Giyilebilir Cihazlardan Gelen Sağlık Verilerinin Kimliksizleştirilmesi Yeterince Güvenli mi?

Veli DURMUŞ*

Öz

Kişinin sağlığı ve günlük aktiviteleriyle ilgili bilgileri toplamak ve izlemek amacıyla, son yıllarda, gelişen teknoloji ile birlikte giyilebilir cihazların yaygınlaşması, kişisel sağlık verilerinin kolaylıkla izlenmesine ve toplanmasına öncülük etmiştir. Dolayısıyla, kişilere yönelik toplanan bu verilerin oluşturulması ve paylaşılması kolaylaşmıştır. Kişiye ait tanımlayıcı bilgilerin kaldırılarak oluşturulan veri setleri toplum sağlığı araştırmalarında, sağlık politikalarının geliştirilmesinde ve değerlendirilmesinde, ayrıca karşılaştırmalı etkinlik çalışmalarında kullanılabilir. Giyilebilir cihazlardan gelen sağlık verilerinin kimliksizleştirilerek erişilebilir olması ya da bir araştırmada kullanılan bu verilerin kimliklerinin belirsizleştirilerek yayınlanması bireysel düzeyde mahremiyeti korumada önemli bir unsur olarak görülmektedir. Ancak tanımlanmamış sağlık verilerinin yeniden tanımlama riski, veri setinin boyutu ve karmaşıklığı, diğer veri setlerinin veya bilgilerin mevcudiyeti ve yeniden tanımlama tekniklerinin kullanımına bağlı olarak değişmektedir. Bu çalışma ile giyilebilir cihazlar aracılığıyla kolaylıkla elde edilebilen sağlık verilerinin yeniden tanımlama riskine yönelik genel bir bakış sağlanması ve bu veriler anonimleştirilse dahi hangi ölçüde bireysel düzeyde mahremiyet riski oluşturabileceği konusu değerlendirilmektedir. Konuyla ilgili güncel ve özgün çalışmalar dikkatle ve önyargısız bir yaklaşımla sistematik olarak taranmış, elde edilen bulgular sentezlenerek bütüncül bir sonuca ulaşılmıştır. Sonuç olarak, sağlık verilerinin kimliksizleştirilmesi, mahremiyeti korumada önemli bir adım olsa da kesin bir çözüm değildir. Giyilebilir cihazlar aracılığıyla veri toplamak ve paylaşmak yaygın olmasından dolayı, mahremiyeti korumak için yeterli politika ve prosedürlerin oluşturulması yeniden tanımlamayla ilişkili potansiyel riskleri en aza indirebilir.

Anahtar Sözcükler: Giyilebilir cihazlar, sağlık verisi, kimliksizleştirme, yeniden tanımlama, mahremiyet.

Is De-identification of Health Data from Wearable Devices Secure Enough?

Abstract

In recent years, the spread of wearable devices with the developing technology has led to the easy monitoring and collection of personal health data in order to collect and monitor information about a person's health and daily activities. Therefore, the creation and sharing of this data collected for individuals has become easier. Datasets created by removing personal identification information can be used in public health research, in the development and evaluation of health policies, as well as in comparative effectiveness studies. Accessibility of health data coming from wearable devices by de-identifying or publishing these data used in a research by anonymizing their identities is seen as an important element in protecting privacy at the individual level. However, the risk of redefinition of unidentified health data varies depending on the size and complexity of the data set, the availability of other data sets or information, and the use of redefinition techniques. With this study, it is evaluated to provide an overview of the risk of redefinition of health data that can be easily obtained through wearable devices and to what extent this data can pose a privacy risk at an individual level, even if anonymized. Current and original studies on the subject were systematically reviewed with careful and unbiased consideration, and the findings were synthesized to reach a comprehensive conclusion. In conclusion, de-identifying health data is an important step in protecting privacy, but it is not a definitive solution. Because it is common to collect and share data through wearable

Derleme Makale (Review Article)
Geliş / Received: 13.07.2024 & **Kabul / Accepted:** 11.11.2024
DOI: https://doi.org/10.38079/igusabder.1326830
* Dr. Öğr. Üyesi, Kütahya Sağlık Bilimleri Üniversitesi, Sağlık Bilimleri Fakültesi, Sağlık Yönetimi Bölümü, Kütahya, Türkiye. E-posta: veli.durmus@ksbu.edu.tr ORCID https://orcid.org/0000-0001-6124-6109

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

devices, establishing adequate policies and procedures to protect privacy can minimize the potential risks associated with redefinition.

Keywords: Wearable devices, health data, de-identification, re-identification, privacy.

