



23. Ulusal ve 6. Uluslararası
Biyoistatistik Kongresi
26-29 Ekim 2022

Ankara Üniversitesi Tıp Fakültesi,
Morfoloji Yerleşkesi

Özet Bildiri Kitabı



23. Ulusal ve 6. Uluslararası Biyoistatistik Kongresi

26-29 Ekim 2022, Ankara Üniversitesi Tıp Fakültesi

Değerli Meslektaşlarımız,

Biyoistatistik Derneği ve Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı olarak **26-29 Ekim 2022** tarihlerinde Ankara'da düzenleyeceğimiz **23. Ulusal ve 6. Uluslararası Biyoistatistik kongresi'**ne sizleri davet etmekten büyük mutluluk duyuyoruz.

Tüm dünyayı etkileyen Covid-19 Pandemisi nedeniyle geçtiğimiz yıl kongremizi çevrim içi olarak gerçekleştirebilmişken, en son 26-29 Ekim 2019'da Antalya'da düzenlenen kongrede bir araya gelebilmiştik. Üç yılın ardından tekrar buluşacak olmanın heyecanı ile düzenleme kurulu olarak bilimsel ve sosyal açıdan zengin içerikli bir kongre hazırlamak için tüm gücümüzle çalışıyoruz.

Kongrelerin bilimsel niteliğini artıran en önemli etkinliklerden birisi, alanında uzman eğitimcilerin güncel konularda kısa süre içinde bilgi paylaşımı yaptıkları kurslardır. Kongremiz kapsamında 26 Ekim 2022 tarihinde uygulamalı olarak düzenlenecek dört farklı kurs ile katılımcıların bilgi ve program kullanma becerilerini geliştirmeyi hedeflemekteyiz. 27-29 Ekim 2022 tarihlerinde ise sözlü ve poster sunumlarının yer aldığı bir bilimsel program ile sosyal gezi programı planlamaktayız.

Siz değerli meslektaşlarımızın katılımı ile bilimsel açıdan zengin ve güncel bilgilerin yer aldığı çalışmaların paylaşıldığı başarılı bir kongre geçireceğimize hiç şüphemiz yok. Birlikteliğimizi pekiştirmek ve gücümüze güç katmak için tüm meslektaşlarımızı kongremizde buluşmaya davet ediyoruz,

Saygılarımızla,

Prof. Dr. Yasemin Yavuz
Kongre Başkanı

Prof. Dr. Erdem Karabulut
Biyoistatistik Derneği Başkanı

Doç. Dr. Derya Gökmen
Kongre Sekreteri

23. Ulusal ve 6. Uluslararası Biyoistatistik Kongresi

26-29 Ekim 2022, Ankara Üniversitesi Tıp Fakültesi

Bilimsel Kurul

Profesör Alan TENNANT
Profesör Ahmet DİRİCAN
Profesör Ahmet ÖZTÜRK
Profesör Atilla Halil ELHAN
Profesör Ayşe Canan YAZICI GÜVERCİN
Profesör Beyza ŞAHİN
Profesör Bülent ÇELİK
Profesör Canan BAYDEMİR
Profesör Celal Reha ALPAR
Profesör Erdem KARABULUT
Profesör Ersin ÖĞÜŞ
Profesör Ferhan ELMALI
Profesör Gülhan TEMEL
Profesör Hafize SEZER
Profesör İmran ÖMÜRLÜ
Profesör Kadir SÜMBÜLOĞLU
Profesör Mehtap AKÇİL OK
Profesör Melody GHAHRAMANI
Profesör Peter W. ROSE
Profesör Recai YÜCEL
Profesör Sanjib BASU
Profesör Seval KUL
Profesör Sıddık KESKİN
Profesör Şengül CANGÖR
Profesör Ünal ERKORKMAZ
Profesör Vildan SÜMBÜLOĞLU
Profesör Yasemin YAVUZ
Profesör Yusuf ÇELİK

