullanarak z_i' den x_i' ye ağırlıklı en küçük kareler regresyonu ile uydurulur.

- Daha sonra ise $F(x) \leftarrow F(x) + \frac{1}{2} f_m(x)$ ve $p(x) \leftarrow (e^{F(x)})/\ (e^{F(x)} + e^{-F(x)})$ olacak şekilde güncellenir.

Adım 3: Sonunda ise sınıflandırıcı çıkışı aşağıdaki şekilde ifade edilir.

$$sign[F(x)] = sign[\sum_{m=1}^{M} f_m(x)] \quad (4)$$

Çalışmada, Logitboost metodunun öğrenme hızı deneysel olarak test edilmiş ve en iyi sonuç 0,5 öğrenme hızı ile elde edilmiştir.

2.2.4. Karar Ağaçları (Decision Trees)

Karar ağaçlarının temel yapısı düğümler, dallar ve yapraklar olarak adlandırılan üç temel bölümden oluşmaktadır. Bu ağaç yapısında, her özellik bir düğüm ile temsil edilmektedir. Dallar ve yapraklar ise, ağaç yapısının diğer bileşenlerini oluşturmaktadır. KA yapısının temel prensibi, veriyi küçük parçalara bölerek en kısa sürede elde etmektir. KA'da, ağaçtaki her bir düğüm bir sınıfı belirtmekte veya düğümdeki test verilerini oluşturan olası çıktılara göre numune alanını ayırarak bir test bölümü oluşturmaktadır. Bölünen her bir alt küme, yeni bir alt ağaçla çözülerek yeni bir alt sınıflandırma problemi ortaya çıkaracaktır. Yapraklar olarak adlandırılan düğümler, sonuç düğümünün sınıfını içerir. Yaprak düğümü olmayan noktalar ise karar düğümleri olarak tanımlanır. Bu karar düğümleri yeni bir özellik oluşturur ve bu nedenle, bu özelliklerin her olası değeri için aynı ağacın dallarını bölerek başka bir karar ağacı oluşturur (Rokach ve Maimon, 2005). Minimum yaprak düğüm sayısı 1 ile 20 arasında ve

maksimum dal düğüm sayısı ise 1 ile 39 arasında seçilir. Bölme kriteri ise, Gini'nin Çeşitlilik İndeksi (GÇİ) sapma parametreleri kullanılarak Bayesian optimizasyon algoritması ile belirlenmektedir.

2.2.5. k-NN Metodu (k-NN Method)

K-NN algoritması, en basit ve temel makine öğrenimi algoritmalarından birisi olarak gösterilmektedir. Metot içindeki k değeri, tam sayı olarak en yakın özellikleri seçmek için belirlenmektedir. Test verileri ile eğitim verileri arasındaki mesafe Euclid, Manhattan, Chebyshev, Minkowski veya Hamming mesafe yöntemleriyle ölçülmektedir. Test verilerine en yakın k-eğitim verileri arasında en yakın komşular arasında en yaygın olan sınıflandırmaya dahil edilmiştir (Mohammed vd., 2016). Bu çalışmada, Bayes optimizasyon algoritması ve 30 tekrardan sonra en düşük hatayı veren değerler sınıflandırıcı parametresi kullanılarak 1'den 5'e kadar değişen k değerleri optimize edilmiş ve optimizasyon sonrası k değeri 1 olarak seçilmiştir. Diğer taraftan mesafe yöntemi de optimizasyon sonrası Minkowski mesafe yöntemi olarak belirlenmiştir. En başarılı doğruluk bu değerler ile elde edilmiştir.