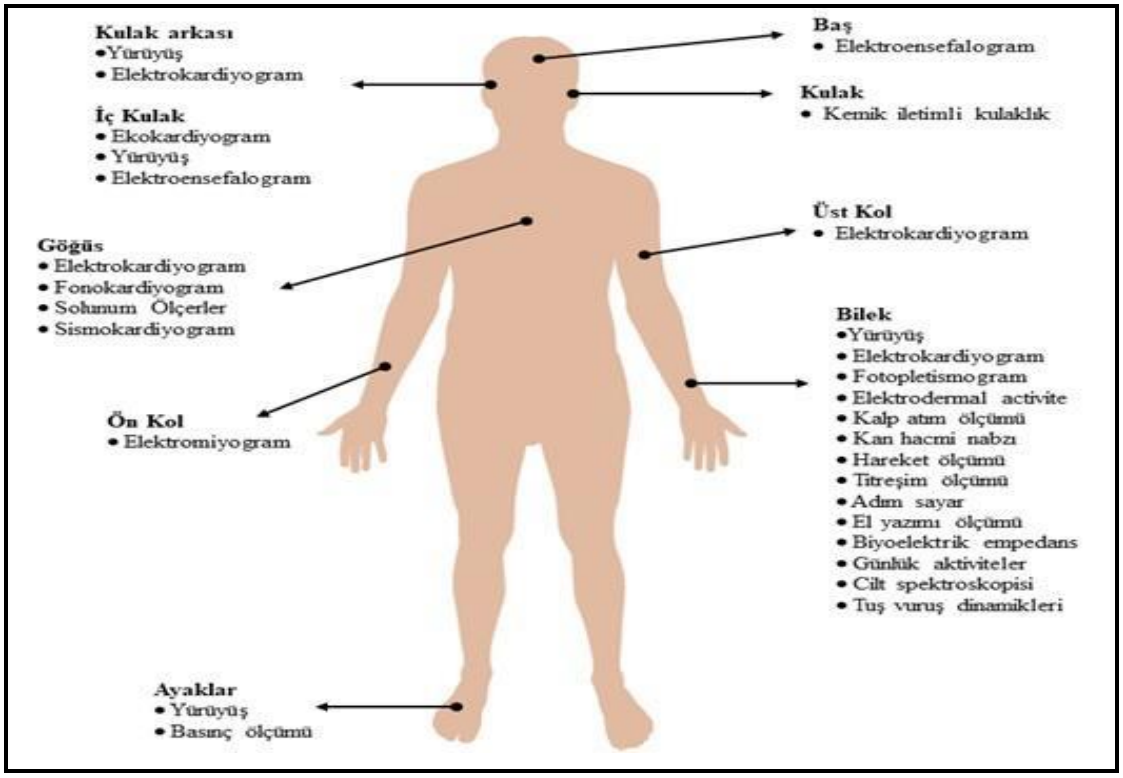
Giriş

Giyilebilir teknoloji kısaca vücudun çeşitli bölgelerine takılan ya da giysi veya aksesuarlara yerleştirilmiş elektronik cihazları ifade eder. Giyilebilir teknolojiler, özellikle geleneksel sabit tıbbi teknolojilerin etkin bir şekilde kullanılamaması durumunda nüfusun geniş kesimleri için kolayca erişilebilir sağlık verileri sağlama potansiyeli ile sağlık değerlendirmesinde önemli bir sınırı temsil etmektedir[1]. İlk nesil giyilebilir teknolojiler, temel olarak adım ve kondisyon takibi amaçlı kullanılıyorken, yeni nesil uygulamalar genişletilmiş işlevsellik ile sağlık hizmetlerinde sıklıkla kullanılmaya başlanmıştır[2]. Bu yüzden, küresel düzeyde hızla büyüyen giyilebilir teknoloji pazarı, 2021 yılında 116,3 milyar Amerikan doları (USD) değerinde iken, bu oranın 2026 yılında 265,4 milyar USD olacağı tahmin edilmektedir[3]. Sağlık hizmet sağlayıcıları, giyilebilir cihazları sağlık kayıt sistemlerine entegre ederek hasta bakımını iyileştirmek, teşhis koymak ve tedavi planları geliştirmek için çok sayıda veriye erişim elde edebilir.

Giyilebilir cihazlar temel olarak kullanıcının yapmış olduğu bazı eylemsel faaliyetleri, sağlığı ve çevresi hakkındaki verileri toplamak ve bunları iletmek için tasarlanmıştır. Bu cihazların, özellikle Kovid-19 hastalığının erken dönem semptomlarının tespit edilmesine yönelik bir araç olarak kullanılabileceği ön görülmesine rağmen[4], literatürde bununla ilgili çalışmalar oldukça sınırlıdır[5,6]. Ancak önceki çalışmaların sonuçları gösterdiği gibi, giyilebilir cihazlar değişken doğruluklarda sonuçlar ortaya koyduğundan bir hastalık hakkında doğrudan teşhis koyma için sınırlı veri sağlar[1,7,8]. Bu cihazlar, bir kişinin periferik dolaşımda kandaki hacimsel değişiklikleri tespit eden (fotopletismografi), kan basıncını (stigmomanometre) ve adım sayısını (ivmeölçer) izleyen ya da kan şekeri konsantrasyonunu sürekli ölçümleyen, aritmileri tespit eden genellikle bir aksesuar veya giyim eşyası olarak vücuda giyilebilen elektronik cihazlardır (Şekil 1).

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

Şekil 1. Giyilebilir cihazların vücuttaki bazı pozisyonları ve yaptıkları sağlık ölçümlerinden bazıları[9]



Bazı giyilebilir cihazlar klinik olarak ilgili olmayan iyi huylu ritim bozukluklarının gereksiz yere tespit edilmesine neden olduğundan, yanlış teşhislere, gereksiz araştırmalara ve hasta kaygısına yol açabilir[1,10]. Ayrıca giyilebilir teknolojilerin maliyetli olması, bu cihazların halk tarafından erişilebilirliğini sınırlayabilir. Örneğin, yaşlılar, teknolojiye aşina olmayanlar veya sosyoekonomik düzeyi düşük olanlar giyilebilir cihazlara erişimi veya uyumluluğu sınırlı olabilir[11].

Giyilebilir dijital sağlık teknolojileri, vücutta bulunduğu konumları, sensör dizilimleri ve yetenekleri açısından giderek daha çeşitli hale gelmektedir. Bu durum beraberinde çok geniş bir sağlık verisinin ortaya çıkmasına neden olmaktadır. Bu sağlık verilerini kullanarak, insan sağlığını iyileştiren araştırmalardaki ilerlemeleri desteklemek için Amerika Birleşik Devletleri (ABD) Ulusal Sağlık Enstitüleri, 2023'ten başlayarak kapsamlı veri paylaşım uygulamalarını teşvik eden politikalar benimsemiştir. Ayrıca çok sayıda kuruluş araştırma verilerini daha etkili ve verimli hale getirmek için bir dizi yol gösterici prensipleri belirlemeye başlamıştır. FAIR (Findable, Accessible, Interoperable, Reusable) adı verilen bu prensiplere bağlı kalarak, araştırma verilerinin daha keşfedilebilir, erişilebilir ve yeniden kullanılabilir hale getirilmesi amaçlanmaktadır[12]. Ayrıca şeffaflığı ve tekrar üretilebilirliği teşvik eden bu prensipler, araştırma verilerinin etkisini ve değerini en üst düzeye çıkarmaya yardımcı olmaya çalışmaktadır[13]. Araştırmanın yayınlanması sırasında verilerin paylaşılması, araştırmanın kesinliğini ve tekrarlanabilirliğini destekler. Analizler için verilerin mevcudiyeti, yeni yöntemlerin keşfini ve geliştirilmesini hızlandırabilir. Ulusal Sağlık Enstitüleri ve diğer araştırma fon

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

sağlayıcıları, nihai araştırma verilerinin paylaşılmasını giderek daha fazla beklemektedir ve araştırma verilerine erişimi kolaylaştırmak için kaynaklar veya araçlar sağlamaktadır[14]. Ancak son zamanlarda kamuoyuna duyurulan gizlilik ihlalleri ve verilerin kötüye kullanılması, hassas ve potansiyel tanımlanabilir bilgilerin paylaşılması riskine yönelik ilgiyi arttırmıştır[15]. Bununla birlikte, verilerin paylaşımı özellikle sağlık ve bilimsel faaliyetler açısından çok önemli faydalar sağlasa da, hastalara ve katılımcılara yönelik mahremiyetin korunması hakkında cevapsız kalan birçok önemli soruyu da beraberinde getirmiştir. Örneğin, elde edilen verilerin hükümetler, şirketler veya bireyler tarafından kötüye kullanılma ihtimali var mı? Eğer varsa, bu risk ne kadar kayda değer ve bunu azaltmanın bir yolu var mı?