Profesör Z. Nazan ALPARSLAN
Profesör Zeki AKKUŞ
Profesör Mithat GÖNEN
Doçent Doktor Adem DOĞANER
Doçent Doktor Alex R. de LEON
Doçent Doktor Beyza DOĞANAY ERDOĞAN
Doçent Doktor Can ATEŞ
Doçent Doktor Derya GÖKMEN
Doçent Doktor Emre DEMİR
Doçent Doktor Erdal COŞGUN
Doçent Doktor Gökmen ZARARSIZ
Doçent Doktor Harika GÖZÜKARA BAĞ
Doçent Doktor İlkay DOĞAN
Doçent Doktor İlker ÜNAL
Doçent Doktor Jale KARAKAYA
Doçent Doktor Kenan KÖSE
Doçent Doktor Leman TOMAK
Doçent Doktor Lynn SLEEPER
Doçent Doktor Mehmet KARADAĞ
Doçent Doktor Mustafa Agah TEKİNDAL
Doçent Doktor Pınar GÜNEL
Doçent Doktor Selcen PEHLİVAN
Doçent Doktor Selçuk KORKMAZ
Doçent Doktor Şirin ÇETİN
Doçent Doktor Umut ÖZBEK
Yrd. Doçent Doktor Duchwan RYU
Dr. Öğr. Üyesi Tolga KASKATI
Dr. Bernd KLAUS