2.2.6. Çok Katmanlı Algılayıcı Yapay sinir Ağları (Multilayer Perceptron Artificial Neural Networks)

Yapay Sinir Ağları, örneklerden olaylar arasındaki ilişkileri belirleyen ve daha önce hiç görmedikleri örnekler hakkında öğrendikleri bilgileri kullanan, birbirine bağlı yapay sinir hücrelerinden oluşan bilgisayar sistemleri olarak tanımlanır. YSA'nın sahip olduğu bilgi, her bir proses elemanıyla olan bağlantılarda ağırlık değerleri ile ağda depolanır ve ağa yayılır (Hu ve Hwang, 2002). Çalışma sırasında farklı YSA metotları test edilmiş ve en başarılı doğruluk değerini veren ÇKAYSA yapısı çalışmaya dahil edilmiştir. ÇKAYSA bir giriş katmanından, bir veya daha fazla gizli katman ve bir çıkış katmanından oluşmaktadır (Hu ve Hwang, 2002). ÇKAYSA' nın çıktısı ise aşağıdaki matematiksel formül ile ifade edilebilmektedir.

$$y = f\left(\sum_{i=1}^{N} X_i W_{ij} + b_j\right),\ (j\ =\ 1,2,\dots,M) \qquad (5)$$

Bu denklemde M katman sayısıdır, giriş katmanındaki nöron sayısı N ile gösterilir, X_i gizli bir katmandaki i. Nöronu ifade eder, W_{ij} her giriş için ağırlık olarak açıklanır, b_j algılayıcı, f aktivasyon fonksiyonudur ve y j. katmandaki algılayıcı çıktısıdır. Çalışmada ÇKAYSA için farklı katman sayıları farklı nöron sayıları ve aktivasyon fonksiyonları test edilmiştir. En yüksek doğruluk 2 katmanlı her iki katmanında 10 nöron bulunan ÇKAYSA yapısı ile elde edilmiştir. Her iki katman ve çıkış katmanında tanjant sigmoid fonksiyonu kullanılmıştır. Tespit edilen özellikler farklı YSA yapıları ile analiz edilmiş ve sonuçlar gözlemlenmiştir. ÇKAYSA'da ağırlık matrisini seçmek için geri yayılım eğitim hatası yöntemi olarak Levenberg-Marquardt algoritması kullanılmıştır.

3. Deneysel Sonuçlar (Experimental Results)

Çalışmada, makine öğrenmesi metotları kullanılmış ve doğruluk değerleri birbirleri ile kıyaslanmıştır. Kıyaslama ölçütlerinde bilindik doğrulama yöntemleri kullanılmıştır. Buna göre, doğrulama için 3, 5 ve 10 kat çapraz doğrulama (KÇD) işlemleri gerçekleştirilmiş ve her doğrulama sonucunda duyarlılık, özgüllük ve doğruluk değerleri hem tüm veri tabanı için hem de sadece test verileri için kaydedilmiştir. Çalışmada hesaplanan duyarlılık, özgüllük ve doğruluk değerleri şu şekilde hesaplanmaktadır.

$$Duyarlılık(\%) = \frac{DP}{DP+YN} \times 100 \qquad (6)$$

$$Ozgulluk(\%) = \frac{DN}{DN+YP} \times 100 \qquad (7)$$

$$Dogruluk(\%) = \frac{DP+DN}{DN+DP+YN+YP} \times 100 \qquad (8)$$

Bu formüllerde geçen parametreler aşağıdaki gibi tanımlanmaktadır;

• DP: Gerçekte pozitif sınıf içinde, tahminde ise pozitif sınıf içinde yer alan değer sayısı.
• YN: Gerçekte pozitif sınıf içinde, tahminde ise negatif sınıf içinde yer alan değer sayısı.
• YP: Gerçekte negatif sınıf içinde, tahminde ise pozitif sınıf içinde yer alan değer sayısı.
• DN: Gerçekte negatif sınıf içinde, tahminde ise negatif sınıf içinde yer alan değer sayısı.

Çalışmada kullanılan farklı makine öğrenmesi yöntemlerinin detaylı doğruluk bilgisi ile performanslarının karşılaştırılması Tablo 2'de gösterilmiştir.

Sonuçlar karşılaştırıldığında DVM metodunun giriş verileri göz önünde bulundurulduğunda diyabet riskinin pozitif veya negatif olacağı yönündeki değerlendirmesinde çok etkili bir metot olduğu görülmektedir. Çapraz doğrulama işlemi 3-KÇD, 5-KÇD ve 10-KÇD olarak değerlendirilmiştir. Ayrıca sonuçlar duyarlılık ve özgüllük yönünden de değerlendirilmiştir.