Sağlık verilerinin kimliksizleştirilmesi (de-identification) bireylerin mahremiyetini korumak ve artık o kişiyi yeniden tanımlamak için kullanılamayacak şekilde sağlık verisinden kişisel tanımlayıcı bilgilerin çıkarılması, değiştirilmesi veya gizlenmesi işlemi olarak tanımlanabilir. Bu işlemlerden bazıları, kişinin ismi ve kimlik numarası gibi doğrudan tanımlayıcı bilgilerin kaldırılması ya da yaş, posta kodu gibi dolaylı tanımlayıcı bilgilerin kaldırılması şeklindedir. Ayrıca sağlık verilerinin anonimleştirilmesi, takma isim verilmesi, maskeleme, kategorikleştirme (yaş veya bölge gibi) veya genelleştirme gibi yöntemlerde uygulanmaktadır[16]. Sağlık verilerinin kimliksizleştirilmesi, toplumsal fayda sağlayan önemli tıbbi araştırmaları mümkün kılmak adına özellikle devlet kurumları, işletmeler ve verileri toplumun kullanımına sunmaya çalışan diğer kuruluşlar için önemlidir. Uygulamada bir veri seti paylaşılmadan önce kişisel olarak tanımlanabilir herhangi bir veri bulundurulmaması esastır. Bu gereklilikler ABD'de Taşınabilir Sağlık Sigortası ve Sorumluluk Yasası (The Health Insurance Portability and Accountability Act-HIPAA), Avrupa'da ise Genel Veri Koruma Düzenlemesi (General Data Protection Regulation-GDPR) tarafından yasal bir zemine dayandırılmaktadır. Diğer yandan, yeniden tanımlama (re-identification) süreci ise anonim veya kimliği belirsizleştirilmiş verileri belirli bir kişiye geri bağlama işlemidir. Yani, kimliği belirsizleştirilmiş veya anonimleştirilmiş verilerden yola çıkarak belirleme eylemi olarak tanımlanabilir. Bu, veri eşleştirme, istatistiksel çıkarım, farklı kaynaklardan gelen verileri birleştirmek ve diğer gelişmiş veri analiz yöntemleri dahil olmak üzere çeşitli tekniklerle yapılabilir. Bir saldırganın tanımlayıcıları içeren verilere erişim elde etme olasılığının önceki zamanlarda düşük olduğuna inanıldığından, yeniden tanımlamaya yönelik mahremiyet endişeleri tarihsel olarak ikna edici olmamıştır. Bununla birlikte, teknolojinin gelişmesiyle giderek artan sayıdaki şirketlerin bazılarında üçüncü şahıs veri paylaşım protokollerinin yeterince güçlü olmaması (politik ve maddi kazanım amacıyla vb.) buna katkı sağlamıştır[17,18].

Yeniden tanımlamanın bir sonucu olarak, görünüşte zararsız olan, kişilerin tespit edilmesini sağlamayan verilerin yayınlanmasının öngörülemeyen sonuçları olabilir. Bununla ilgili kayda değer bir örnek, 1990'lı yıllarda ABD'de yaşanmıştır. Buna göre, Massachusetts valisi William Weld tarafından çeşitli sebeplerle hastanede tedavi görmüş çalışanların, tıbbi kayıt bilgilerini içeren araştırma veri seti kamu ile paylaşılmıştır[9]. Kişilerin mahremiyetini korumak için isimlerin veri setinden tamamen çıkarılmış olmasına rağmen, doğum tarihi, posta kodu ve cinsiyeti gibi veriler istatistiksel analizi mümkün kılmak için muhafaza edilmiştir. Latanya Sweeney adlı bir araştırmacı bu

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

bilgiler ile Cambridge seçmen kayıt listesinde yer alan bilgileri kullanarak mevcut kayıtlar yeniden tanımlamıştır. Sweeney daha sonra bulgularını genelleştirerek, 1990 nüfus sayımına göre ABD nüfusunun %87'ye varan kısmının 5 haneli posta kodu, doğum tarihi ve cinsiyetiyle benzersiz bir şekilde tanımlanabileceğini savunmuştur[19]. Bunun üzerine, 1996 yılında HIPAA yürürlüğe girerek çalışanları iş değişikliğinde sağlık sigorta kapsamını sürdürebilmesine imkân tanınmış hem de bireysel düzeyde sağlık sigortası verilerinin yeniden tanımlamaya daha az elverişli olmasını sağlayacak düzenlemeler getirilmiştir[20]. Bu örnek aslında yasal düzenlemelerin, verilerin gerçek dünyadaki yeniden tanımlama olaylarının ve bunların sonuçlarının gerisinden geldiğini de göstermektedir.

Bu derleme çalışmasıyla, günümüzde yaygın kullanım alanı bulan giyilebilir cihazlardan elde edilen sağlık verilerinin kimliksizleştirilerek kullanılması ve kamuya açık hale getirilmesi sonucu çeşitli yöntemlerle yeniden tanımlanabilme riskine yönelik genel bir bakış sağlanması amaçlanmıştır. Özellikle giyilebilir cihazlar yoluyla elde edilen sağlık verilerinin anonimleştirilerek kullanılsa dahi bireysel düzeyde nasıl mahremiyet riski oluşturabileceği konusu değerlendirilmiştir. Bu bağlamda başlıca şu sorulara cevap aranmıştır: 1) Sağlık verileri kimliksizleştirilerek çeşitli yollarla (araştırma, rapor vb.) kamuya açık hale gelse bile, kimliğin yeniden tespiti mümkün müdür? 2) Giyilebilir cihazlardan ne tür sağlık verileri elde edilmektedir? 3) Güncel literatür ışığında, giyilebilir cihazlardan gelen verilerin kimliğini gizlemek, veri setlerindeki kişilerin mahremiyetini korumak için yeterli midir? Bu araştırma sorularına cevap bulmak için, sistematik derleme yapılarak, alanda yayınlanmış orijinal çalışmalar sistemli ve yan tutmadan taranmış, bulunan çalışmalar sentezlenerek birleştirilmiştir.

Kimliksizleştirilen Sağlık Verilerinin Önemi

Hastaya ait belirli tanımlayıcı bilgilerin (adı, soyadı, kimlik numarası vb.) veri setinden çıkarılmasıyla veya o kişiye aitliği belli olmayacak bir halde anonimleştirilmesiyle oluşturulan sağlık verileri, bireylerin mahremiyetini koruyarak bu verilerin toplum sağlığı araştırmalarında, sağlık politikalarının geliştirilmesinde ve değerlendirilmesinde, ayrıca karşılaştırmalı etkinlik çalışmalarında kullanılabilir. Örneğin, sahada yapılan aşı çalışmalarının etkinliğinin toplum sağlığı açısından ölçülmesi için aşı olunan bireylerin sosyodemografik bilgilerinin yanı sıra, sağlık ve ekonomik bilgilerine de ihtiyaç duyulabilir. Bu durumda, asıl değerlendirilmesi gereken bireylerin tanımlayıcı bilgileri olmayacağından, bunların kimliksizleştirilerek işlenmesi ve değerlendirilmesi sonucu sağlık politikasının belirlenmesi mümkün olabilecektir.

Kovid-19 hastalığı ve bunun gibi geniş kitlelere yayılma özelliği gösterebilecek hastalıkların geriye dönük olarak asemptomatik ve pre-semptomatik takip ve tespit potansiyeli giyilebilir cihazlar sayesinde mümkün olabilir[21,22]. Ayrıca bu cihazlar pandemi sırasında salgının yayılmasını azaltmak ve olası enfeksiyonları izlemeye yardımcı olmak için çeşitli şekillerde kullanılmıştır[21,23]. Binlerce hastadan kimliği belirsizleştirilmiş çok sayıda göğüs röntgen filmi veri seti kullanılarak makine öğrenimi algoritmaları oluşturulmuştur. Bu sayede klinik tahminleme modelleri oluşturularak teşhislerin doğruluğu arttırılmaya çalışılmıştır[24]. Dolayısıyla çok sayıdaki bu veriler, hastalığın azaltılması, kontrol altına alınması ve tedavisi için çeşitli yöntemlerle

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

anonimleştirilmiş veya kimliksizleştirilmiş olsa bile, yeniden tanımlama riskinin tamamen mevcut olmadığını söylemek güçtür.