Kongre Başkanı

Prof. Dr. Yasemin YAVUZ, Ankara Üniversitesi

Düzenleme Kurulu

Prof. Dr. Yasemin YAVUZ, Ankara Üniversitesi

Prof. Dr. Erdem KARABULUT, Hacettepe Üniversitesi

Prof. Dr. Mehtap AKÇİL OK, Başkent Üniversitesi

Prof. Dr. Seval KUL, Gaziantep Üniversitesi

Doç. Dr. Derya GÖKMEN, Ankara Üniversitesi

Doç. Dr. İlker ÜNAL, Çukurova Üniversitesi

Dr. İrem KAR, Ankara Üniversitesi

Dr. Batuhan BAKIRARAR, Ankara Üniversitesi

23. Ulusal ve 6. Uluslararası Biyoistatistik Kongresi

26-29 Ekim 2022, Ankara Üniversitesi Tıp Fakültesi

Program

26.10.2022 Çarşamba			
09.00-17.00	Kongre Kayıt		
Salon	Prof. Dr. Lütfi Tat Konferans Salonu (Mavi Salon)	Pembe Salon	
09.30-13.00	RevMan ve R ile Uygulamalı Meta Analizi Prof. Dr. Erdem KARABULUT, Doç. Dr. Can ATEŞ	JAMOVİ ve JASP ile Veri Analitiği Doç. Dr. Beyza DOĞANAY ERDOĞAN, Doç. Dr. Mustafa Agah TEKİNDAL	
13.00-14.00	Öğle Yemeği Arası		
14.00-17.30	R ve SPSS uygulamaları ile Propensity Skor Eşleştirme Analizi Doç. Dr. Emre DEMİR	Yüksek Boyutlu Verilerin R Programında Analizi Prof. Dr. Seval KUL	
27.10.2022 Perşembe			
09.00-17.00	Kongre Kayıt		
09.30-10.45	Açılış Programı (Ord. Prof. Dr. Abdülkadir Noyan Konferans Salonu)		
10.45-11.00	Çay/Kahve Arası		
11.00-12.00	Bayesçi Yöntemlerin Klinik Denemelerdeki Uygulamaları (Ord. Prof. Dr. Abdülkadir Noyan Konferans Salonu) Konuşmacı: Prof. Dr. Mithat Gönen, Memorial Sloan Kettering Cancer Center		
12.00-13.30	Öğle Yemeği Arası		
Salon	Ord. Prof. Dr. Abdülkadir Noyan Konferans Salonu	Sarı Salon	Pembe Salon
Oturum Başkanları	Prof. Dr. Seval KUL Prof. Dr. Ayşe Canan YAZICI GÜVERCİN Doç. Dr. Didem DERİCİ YILDIRIM	Prof. Dr. Kevser Setenay ÖNER Prof. Dr. Pınar ÖZDEMİR Doç. Dr. İlkey DOĞAN	Prof. Dr. G. Nural BEKİROĞLU Prof. Dr. Duru KARASOY Doç. Dr. S. Kenan KÖSE
13.30-13.45	Genom-boyu ilişki Çalışmalarında Poligenik Risk Skorunun Makine Öğrenimi ve Derin Öğrenme Yöntemleri ile Tahmin Edilmesi Ragıp Onur ÖZTORNACI, Erdal COŞGUN, Bahar TAŞDELEN, Cemil ÇOLAK	Türkiye Covid-19 Verileri için Parçalı Fonksiyona Dayalı Büyüme Modelleri Leyla BAKACAK KARABENLİ, Serpil AKTAŞ ALTUNAY	Periton Diyaliz Hastalarında Sağkalım ve Uzunlamasına Verilerin Birlikte Modellemesi: Kişiyi Özgü Dinamik Risk Kestirimleri Merve Başol GÖKSÜLÜK, Dinçer GÖKSÜLÜK, Ergun KARAAĞAOĞLU
13.45-14.00	Recurrent Neural Network for Complex Survival Problems Pius MARTIN	Psikiyatrik Hastalıklarda, Psikopatoloji ve Çevresel Etken İlişkisinin Özyinelemesiz Model ile Değerlendirilmesi Selen Begüm UZUN, Derya GÖKMEN, Meram Can SAKA	Boylamsal ve Yaşam Verilerinin Parametrik Bileşik Modellemesinde Parametre Tahmin Yaklaşımlarının Karşılaştırılması Elif YILDIRIM, Duru KARASOY
14.00-14.15	RNA-Dizileme Verilerinde Sağkalım Algoritmalarının Kapsamlı Bir Karşılaştırması Ahu CEPHE, Ahmet SEZGİN, Necla KOÇHAN, Gözde ERTÜRK ZARARSIZ, Vahap ELDEM, Erdem KARABULUT, Gökmen ZARARSIZ	The Interaction Term and Its Graphical Interpretation In Regression Models Handan ANKARALI, Özge PASİN, Senem GÖNENÇ	Sağkalım Analizi için Veri Simülasyonu İrem KAR, Erdal COŞGUN, Atilla Halil ELHAN
14.15-14.30	Biocrates Mxp® Quant 500 Kiti Kullanılarak Elde Edilen Metabolomik Verilerinin Laboratuvarlar Arası Karşılaştırması Gözde ERTÜRK ZARARSIZ, Alexander CECIL, Jutta LINTELMANN, Gernot POSCHET, Jennifer KIRWAN, Sven SCHUCHARDT, Xue L. GUAN, Daisuke SAIGUSA, David WISHART, Jiamin ZHENG, Rupasri MANDAL, Lisa ST. JOHN-WILLIAMS, Kendra ADAMS, J. Will THOMPSON, Michael P. SNYDER, Kevin CONTREPOIS, S Songjie CHEN, Nadia ASHRAFI, Sumeyya AKYOL, Ali YILMAZ, Stewart GRAHAM, Thomas M. O'CONNELL, Karel KALECKÝ, T. Hai PHAM, Therese KOAL, Teodoro BOTTIGLIERI, Jerzy ADAMSKI, Gabi KASTENMÜLLER	Kantil Regresyon Yöntemi ile Oluşturulan Pediatrik Büyüme Eğrilerinin Model Uyumunun Worm Grafiği ile Değerlendirilmesi Eda ÇAKMAK, Ergun KARAAĞAOĞLU, Serhat KILIÇ, Pınar ÖZDEMİR	Sağkalım Analizinde Kopula Entropi Tabanlı Değişken Seçimi Aslıhan ALHAN