Tablo 2. Metotların k-Kat Çapraz Doğrulama Karşılaştırmaları (k-Fold Cross Validation Comparisons of Methods)

KÇD	METOT	VERİLERİN TÜMÜ			TEST VERİLERİ		
		Duy. (%)	Özg. (%)	Doğ. (%)	Duy. (%)	Özg. (%)	Doğ. (%)
3-KÇD	DAA	93,76	96,59	94,74	90,51	95,67	92,12
	DVM	98,76	97,72	98,33	96,60	93,53	95,19
	TÖA	99,48	97,86	98,85	98,41	93,76	96,54
	KA	96,57	93,03	95,13	94,54	89,87	92,50
	k-NN	99,79	97,87	99,04	99,36	93,85	97,12
	ÇKAYSA	99,06	97,86	98,59	97,17	93,71	95,77
5-KÇD	DAA	96,71	95,76	96,35	94,57	95,42	94,81
	DVM	99,37	98,51	99,04	96,86	92,75	95,19
	TÖA	99,56	99,40	99,50	97,86	97,01	97,50
	KA	97,29	94,64	96,12	94,47	90,19	91,92
	k-NN	99,87	99,01	99,54	99,37	95,26	97,69
	ÇKAYSA	99,24	93,29	96,27	97,40	88,31	93,08
10-KÇD	DAA	96,89	95,53	96,37	94,48	95,25	94,42
	DVM	99,66	99,50	99,60	96,77	95,21	95,96
	TÖA	99,81	99,65	99,75	98,17	96,63	97,50
	KA	98,12	95,60	97,12	96,56	91,62	94,42
	k-NN	99,91	99,65	99,81	99,07	96,66	98,08
	ÇKAYSA	99,75	99,30	99,58	97,46	93,42	95,77

Her KÇD bölümünde elde edilen duyarlılık ve özgüllük değerleri kabul edilebilecek seviyede olup gayet iyi sonuç vermişlerdir. Diğer taraftan doğruluk değerleri karşılaştırıldığında en yüksek tahmin doğruluk değerinin her KÇD bölümünde k-NN metoduna ait olduğu ve %99,81 şeklinde doğruluk elde edildiği aşikârdır. Dolayısıyla, çalışmada gerçekleştirilen optimizasyon teknikleri ve deneysel gözlemler diyabet riski sınıflandırılmasında k-NN yönteminin çok etkili olduğunu ortaya koymaktadır. Tablo 2'ye göre, DVM den başka, DVM ve TÖA yöntemlerinin de gayet başarılı sonuçlar verdiği görülmektedir. Diğer taraftan bakıldığında ise DAA yönteminin bu tür bir sınıflandırmada diğer metotlara göre hayli yetersiz kaldığı görülmektedir.

Bu çalışmada elde edilen sonuçlar literatürde gerçekleştirilmiş olan diğer çalışmalar ile karşılaştırıldığında, DVM yönteminin hayli başarılı bir yöntem olduğu burada da ortaya çıkmaktadır. Sonuçların diğer metotlarla karşılaştırılabilmesi açısından kesinlik ve F1 skoru değerleri de hesaplanmıştır. Bu değerler aşağıdaki formüller ile ifade edilebilmektedir.

$$Kesinlik(\%) = \frac{DP}{DP+YP} \times 100 \quad (9)$$

$$F1\ skoru(\%) = 2 \times \frac{Kesinlik \times Duyarlılık}{Kesinlik + Duyarlılık} \quad (10)$$

Elde edilen kesinlik, hassasiyet, F1 ve doğruluk skoru değerleri literatürdeki diğer çalışmalar ile karşılaştırılmış ve Tablo 3'te gösterilmiştir.