Bireylerin sağlık kayıtları ve diğer tanımlanabilir sağlık bilgileri hakkında mahremiyeti korumak amacıyla ABD tarafından oluşturulan yasal düzenleme (HIPAA) ile sağlık verilerinin kimliksizleştirmek için başlıca iki farklı yöntemden bahsedilmiştir[25]. Bunlardan birincisi, alanında bilgi ve deneyime sahip uzman bir kişi (istatistikçi, araştırmacı vb.) tarafından yeniden tanımla riskinin en aza indirgenmesidir (The Expert Determination Method). Bunun için uzman personel, HIPAA tarafından belirlenen yeniden tanımlama potansiyeli olan 18 farklı bilginin tespitini yapmak zorundadır. Bu bilgilerden bazıları, hastanın adı ve soyadı, telefon numarası, elektronik posta adresi, sosyal güvenlik numarası, tıbbi kayıt numarası, sağlık sigorta numarası, araç plaka numarası, biyometrik tanımlayıcılar (parmak izi, retina kaydı, ses izi vb.), tam yüzü gösterir görüntüler, sertifika numarası gibi verilerden oluşmaktadır. Eğer uzman kişi kimliğin yeniden belirlenmesi riskinin çok düşük seviyede olduğunu belirlerse, bu sağlık bilgileri kimliksizleştirilmiş olarak kabul edilir ve HIPAA kapsamında herhangi bir kısıtlama olmaksızın kullanılabilir. Ancak mahremiyetin ihlali noktasında bir risk varsa bunu kabul edilebilir düzeye indirmek için bazı bilgiler veri setinden çıkarılması gerekmektedir.[9]

Diğer bir uygulama ise güvenli liman (The Safe Harbor Method) yaklaşımıdır. Bu yöntemde de HIPAA tarafından belirlenen 18 tanımlayıcı bilginin veri setinden tamamen çıkarılması işlemi ön plandadır. Bu işlem sonucunda veri kimliksizleştirilmiş olduğu kabul edilir ve herhangi bir kısıtlama olmadan kullanılabilir. Dolayısıyla, uzman tarafından belirlenen bilgilerin çıkarılması yöntemi ile güvenli liman uygulaması arasındaki esas fark verilerin kimlik bilgilerinin kaldırılıp kaldırılmadığını belirlemek için izledikleri yaklaşımdır[25]. Güvenli liman uygulaması daha katı ve kuralcıdır. Çünkü bir veri setindeki 18 farklı bilgiyi kaldırmak her zaman mümkün olmadığından pratik kabul edilmemektedir. Bu, gerçekte yeniden tanımlama riski oluşturmayan verilerin kaldırılmasına neden olabilir[9]. Bu durumda, uzmanlık ve muhakeme gerektiren diğer yöntemin (The Expert Determination Method) kullanılması daha uygun olabilir[24]. Ayrıca unutulmamalıdır ki, veri setinden 18 tanımlayıcı bilginin tamamı kaldırılsa bile, kalan bilgiler bir kişiyi tanımlamak için diğer kamuya açık bilgilerle birleştirilebiliyorsa, yine de yeniden tanımlama riski olabilecektir. Hiçbir yöntemin doğası gereği diğerinden daha iyi olmadığına ve uygun yöntemin verilerin özel koşullarına ve bağlamına bağlı olacağına dikkat etmek gereklidir. Hangi yöntemin kullanılacağına ilişkin karar, her yöntemin risk ve yararlarının yanı sıra verileri işleyen kuruluşun teknik ve idari yeterliliklerinin dikkatli bir analizine dayanmalıdır[9,25].

Sağlık Verilerinde Yeniden Tanımlama Riski

Sağlık verileri, sağlık hizmetlerinin kalitesini, güvenliğini ve hasta merkezli hizmet sunumunu geliştirmek, bilimsel yeniliği desteklemek, yeni tedavilerin keşfedilmesi ve değerlendirilmesi ile birlikte güncel sağlık hizmeti sunumu modellerinin tasarlanması ve değerlendirilmesi için gereklidir. Biyometrik uygulamalar, davranışsal ve çevresel izleme cihazları ve uygulamalardaki teknolojiler arttıkça bunlardan elde edilen elektronik formdaki kişisel sağlık verilerinin hacmi de giderek büyümektedir[26]. Ancak sağlık hizmetlerinde makine öğrenimi geliştikçe bu verilerin yeniden tanımlanma riskini de

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

arttırarak bireysel mahremiyetle ilgili endişeleri beraberinde getirecektir[27]. Son yıllarda dijital sağlık teknolojileri alanında giyilebilir cihaz pazarının genişlemesi ile birlikte bu tür cihazların kullanımı yaygınlaşmıştır[18]. Dolayısıyla giyilebilir cihazlardan elde edilen sağlık verilerinin kimliksizleştirilmesi kişilerin mahremiyetini korumada yeterli olup olmadığı günümüzde tartışılan bir konu haline gelmiştir[7,15,28,29].

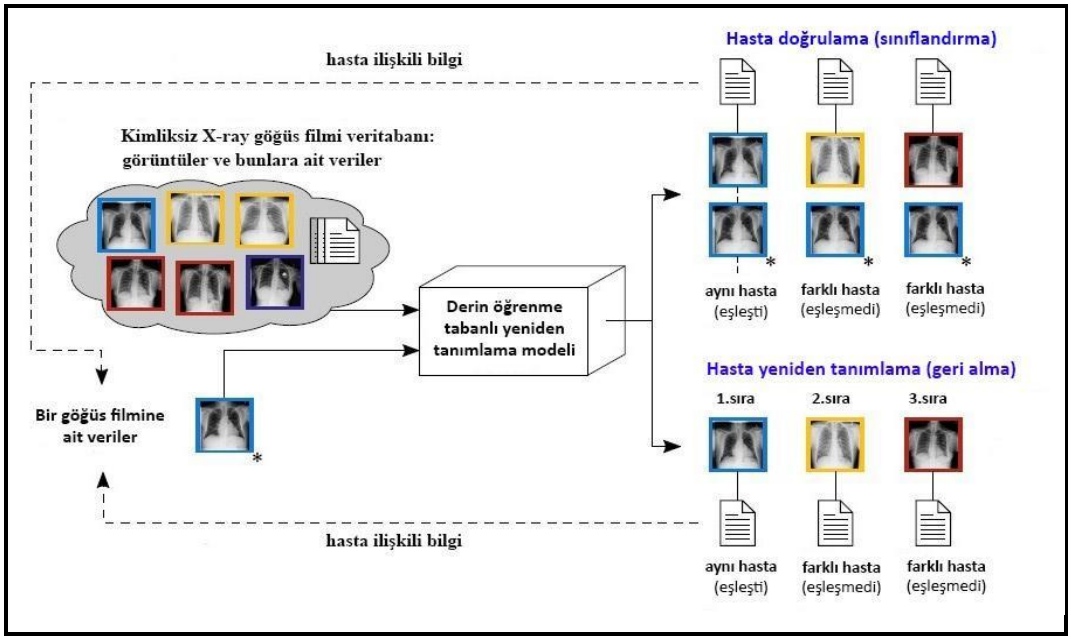
Aslında herhangi bir algılama yönteminden (giyilebilir sağlık cihazlarından elde edilen harekete, sese, kalp atım ritmine dayalı) elde edilen veriler potansiyel olarak biyometrik tanımlama ve kimlik doğrulama için kullanılabileceği gibi bir kişiyi yeniden tanımlamak için de kullanılabilir[8]. Bu yüzden bir veri setinden tanımlayıcı bilgilerin çıkarılmasıyla elde edilen verilerin anonimleştirilmesi kişilerin mahremiyetini yeterince koruduğunu söylemek güçtür.