14.30-14.45	RNA-Dizi Verilerinde İntron Tutulumunun Belirlenmesinde Kullanılan Biyoinformatik Algoritmaların Değerlendirilmesi Esmâ Gamze AKSEL, Vahap ELDEM, Gökmen ZARARSIZ	Küçük Küme ve Periyotlu Boylamsal Çalışmalarda Değişken ve Model Seçimi Merve TURKEGÜN, Bahar TAŞDELEN, Saim YOLOĞLU	Kadaverik Böbrek Naklinde Yaşam Süresini Belirleyen Nakil Sırasında ve Sonrasındaki Önemli Değişkenler Pembe KESKİNOĞLU, Kemal OLÇA, Duygu KARAKAYA, Ebru ÖZTÜRK, Mehmet ORMAN, Timur KÖSE, Erdem KARABULUT, Ahmet DIRİCAN
14.45-15.00	Çay/Kahve Arası		
Oturum Başkanları	Prof. Dr. Canan BAYDEMİR Prof. Dr. Gökmen ZARARSIZ Doç. Dr. Pınar GÜNEL	Prof. Dr. Ahmet ÖZTÜRK Prof. Dr. Hafize SEZER Doç. Dr. Harika Gözde GÖZÜKARA BAĞ	Prof. Dr. Ertuğrul ÇOLAK Prof. Dr. Fezan MUTLU Doç. Dr. Emre DEMİR
15.00-15.15	Machine Learning Sensors for Diagnosis of COVID-19 Disease Using Routine Blood Values for Internet of Things Application Mehmet Tahir HUYUT, Andrei VELICHKO, Belyaev MAKSIM	Efor Kullanımı ve Üç Parametrelili Lojistik Madde Yanıt Teorisi Modellerinin Bilgisayar Uyarlamalı Test Performanslarının Karşılaştırılması Yusuf Kemal ARSLAN, Afra ALKAN, Atilla Halil ELHAN	Referans Aralığı Belirlemede Kullanılan Tek Değişkenli ve Çok Değişkenli Yöntemlerin Karşılaştırılması Esra Kutsal MERGEN, Sevilay KARAHAN
15.15-15.30	Makine Öğrenmesinde Sınıf Dengesizliği Sorunu için Bazı Veri Ön İşleme Yöntemlerinin Karşılaştırılması: Bir Uygulama Hülya ÖZEN, Doğan ÖZEN	Ölçme Araçlarının Boyutluluk Yapısının Belirlenmesinde Yapısal Eşitlik Modellemesine Dayalı Yöntemler Serhat HAYME, Derya GÖKMEN, Şehim KUTLAY, Ayşe Adile KÜÇÜKDEVECİ	Mamografi Görüntülerinde Farklı Ön İşleme Kombinasyonlarının Sınıflandırma Performansları Üzerine Etkisi Hanife AVCI, Jale KARAKAYA
15.30-15.45	Makine Öğrenme Yöntemlerinin Başarılarını Değerlendirirken Hangi Performans Ölçütleri Kullanılmalıdır. Yaşar SERTDEMİR, İlker ÜNAL, Hülya BİNOKAY	Doğrusal Olmayan Yapısal Eşitlik Modellemesi Yaklaşımlarının İncelenmesi ve Bir Uygulama Nur Efşan TİĞLİ, Şengül CANGÜR	Obuchowski Parametrik olmayan ROC Eğrisi Analizi ve Küme Bootstrap Yaklaşımları ile İlişkili Verilerin Değerlendirilmesi: Multipl Skleroz Hastalarında Optik Nöropati Teşhisi Üzerine Bir Uygulama Büşra EMİR, Fatma Ezgi CAN, Mustafa Agah TEKİNDAL, Ferhan ELMALI, Ertuğrul ÇOLAK
15.45-16.00	Veri Madenciliğinde Hibrit Model Yaklaşımı Batuhan BAKIRARAR, Atilla Halil ELHAN	Biyoistatistik Dersine Yönelik İstatistik Öz Yeterlilik İnanç Ölçeği, İstatistik Kaygı Ölçeği ve İstatistik Tutum Ölçeği Arasındaki Aracılık İlişkisinin Araştırılması: Bir Yapısal Eşitlik Modellemesi Çalışması (Path Analizi) Kamber KAŞALI, Senem GÖNENÇ	Sıralı Bir Yanıt Değişkenini Sınıflamada Sayısal Bir Ölçüm İçin Optimal Kesim Noktalarını Belirleme İlker ÜNAL, Yaşar SERTDEMİR, H. Esin ÜNAL
16.00-16.15	Yayın Yanılgısını Önlemede P-Eğrisi Yöntemi Arzu BAYGÜL EDEN, Alev BAKIR KAYI	Yapısal Eşitlik Modellemesi Uygulamalarının Raporlanması: Bir Kontrol Listesi Zeliha AYDIN KASAP, Burçin KURT, Kemal TURHAN	The Effectiveness of Blood Routine Parameters and Some Biomarkers as a Potential Diagnostic Tool in the Diagnosis and Prognosis of COVID-19 Disease Mehmet Tahir HUYUT, Sıddık KESKİN