Bildirilen çalışmaların büyük çoğunluğunda, diyabet riskinin öngörülmesinde sınıflandırma doğruluğunu %80'lerin üzerinde olduğunu belirtmek gerekir.
Tabloda görüldüğü üzere, Islam vd., 4 farklı metot kullanarak, en yüksek doğruluğu Rasgele orman (RO) yöntemi ile elde etmişler ve bu yöntemle %97,4 doğruluğu yakalamışlardır (Islam vd., 2020). Diğer taraftan Özer, yaptığı çalışmada, uzun kısa dönem bellek ağlarını (LTSM) kullanarak aynı veri seti üzerinde %98,9 başarı elde etmişlerdir. Frimpong vd. 2021' deki çalışmasında İleri Beslemeli YSA modeli kullanarak %96,09 doğruluk ve %94,88 F1 skorunu yakalamıştır (Frimpong vd., 2021). Bunlara karşın, gerçekleştirilen bu çalışmada ise k-NN yöntemi kullanılmış ve bu yöntem ile %99,81 oranında bir başarı elde edilmiştir. Ayrıca %99,83 F1 skoru yakalama başarısı gösterilmiştir. Ayrıca, alışılagelmiş öznitelikler kullanılarak yapılan sınıflandırma çalışmalarından farklı olarak Fan ve arkadaşları dil görüntülerinden elde ettikleri öznitelikleri kullanarak yaptıkları çalışmalarında 3 birleşik özellik seti kullanmıştır.

Tablo 3. Metotların k-Kat Çapraz Doğrulama Karşılaştırmaları (k-Fold Cross Validation Comparisons of Methods)

Çalışma	Yöntem	Has. (%)	Kes. (%)	F1 Skoru (%)	Doğ. (%)
(Kaur ve Kumari, 2018)	DVM-Lineer	87	88	87	89
	k-NN	90	87	88	88
	YSA	88	85	86	86
	RBF-DVM	83	85	83	84
	MDR	83	82	84	83
(Mujumdar ve Vaidehi, 2019)	Lojistik Regresyon	-	-	-	96
	DAA	-	-	-	94
	Gradient Boost Sınıflandırıcı	-	-	-	93
	AdaBoost Sınıflandırıcı	-	-	-	93
	Gaussian Naive Bayes	-	-	-	93
	Rastgele Orman	-	-	-	91
(Islam vd., 2020)	Naive Bayes	87,4	87,9	87,5	-
	Lojistik Regresyon	92,4	92,4	92,4	-
	J48	95,6	95,7	95,6	-
	Rasgele Orman	97,4	97,4	97,4	-
(Özer, 2020)	LSTM	99,37	98,43	98,9	-
(Frimpong, 2021)	İleri Beslemeli YSA	93,28	96,53	94.88	96,09
(Fan vd., 2021)	Rastgele Orman	93,5	-	-	94,2
	DVM-Lineer	79,8		-	88,8
(Tiwari ve Singh, 2021)	XGBoost Sınıflandırıcı	59,33	-	-	78,91
	YSA	45,22	-	-	71,35
Bu çalışma	**k-NN**	**99,91**	**99,75**	**99,83**	**99,81**

Rastgele Orman (%94,2) ve DVM-Lineer (%88,8) metotlarında en yüksek skorlara ulaşmışlardır (Fan vd., 2021).

Birçok sınıflandırma probleminde olduğu gibi diyabetin erken tanısı için gerçekleştirilen bu çalışmada da veri setlerinin niteliği ve miktarı önem kazanmaktadır. Güncel bir çalışmada 9 farklı öznitelik kullanılmış XGBoost Sınıflandırıcı ve YSA yöntemleri ile yüksek başarılar elde edilmiştir. XGBoost ile 78,91 ve YSA ile 71,35' tir (Tiwari ve Singh, 2021). Kaur ve Kumari yapmış olduğu çalışmada makine öğrenmesi için 8 öznitelik kullanmış, en yüksek doğruluk değerlerine sırasıyla DVM-Lineer ve k-NN metodları ile %89 ve %88 elde etmiştir. Aynı çalışmada en yüksek F1 skoru değeri k-NN ile %88' dir (Kaur ve Kumari, 2018). Bir başka çalışmada Mujumbar ve Vaidehi, 10 farklı öznitelik kullanmış en yüksek skorları %96, %94 ve %93 olarak sırasıyla Lojistik Regresyon, DAA ve Gradient Boost Sınıflandırıcı algoritmaları ile elde etmişlerdir (Mujumbar vd., 2019). Buna karşın, bu çalışma, öznitelik çıkarımı için 16 farklı özellik kullanılarak ön plana çıkmış ve %99,81 doğruluk oranına ulaşmıştır. Bu sonuçlar ile yalnız kullanılan sınıflandırma metotlarının değil, öznitelik çıkarımı için kullanılan özelliklerin niteliğinin de sonuçta ne kadar etkili olduğu ortaya konulmuştur.