Bu yöntemle ABD vatandaşlarının %87'sinin yalnızca doğum tarihi, cinsiyeti ve posta kodu ile benzersiz bir şekilde tanımlanabilme ihtimali bulunmaktadır[8,19,30]. Bu demografik bilgilerin her biri tek başına (yarı tanımlayıcılar) birini tanımlamak için yeterli olmasa da, bunların kombinasyonu nüfusun önemli bir yüzdesi için benzersiz bir tanımlama olabilir. Bununla birlikte, bazı belirli coğrafi bölgeler, bu tanımlayıcılar için çok daha yüksek benzersizlik oranlarına sahiptir. Örneğin, 90 yaşın üzerindeki hastalar ve bir posta kodunda yaşayan çok küçük bir etnik/ırk nüfusu gibi nadir senaryolar, yeniden tanımlama saldırılarına karşı onları daha savunmasız hale getirebilir.

Derin öğrenme (Deep learning) tekniklerinin son yıllardaki yükselişi ve sürekli artan potansiyeliyle birlikte, halka açık tıbbi veri kümeleri, tıp alanında teşhis algoritmalarının tekrarlanabilir şekilde geliştirilmesini sağlamak için önemli bir unsur haline gelmiştir[31]. Bugüne kadar derin öğrenme teknikleri kullanılarak büyük tıbbi veri setlerinde hastaların yeniden tanımlanmasına yönelik olası senaryo ihtimaline yeterince dikkat edilmediği iddia edilmektedir[7,9,20,32]. Ancak teoride, Şekil 2'de gösterildiği gibi, potansiyel saldırganlar için tıbbi verilerin yeniden tanımlanması, uygun derin öğrenme yaklaşımları kullanılarak mümkün olabilir.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

Şekil 2. Derin öğrenme tekniğiyle göğüs radyografisi verilerinin olası yeniden tanımlama senaryosu[7]



Anonimleştirildiği varsayılan ancak teşhis, tedavi geçmişi ve tedavi edilen kurum gibi hastayla ilgili daha fazla hassas bilgi içeren, halka açık bir veri kümesi düşünüldüğünde, kimliği bilinen bir radyografiye potansiyel bir saldırgan erişebilirse ve düzgün çalışan bir doğrulama veya yeniden tanımlama modeli varsa, bu durumda bu model, verilen radyografiyi veri kümesindeki her bir görüntüyle karşılaştırmak için kullanılabilir[33]. Sonuçta, bu radyografiye sahip olan aynı hasta ortaya çıkarılabilir (hasta tanımlama) ya da ait görüntülerin ortaya çıkması veya verilen radyografiye en çok benzeyen görüntülerin sıralı bir listesi oluşturulabilir (hasta yeniden tanımlama). Bu şekilde, hastanın kimliği, veri setinde yer alan hassas verilerle ilişkilendirilebilir. Ayrıca daha yaygın olarak bulunan ve daha az hassas olan verilerin (demografik özellikler, doğumlar, ölümler vb.) hassas bilgilerin yeniden tanımlanmasına yönelik yeni fırsatlar yaratma olasılığı da yüksektir[13]. 2022 yılında yapılan bir çalışmada[7], 30 binden fazla hastaya ait olan anonimleştirilmiş 112 bin önden görünüşlü göğüs röntgeni görüntüsü veri seti kullanılmıştır. İyi eğitilmiş derin öğrenme uygulaması aracılığıyla bu görüntüler kullanılarak, %95 doğruluk oranı ile hastanın kimlik bilgileri yeniden tanımlanabilmiştir. Benzer başka bir çalışmada, 499727 hastadan elde edilen 12 uçlu Elektrokardiyografi (EKG) sinyal verileri kullanılarak hastaların cinsiyeti % 90,4 oranında ve yaşı 6,9 – 5,6 yıl hata payı ile tahmin edilebileceği belirtilmiştir[34].

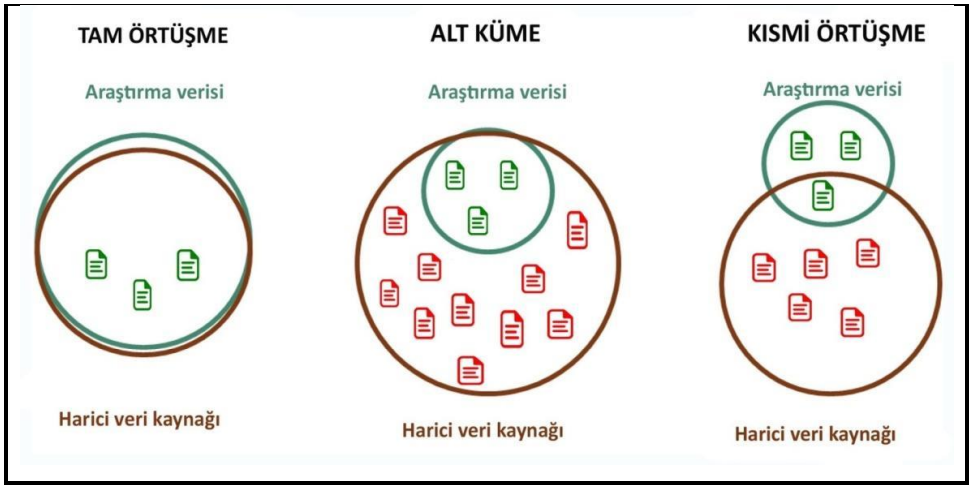
HIPAA tarafından belirlenen 18 tanımlayıcı bilginin doğrudan kişiyi tanımlamaya yönelik olmasının yanında, dolaylı tanımlayıcılarında diğer bilgilerle birleşerek kişiyi yeniden tanımlamaya önemli ölçüde katkı sağlayacağı değerlendirilmektedir[35]. Dolayısıyla, kişiye yönelik dolaylı tanımlayıcı özellikler arttıkça mahremiyet güvenliğinin tehdit altında olduğu anlaşılabilir. Örneğin, doktor muayene tarihleri ile yerlerini ve ayrıca bu muayene sırasında tedavi edilen rahatsızlıkları içeren bir veri kümesinin olduğunu varsayalım. Bu veri setinde hasta isimleri ve diğer doğrudan tanımlayıcılar

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

kaldırılmış olsa bile, bir saldırgan, cinsiyet, yaş ve posta kodu gibi diğer dolaylı tanımlayıcı bilgiler aracılığıyla tıbbi durumları belirli kişilerle ilişkilendirerek potansiyel olarak kişileri yeniden tanımlayabilir[7,15,36]. Buradaki dolaylı tanımlayıcılar ise bir kişiyi tümdengelim yoluyla tanımlamayı mümkün kılan veri öğeleridir. Örneğin, bir kişinin cinsiyet, ırk ve etnisite, dini inanç, yaş, medeni durum, doğum yeri, eğitim durumu, gelir düzeyi, iş ünvanı, çalışma yeri gibi bilgileridir[20].