Diyabet riski erken tahmininde, kişilerin konumu, yaşı veya eğitim geçmişi ne olursa olsun, bu kişilere küresel olarak erişilebilir bir online sistem tasarımı şüphesiz ki önemli bir katkı olacaktır. Güncel teknolojide, internet, veri ve hizmet aramanın en yaygın yolu haline gelmektedir. Diyabet riskinin erken tespitinde, kullanıcı semptomlarının sisteme girilmesini sağlayacak basit bir web sitesi yapılabilir. Bu web sitesi hem diyabete yakalanma olasılığı tahminlerini yapabilir hem de diyabetik belirtiler ile ilgili bazı yararlı sağlık ipuçları sağlayabilir nitelikte olabilir. Bu ihtiyaç için çalışma kapsamında geliştirilen bir arayüz çalışması ana sayfası Şekil 4'te gösterilmektedir.

Şekil 4. Demo Arayüz Tasarımı (Demo Interface Design)

4. Sonuç ve Tartışma (Result and Discussion)

Farklı yaş gruplarından insanlar her geçen gün diyabet riski ile karşı karşıyadır. Bu çalışmada, diyabetin erken aşamada tespit edilmesinin tedavide çok önemli bir rol oynadığını ve erken teşhis için ipuçlarının yakalanmasının çok önemli olduğu vurgulanmaktadır. Beslenmede şeker dengesinin sağlanması, düzenli fiziksel aktivite gerçekleştirilmesi ve sağlıklı yaşam tarzına geçiş yapmak gibi bazı basit bilgilendirmeler kısmen de olsa obeziteyi önleyebilecek imkânlar sağlayabilecektir. Diyabeti tahmin etmek için geliştirilen makine öğrenmesi yöntemleri ve araçları geliştikçe ve bu yöntemlerin girişine uygulanacak veri sayısı ve denek sayısı arttıkça sistemin gerçekliği ve erken tanı ve teşhise maliyeti düşük bir şekilde vereceği destekte, bu tıbbi sağlık hizmetindeki rolünü yadsınamaz hale getirecektir. Bu çalışmanın literatüre olan temel katkısı, diyabetik riskinin erken tahmini için yapılan en iyi algoritmayı bulmaktır. Çalışmada, k-NN algoritmasının yüzdelik çapraz doğrulama değerlendirme testinde en iyi doğruluk performansı sergilediği gösterilmiştir. Çalışma sonucunda, k-NN algoritmasının doğruluk performans değeri %99,81 olarak hesaplanmıştır. Bununla beraber F1 skoru % 99,83 olarak elde edilmiş ve diğer çalışmalara göre üstünlüğü ifade edilmiştir. Diğer taraftan ise, son kullanıcı tarafında küçük anket tarzı bir sorgulama ile bilgisayar arayüzü üzerinden diyabet ön tahmini yapılması sağlanmıştır.

Teşekkür (Acknowledgment)

Bu çalışma Burdur Mehmet Akif Ersoy Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi tarafından desteklenmiştir.

Kaynaklar (References)

Ahmed, T. M., 2016. Developing a predicted model for diabetes type 2 treatment plans by using data mining. Journal of Theoretical and Applied Information Technology, 90(2), 181.

Al Helal, M., Chowdhury, A. I., Islam, A., Ahmed, E., Mahmud, M. S., & Hossain, S., 2019. An optimization approach to improve classification performance in cancer and diabetes prediction. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 1-5, IEEE.

Bi, S., Ding, X., Yu, S., Guo, B., Mu, L., Wang, B., 2021. A machine learning model for quantifying the effect of lifestyle interventions for patients with type 2 diabetes mellitus. In Journal of Physics, Conference Series, Vol. 1732, No. 1, p. 012006, IOP Publishing.