Hassas veri setinin kapsadığı popülasyon ile harici bir veri kaynağının kapsadığı popülasyon uyumluysa veya tamamen örtüşüyorsa yeniden tanımlama riskinin oldukça yüksek olduğundan söz edilebilir. Hassas bilgileri içeren veri tabanı, daha genel bilgileri içeren veri tabanının bir alt kümesini temsil ediyorsa, yeniden tanımlama riskinin biraz daha düşük olduğundan söz edilebilir. Çünkü alt kümede yer alan verilerin doğru bir şekilde yeniden tanımlanması için harici veri kaynağındaki diğer verilerden ayrıştırılması ve bağlantının doğru sağlanması gerekir. Araştırma verisi ile harici veri kaynağının kısmi olarak örtüşmesi durumunda, yeniden tanımlama riski diğerlerine göre biraz daha düşük olacaktır. Çünkü elde edilmek istenen araştırma verisinin küçük bir kısmına ulaşılmıştır ve verinin tamamı tanımlanamamıştır[13,37]. Yeniden tanımlama riski ayrıca bir araştırma veri setinin kapsadığı popülasyon ile bağlantılı olabileceği gibi harici bir veri setinin kapsadığı popülasyon arasındaki örtüşme derecesine de bağlı olabilir (Şekil 3).

Şekil 3. Hassas veri setinin kapsadığı popülasyonlar ile harici veri seti arasındaki olası ilişkiler[13]



Sonuç olarak kabul edilebilir bir yeniden tanımlama riski için evrensel bir eşik yoktur. Risk tahmin edilmeye çalışılsa da, başarılı bir yeniden tanımlama saldırı olasılığı aynı zamanda saldırganın motivasyonuna ve elindeki kaynaklara da bağlı olacaktır. Herhangi bir bireysel kayıt için yeniden tanımlama riski, örtüşen veri öğeleri ile dış verilerin mevcudiyetinden, araştırma veri setinin kapsadığı ve dış veri kaynağının kapsadığı popülasyondan, söz konusu kaydın örtüşen öğeler tarafından tanımlama derecesinden ve boyutundan, ayrıca popülasyondaki örtüşmeye nasıl bağlı olduğundan da etkilenecektir[7,9].

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

Sonuç ve Öneriler

Kimliksizleştirilmiş geniş sağlık veri tabanlarının varlığı hastaları ve hastalıkları daha iyi anlamak için makine öğrenme yöntemi aracılığıyla önemli bir katkı sağlamasına rağmen, sağlık verilerinde mahremiyetin yeterince sağlanması ve bunların düzenlenmesi hala bir sorun olarak devam etmektedir. Özellikle, giyilebilir cihazlar sayesinde bireysel düzeyde sağlık verilerinin oluşturulması ve paylaşılması kolaylaşmıştır[7,38]. 2023 yılında yapılan bir sistematik literatür araştırmasında[9], vücudun çeşitli bölgelerine yerleştirilen birbirinden farklı sensörlerden elde edilen bilgiler kullanılarak yapılan çalışmalardaki verilerin kimliksizleştirilmesinin, bireylerin mahremiyetini korumak için yeterli olup olmadığını tartışılmıştır. Bu çalışmanın sonucuna göre, çeşitli yapay zeka uygulamaları kullanılarak yeniden tanımlama riskinin oldukça yüksek düzeylere çıkabildiği vurgulanmıştır. Üstelik bu yeniden tanımlama uygulamasının doğru tanımlama oranlarının % 86 ile % 100 arasında değişkenlik gösterdiği tespit edilmiştir[9]. Çalışmanın diğer bir sonucuna göre, daha az miktardaki bilgiler kullanılarak kimliksizleştirilmiş biyolojik sensörlerden elde edilen verilerin yeniden tanımlanabilme riskinin yüksek olabileceği ve bu sayede mahremiyetin ihlal edilebileceği değerlendirilmiştir. Kimliksizleştirilen sağlık verilerinin çeşitli sebeplerle (araştırma, analiz, yorum vb.) kamuyla paylaşılmasının giderek artması mahremiyetin sağlanması bakımından endişe verici olabilir.

Yeniden tanımlama ve mahremiyet riski bakımından, genel olarak giyilebilir cihazlardan elde edilen sağlık verileri, son teknoloji ürünü giyilebilir olmayan cihazlarla ilgili benzer araştırmalarla uyumludur. Örneğin, kamuya açık erişimi olan iki veri tabanından alınan 12 derivasyonlu EKG verileri, 40000 hastadan alınan diğer elektronik sağlık kayıtları verileriyle birleştiğinde, giyilebilir EKG verileri kullanılarak incelenen çalışmalarda bildirilenlere benzer yeniden tanımlama risk oranına sahip olduğu ortaya çıkmıştır[26].

Nihayetinde, yeniden tanımlamanın gerçekleşebilmesi için tanımlayıcılara sahip olmak gerekir. Bu nedenle yalnızca kimliği belirsizleştirilmiş veya anonimleştirilmiş veri kümelerindeki bireyleri eşleştirmek, gerçek yeniden tanımlamayı oluşturmayacaktır. Bu çalışma ile amaçlanan aslında giyilebilir cihazlardan biyometrik verilerin paylaşımının engellenmesini haklı çıkarmaya çalışmak değildir. Aksine, verilerin nasıl paylaşılması gerektiğine dair daha dikkatli değerlendirme yapılması gerektiğini ortaya koymaktır. Geniş kapsamlı çalışmaların sonuçları, bilimin gelişimi için sağlık verileri kullanarak mahremiyeti koruyan yöntemlere ihtiyaç duyulduğunu göstermektedir[7,8,9,23].

Giyilebilir cihazlardan elde edilen kimliksizleştirilmiş sağlık verilerinin kullanılarak yeniden tanımlama sonucu mahremiyetin ihlalini belirlemeye yönelik çoğu çalışmaların değerlendirmeye aldığı verilerin tek oturumda ya da tek bir zaman diliminde toplanmış olması bu türdeki çalışmaların bir sınırlılığını oluşturmaktadır. Örneğin, bu sınırlılık, yaşlanmayla birlikte ya da hastalık sayesinde kişilerin ses ve ciltteki değişikliklerine bağlı olarak zamanla değişebileceği için buna bağlı verilerin sonraki yıllarda da yeniden tanımlama için kullanılmasını güçleştirir[39,40]. Ayrıca bilinmelidir ki, bir hastalık yaygın değilse ve bir biyosensörle kolaylıkla tanımlanabiliyorsa, söz konusu sensör verilerinden bir bireyin yeniden tanımlanması nispeten daha kolay olabilir.

Kimliği belirsizleştirilmiş sağlık verilerinin yeniden tanımlanmasına yönelik mahremiyet endişeleri hala mevcut olsa da[9], bu verilerin aynı zamanda sağlık hizmetlerinde yenilikçi

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

yaklaşımların gelişimini hızlandırdığı yönünde tartışmalar da mevcuttur. Örneğin, Güney Kore, ABD ve Birleşik Krallık'tan elde edilen kimliksizleştirilmiş mamografi görüntü veri seti kombinasyonu sayesinde, radyologların tespitine kıyasla daha iyi meme kanseri tespitinin yapay zeka uygulamasıyla mümkün olabileceği iddia edilmiştir[41]. Hastalar ayrıca verilerinin araştırma veya ticari amaçla kullanılıp kullanılmadığını zamanla daha fazla sorgulayarak, bu verilerin kontrolü ve nasıl kullanıldığı ile ilgili konularda daha fazla ilgilenmeye başlamıştır[42]. Çünkü kimliği belirsiz hasta verilerinin artan kullanılabilirliği, makine öğrenimi odaklı yazılım inovasyonunda küresel düzeyde kayda değer gelişmelere yol açmıştır[43]. Kamuya açık kimliksizleştirilmiş veri kümeleri, yapay zekanın sağlık hizmetlerinde yaygın olarak uygulanmasının ve benimsenmesinin yanı sıra küresel olarak heterojen ve çeşitli hasta popülasyonlarına ilişkin anlayışı geliştirebilir. Tanımsızlaştırılan topluma açık sağlık verilerinin kullanımın kısıtlanması ya da erişimin zorlaştırılmasının maliyeti gelecekteki tıbbi yeniliklerin gelişimi üzerine olumsuz etkiler yaratabileceği düşünülmektedir[24].

2011 yılında yapılan bir sistematik literatür çalışmasına göre, sağlık verilerine yönelik yeniden tanımlamaya ilişkin girişimlerin başarı oranının sanılanın aksine düşük olduğu[44], bunun temel nedenlerinden birisinin de potansiyel saldırganın kaynak ve yeteneklerinin yeterince gelişmiş ve organize olamaması olduğundan bahsedilmiştir[45]. Aslında, sağlık verilerinde yeniden tanımlama riski, bireylerin mahremiyetini korumak için teknik, politika ve yasal önlemlerin bir kombinasyonunu gerektiren konudur.

Yeniden kimlik belirleme riskinin en aza indirilmesini sağlamak için kimlik gizleme sürecinde dolaylı tanımlayıcılar dikkatle değerlendirilmelidir. Doğrudan tanımlayıcıları kaldırarak veya gizleyerek ve dolaylı tanımlayıcıları değiştirerek, yeniden tanımlama riski önemli ölçüde azaltılabilir. Ayrıca giyilebilir verilerin saklanması için katı güvenlik önlemlerinin uygulanması kritik öneme sahiptir. Buna, yetkisiz erişime veya veri ihlallerine karşı koruma sağlamak için şifreleme, erişim kontrolleri ve düzenli güvenlik denetimleri dahildir. Verileri paylaşmaya karşı paylaşmamanın potansiyel risklerini ve faydalarını dengelemek için en iyi uygulamaları keşfetmek ve tartışmak üzere literatürde daha fazla çalışmanın yapılmasına ihtiyaç duyulmaktadır. Uygulamalardaki riskler ve faydalar hakkında yeni çalışmalar ortaya çıktıkça, veri paylaşım politikalarının gizlilik ve araştırma ilkeleri bağlamında yeniden değerlendirilmesine devam edilmelidir.

Hem kimliği belirsizleştirilmiş verilerin açık paylaşımını kalıcı olarak koruyan hem de hastanın yeniden tanımlanmasını güçlü bir şekilde cezalandıran sağlam bir düzenleyici çerçeve, veri paylaşımını geniş ölçüde sınırlandırmaya çalışmaktan daha ölçülü bir çözüm olabilir. Bununla birlikte veri minimizasyonu yaklaşımını benimsemek, giyilebilir cihazlar tarafından toplanan ve depolanan kişisel veri miktarını azaltmaya yardımcı olur. Yalnızca amaçlanan amaçlar için gerekli olan temel verilerin toplanması ve saklanması, yeniden tanımlamayla ilişkili potansiyel riskleri en aza indirebilir.

Sonuç olarak, giyilebilir cihaz sensör verileri paylaşıldığında gerçek bir yeniden tanımlama riskinden söz etmek mümkündür. Bu risk en aza indirilebilse de tamamen azaltılamaz. Giyilebilir cihazlarla veri toplama ve paylaşmanın yaygınlığı göz önüne alındığında, mahremiyeti korumak için yeterli politika ve prosedürlerin oluşturulmasına rehberlik edecek daha fazla araştırmaya ihtiyaç vardır.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

Verileri paylaşmama riski hayat kurtarabilecek yeni algoritmik araçlar geliştirilmesine engel olabileceğinden, yaratacağı etki yeniden tanımlama riskinden bile daha büyük olabilir. Bu nedenle, giyilebilir cihazların yararları ile mahremiyet hususlarını dengelemek, bireylerin hassas sağlık verilerini korumak ve kullanıcıların bu teknolojilere olan güvenini sürdürmek için çok önemlidir. Güçlü güvenlik önlemleri ve mahremiyet korumaları uygulanarak, yeniden tanımlama riski en aza indirilebilir.

KAYNAKLAR

1. Cheung CC, Krahn AD, Andrade JG. The emerging role of wearable technologies in detection of arrhythmia. *Can J Cardiol*. 2018;34(8):1083-1087. doi: 10.1016/j.cjca.2018.05.003.
2. Jia S, Gao H, Xue Z, Meng X. Recent advances in multifunctional wearable sensors and systems: design, fabrication, and applications. *Biosensors*. 2022;12(11):1057. doi: 10.3390/bios12111057.
3. Market Research Report: Wearable Technology Market. Report Code SE 2816, 2023. https://www.marketsandmarkets.com/Market-Reports/wearable-sensor-market-158101489.html. Erişim tarihi 18 Temmuz 2023.
4. Burki T. Wearable technology and COVID-19. *Lancet Respir Med*. 2022;10(10):934-935. doi: 10.1016/S2213-2600(22)00351-4.
5. Santos MD, Roman C, Pimentel MAF, et al. A real-time wearable system for monitoring vital signs of COVID-19 patients in a hospital setting. *Front Digit Heal*. 2021;3. doi: 10.3389/fdgth.2021.630273.
6. Cheong SHR, Ng YJX, Lau Y, Lau ST. Wearable technology for early detection of COVID-19: A systematic scoping review. *Prev Med (Baltim)*. 2022;162:107170. doi: 10.1016/j.ypmed.2022.107170.
7. Packhäuser K, Gündel S, Münster N, et al. Deep learning-based patient re-identification is able to exploit the biometric nature of medical chest X-ray data. *Sci Rep*. 2022;12(1):14851. doi: 10.1038/s41598-022-19045-3.
8. Ghazarian A, Zheng J, Struppa D, Rakovski C. Assessing the reidentification risks posed by deep learning algorithms applied to ECG Data. *IEEE Access*. 2022;10:68711-68723. doi:10.1109/ACCESS.2022.3185615.
9. Chikwetu L, Miao Y, Woldetensae MK, et al. Does deidentification of data from wearable devices give us a false sense of security? A systematic review. *Lancet Digit Heal*. 2023;5(4):e239-e247. doi: 10.1016/S2589-7500(22)00234-5.
10. Wang R, Blackburn G, Desai M, et al. Accuracy of wrist-worn heart rate monitors. *JAMA Cardiol*. 2017;2(1):104. doi: 10.1001/jamacardio.2016.3340.
11. Gillinov S, Etiwy M, Wang R, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. *Med Sci Sport Exerc*. 2017;49(8):1697-1703. doi: 10.1249/MSS.0000000000001284.
12. Boeckhout M, Zielhuis GA, Bredenoord AL. The FAIR guiding principles for data stewardship: fair enough? *Eur J Hum Genet*. 2018;26(7):931-936. doi: 10.1038/s41431-018-0160-0.
13. Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al. The FAIR Guiding Principles

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr

for scientific data management and stewardship. *Sci Data*. 2016;3(1):160018. doi: 10.1038/sdata.2016.18.