Bishop, C. M., 2006. Pattern Recognition and Machine Learning Springer-Verlag New York. Inc. Secaucus, NJ, USA.

Cichosz, S. L., Johansen, M. D., Hejlesen, O., 2016. Toward big data analytics: review of predictive models in management of diabetes and its complications. Journal of diabetes science and technology, 10(1), 27-34.

Dua, D., Graff, C., 2019. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Failure to detect type 2 diabetes early costing $700 million per year, Diabetes Australia, 8 July 2018. https://www.diabetesaustralia.com.au

Fan, S., Chen, B., Zhang, X., Hu, X., Bao, L., Yang, X., Liu, Z., Yu, Y., 2021. Machine Learning Algorithms in Classifying TCM Tongue Features in Diabetes Mellitus and Symptoms of Gastric Disease. European Journal of Integrative Medicine, 101288.

Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting. Annals of Statistics, 28 (2), 337–407.

Frimpong, E. A., Oluwasanmi, A., Baagyere, E. Y., Zhiguang, Q., 2021. A feedforward artificial neural network model for classification and detection of type 2 diabetes. In Journal of Physics: Conference Series, Vol. 1734, No. 1, p. 012026, IOP Publishing.

George, T., Rufus, E., Alex, Z.C., 2015. Simulation of microwave induced thermo-accoustical imaging technique for cancer detection. Journal of Engineering and Applied Sciences (ARPN), 10.

Goetz, T., 2010. The Decision Tree: Taking Control of Your Health in the New Era of Personalized Medicine, New York, NY, USA: Rodale.

Goldenberg, R., Punthakee, Z., 2013. Definition, classification and diagnosis of diabetes, prediabetes and metabolic syndrome. Canadian journal of diabetes, 37, S8-S11.

Harding, J. L., Pavkov, M. E., Magliano, D. J., Shaw, J. E., Gregg, E. W., 2019. Global trends in diabetes complications: a review of current evidence. Diabetologia, 62(1), 3-16.

Harris, M.I., Klein, R., Welborn, T.A., Knuiman, M. W., 1992. Onset of NIDDM occurs at least 4–7 yr before clinical diagnosis. Diabetes Care, 15(7), 815–819.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction.

Hu, H.Y., Hwang, J.N., 2002. Handbook of Neural Network Signal Processing, New York, NY, USA: CRC Press.

Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., Xu, W., 2018. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics-Proteomics, 15(1), 41-51.

Islam, M. F., Ferdousi, R., Rahman, S., Bushra, H. Y., 2020. Likelihood prediction of diabetes at early stage using data mining techniques. In Computer Vision and Machine Intelligence in Medical Image Analysis pp. 113-125, Springer, Singapore.

Joshi, T. N., Chawan, P. P. M., 2018. Diabetes Prediction Using Machine Learning Techniques. International Journal of Engineering Research and Application (Ijera), vol. 8, no.1, pp. 9-13, 2018.

Jutel, A., 2011. Classification, disease, and diagnosis. Perspectives in biology and medicine, 54(2), 189-205.

Karthikeyani, V., Begum, I. P., 2013. Comparison a performance of data mining algorithms (CPDMA) in prediction of diabetes disease. International journal on computer science and engineering, 5(3), 205.

Kaur, H., Kumari, V., 2018. Predictive modelling and analytics for diabetes using a machine learning approach. Applied Computing and Informatics.

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., Chouvarda, I., 2017. Machine learning and data mining methods in diabetes research. Computational and structural biotechnology journal, 15, 104-116.

Kononenko, I., 2001. Machine learning for medical diagnosis: history, state of the art and perspective. Artificial Intelligence in Medicine, vol. 23, no. 1, pp. 89-109.

Mercaldo, F., Nardone, V., Santone, A., 2017. Diabetes Mellitus Affected Patients Classification and Diagnosis through Machine Learning Techniques. Procedia Computer Science, 112, 2519-2528.

Mohammed, M., Khan, M. B., Bashier, E. B. M., 2016. Machine learning: algorithms and applications. Crc Press.