14. Simon GE, Shortreed SM, Coley RY, et al. Assessing and minimizing re-identification risk in research data derived from health care records. *eGEMs (Generating Evid Methods to Improv patient outcomes)*. 2019;7(1):6. doi: 10.5334/egems.270.

15. Raghupathi W, Raghupathi V, Saharia A. Analyzing health data breaches: a visual analytics approach. *AppliedMath*. 2023;3(1):175-199. doi: 10.3390/appliedmath3010011.

16. Ahmed T, Aziz MMA, Mohammed N. De-identification of electronic health record using neural network. *Sci Rep*. 2020;10(1):18600. doi: 10.1038/s41598-020-75544-1.

17. Fuller M. Big data and the Facebook scandal: Issues and responses. *Theology*. 2019;122(1):14-21. doi: 10.1177/0040571X18805908.

18. Schneble CO, Elger BS, Shaw D. The Cambridge Analytica affair and Internet-mediated research. *EMBO Rep*. 2018;19(8). doi: 10.15252/embr.201846579.

19. Sweeney L. *Simple Demographics Often Identify People Uniquely*. Yayınlanma tarihi: 2000. http://dataprivacylab.org/projects/identifiability/paper1.pdf. Erişim tarihi: 23 Temmuz 2023.

20. Garfinkel SL. *De-Identification of Personal Information*. 2015. doi: 10.6028/NIST.IR.8053

21. Mishra T, Wang M, Metwally AA, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. *Nat Biomed Eng*. 2020;4(12):1208-1220. doi: 10.1038/s41551-020-00640-6.

22. Ates HC, Yetisen AK, Güder F, Dincer C. Wearable devices for the detection of COVID-19. *Nat Electron*. 2021;4(1):13-14. doi: 10.1038/s41928-020-00533-1.

23. Ghiță AŞ, Florea AM. Real-time people re-identification and tracking for autonomous platforms using a trajectory prediction-based approach. *Sensors*. 2022;22(15):5856. doi: 10.3390/s22155856.

24. Seastedt KP, Schwab P, O'Brien Z, et al. Global healthcare fairness: We should be sharing more, not less, data. *PLOS Digit Heal*. 2022;1(10):e0000102. doi: 10.1371/journal.pdig.0000102.

25. Health Information Privacy. The HIPAA Privacy Rule. United States (U.S.) Department of Health and Human Services. https://www.hhs.gov/hipaa/for-professionals/privacy/index.html. Published 1996. Erişim tarihi 4 Mayıs 2023.

26. Ghazarian A, Zheng J, El-Askary H, et al. Increased Risks of Re-identification For Patients Posed by Deep Learning-Based ECG Identification Algorithms. In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE; 2021:1969-1975. doi: 10.1109/EMBC46164.2021.9630880.

27. Kancherla J. Re-identification of health data through machine learning. *SSRN Electron J*. Published online 2020. doi: 10.2139/ssrn.3794927.

28. Haley DF, Matthews SA, Cooper HLF, et al. Confidentiality considerations for use of social-spatial data on the social determinants of health: Sexual and reproductive health case study. *Soc Sci Med*. 2016;166:49-56. doi:

10.1016/j.socscimed.2016.08.009.

29. Spengler H, Prasser F. Protecting Biomedical Data Against Attribute Disclosure. In: (Eds.) RR et al., ed. *German Medical Data Sciences: Shaping Change – Creative Solutions for Innovative Medicine*. IOS Press; 2019:207-214. doi: 10.3233/SHTI190829

30. Hayes B. Uniquely Me! *Am Sci*. 2014;102(2):106-109.

31. Willemink MJ, Koszek WA, Hardell C, et al. Preparing medical imaging data for machine learning. *Radiology*. 2020;295(1):4-15. doi: 10.1148/radiol.2020192224.

32. Lee I. Analysis of insider threats in the healthcare industry: a text mining approach. *Information*. 2022;13(9):404. doi: 10.3390/info13090404.

33. Vincent J. Google scrapped the publication of 100,000 chest X-rays due to last-minute privacy problems. The Verge. https://www.theverge.com/2019/11/15/20966460/google-scrapped-publication-100000-chest-x-rays-nih-project-2017. Published 2019. Erişim tarihi 10 Mayıs 2023.

34. Attia ZI, Friedman PA, Noseworthy PA, et al. Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circ Arrhythmia Electrophysiol*. 2019;12(9). doi: 10.1161/CIRCEP.119.007284.

35. Lubarsky B. Re-identification of "anonymized" data. *Georg Law Technol Rev*. 2017;202. doi: 10.48550/arXiv.1909.09675.

36. Rocher L, Hendrickx JM, de Montjoye YA. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun*. 2019;10(1):3069. doi: 10.1038/s41467-019-10933-3.

37. El Emam K, Dankar FK, Neisa A, Jonker E. Evaluating the risk of patient re-identification from adverse drug event reports. *BMC Med Inform Decis Mak*. 2013;13(1):114. doi: 10.1186/1472-6947-13-114.

38. Richens JG, Lee CM, Johri S. Improving the accuracy of medical diagnosis with causal machine learning. *Nat Commun*. 2020;11(1):3923. doi: 10.1038/s41467-020-17419-7.

39. Matveev Y. The Problem of Voice Template Aging in Speaker Recognition Systems. In: The 15th International Conference SPECOM 2013; September 1-5, 2013; Plzeň, Czech Republic. doi: 10.1007/978-3-319-01931-4_46.

40. Manjani I, Sumerkan H, Flynn PJ, Bowyer KW. Template aging in 3D and 2D face recognition. In: 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS). IEEE; September 6-9, 2016; New York, ABD. doi:10.1109/BTAS.2016.7791202.

41. Kim HE, Kim HH, Han BK, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit Heal*. 2020;2(3):e138-e148. doi: 10.1016/S2589-7500(20)30003-0.

42. Tosoni S, Voruganti I, Lajkosz K, et al. The use of personal health information outside the circle of care: consent preferences of patients from an academic health care institution. *BMC Med Ethics*. 2021;22(1):29. doi: 10.1186/s12910-021-00598-3.

43. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ Digit Med*.

2020;3(1):118. doi: 10.1038/s41746-020-00324-0.

44. El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. Scherer RW, ed. *PLoS One*. 2011;6(12):e28071.

45. Xia W, Liu Y, Wan Z, et al. Enabling realistic health data re-identification risk assessment through adversarial modeling. *J Am Med Informatics Assoc*. 2021;28(4):744-752.

Istanbul Gelisim University Journal of Health Sciences (IGUSABDER) is indexed by TUBITAK ULAKBIM TR Index.
Web site: https://dergipark.org.tr/en/pub/igusabder
Contact: igusabder@gelisim.edu.tr