Mujumdar, A., Vaidehi, V., 2019. Diabetes Prediction Using Machine Learning Algorithms. Procedia Computer Science, 165, 292–299.

Ogurtsova, K., da Rocha Fernandes, J. D., Huang Y., Linnenkamp, U., Guariguata, L., Cho, N. H., Cavan, D., Shaw, J. D., Makaroff, L. E., 2017. IDF Diabetes Atlas: global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes research and clinical practice, 128, 40-50.

Özer, İ., 2020. Uzun Kısa Dönem Bellek Ağlarını Kullanarak Erken Aşama Diyabet Tahmini. Mühendislik Bilimleri ve Araştırmaları Dergisi, 2 (2) , 50-57.

Parashar, A., Burse, K., Rawat, K., 2014. A Comparative approach for Pima Indians diabetes diagnosis using lda-support vector machine and feed forward neural network. International Journal of Advanced Research in Computer Science and Software Engineering, 4(11), 378-383.

Parikh, R. B., Kakad, M., Bates, DW., 2016. Integrating predictive analytics into high-value care: the dawn of precision delivery. JAMA, 315, 651-652.

Rokach, L., Maimon, O., 2005. Decision trees. In Data mining and knowledge discovery handbook, pp. 165-192, Springer, Boston, MA.

Sapon, M. A., Ismail, K., Zainudin, S., 2011. Prediction of diabetes by using artificial neural network. In Proceedings of the 2011 International Conference on Circuits, System and Simulation, Singapore Vol. 2829.

Specht, D., 1991. A general regression neural network, IEEE Trans. Neural Netw. 2, 568-576.

Statistics About Diabetes: American Diabetes Association, 22 Mar 2018. https://www.diabetes. Org.

Thaiyalnayaki, K., 2021. Classification of Diabetes Using Deep Learning and SVM Techniques. International Journal of Current Research and Review, Vol, 13(01), 146.

The 6 Different Types of Diabetes: (5 Mar 2018). The diabetic journey. https:// thediabeticjourney.com/the-6-different-types-of-diabetes.

Tiwari, P., Singh, V., 2021. Diabetes disease prediction using significant attribute selection and classification approach. In Journal of Physics: Conference Series, Vol. 1714, No. 1, p. 012013.

Tran, B. X., Latkin, C. A., Giang, V. T., Huong, L. T. N., Son, N., Ming-Xuan, T., Zhi-Kai, L., Cyrus, S. H. H., Roger, C. M. H., 2019. The Current Research Landscape of the Application of Artificial Intelligence in Managing Cerebrovascular and Heart Diseases: A Bibliometric and Content Analysis. International Journal of Environmental Research and Public Health, 16,2699.

Türk Endokrinoloji ve Metabolizma Derneği (TEMD), 2013. Diabetes mellitus ve komplikasyonlarının tanı, tedavi ve izlem kılavuzu (6.baskı). Ankara, BAYT Bilimsel Araştırmalar Basın Yayın, 2012, 15-42.

TC Sağlık Bakanlığı Temel Sağlık Hizmetleri Genel Müdürlüğü (TSHGM), 2011. Türkiye Diyabet Önleme ve Kontrol Programı Eylem Planı (2011- 2014). Ankara: Sağlık Bakanlığı Yayın No:816.

Vaibhaw, Jay Sarraf, P.K. Pattnaik, Chapter 2 - Brain–computer interfaces and their applications, Editor(s): Valentina Emilia Balas, Vijender Kumar Solanki, Raghvendra Kumar, An Industrial IoT Approach for Pharmaceutical Industry Growth, Academic Press, 2020, Pages 31-54, ISBN 9780128213261.

Wang, C., Long, Y., Li, W., Dai, W., Xie, S., Liu, Y., Zhang, Y., Liu, M., Tian, Y., Li, Q., Duan, Y., 2020. Exploratory study on classification of lung cancer subtypes through a combined K-nearest neighbor classifier in breathomics. Scientific reports, 10(1), 1-12.

Zhong, G., Ling, X., Wang, L. N., 2019. From shallow feature learning to deep learning: Benefits from the width and depth of deep architectures. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(1), e1255.