28.10.2022 Cuma			
Salon	Ord. Prof. Dr. Abdülkadir Noyan Konferans Salonu	Sarı Salon	Pembe Salon
Oturum Başkanları		Prof. Dr. Atilla Halil ELHAN Prof. Dr. Şengül CANGÜR Doç. Dr. Özlem KAYMAZ	Prof.Dr. Gülşah SEYDAOĞLU Prof. Dr. Sinan SARAÇLI Doç. Dr. Mehmet KARADAĞ
09.00-09.05		Sobel ve Baron-Kenny Yöntemlerinin Performanslarının Karşılaştırılması Hülya BİNOKAY, Yaşar SERT-DEMİR	Lökosit Antijen Molekülleri Toplum Doku Tipi Sıklığı Kemal OLÇA, Ahmet KESKİNOĞLU, Duygu KARAKAYA, Ebru ÖZTÜRK, Pembe KESKİNOĞLU, Mehmet ORMAN, Timur KÖSE, Ahmet DİRİCAN, Erdem KARABULUT
09.05-09.10		Mamografi Görüntülerinde K-ortalama Kümeleme ve Canny Kenar Algılama Bölütleme Yöntemlerinin Sınıflandırma Performanslarının Karşılaştırılması Jale KARAKAYA, Hanife AVCI	Çok Boyutlu Ölçekleme Yöntemi ile Trakya'da Tarımsal Üretim Bölgelerinde Ağır Metal Yoğunluğunun Gösterimi Büşra AYDIN, Özgecan KORKMAZ AĞAOĞLU, İsmayıl Safa GURCAN
09.10-09.15		Log Multinomial Regresyon Modeline Hosmer-Lemeshow Uyum İyiliği Testinin Uyarlanması Yasemin ÖZTÜRK, Erdem KARABULUT	Lojistik Regresyon Modeli ile Kardiyovasküler Kalp Hastalığı Sınıflandırılması ve Tahmini Mehmet KIVRAK
09.15-09.20		Diyabetik Gözlerde Multifokal Elektoretinografide Halka Amplitüdünün ve Halka Amplitüdünü Etkileyen Faktörlerin Değerlendirilmesi Ayşe Büşra GÜNAY ŞENER, Hidayet ŞENER	Önceden Eğitilmiş Evrişimsel Sinir Ağlarından Elde Edilen Öznitelikler ile Sperm Morfolojisinin Julia Tabanlı Web Uygulaması ile Sınıflandırılması Hüseyin KUTLU, Emek GÜLDOĞAN, Cemil ÇOLAK
09.20-10.00	Çay/Kahve Arası		
10.00-12.00	Panel (Ord. Prof. Dr. Abdülkadir Noyan Konferans Salonu) Moderatörler: Prof. Dr. Z. Nazan ALPARSLAN, Prof. Dr. Yasemin YAVUZ		
12.30-16.30	Gezi Programı		
29.10.2022 Cumartesi			
Salon	Ord. Prof. Dr. Abdülkadir Noyan Konferans Salonu	Sarı Salon	Pembe Salon
Oturum Başkanları	Prof. Dr. Mehtap AKÇİL OK Doç. Dr. Jale KARAKAYA KARABULUT Doç. Dr. Beyza DOĞANAY ERDOĞAN	Prof. Dr. Yusuf ÇELİK Prof. Dr. Nurhan DOĞAN Doç. Dr. Cengiz BAL	Prof. Dr. İsmayıl Safa GÜRCAN Prof. Dr. S. Yavuz SANİSOĞLU Doç. Dr. Yaşar SERTDEMİR
09.30-9.45	Makine Öğrenmesi Algoritmalarının Performanslarının Farklı Koşullar Altında Karşılaştırılması: Bir Monte Carlo Benzetim Çalışması Müge COŞKUN, Osman DAĞ	dtComb: İki Sürekli Tanı Testinin Birleştirilmesine Yönelik Geliştirilmiş Kapsamlı Bir R Paketi Serra İlayda YERLİTAŞ, Serra Bersan GENGEÇ, Gözde ERTÜRK ZARARSIZ, Selçuk KORKMAZ, Gökmen ZARARSIZ	Çok Değişkenli Varyans Analizinde Kullanılan Test İstatistiklerinin Tip I Hata Oranları Bakımından Karşılaştırılması Fatma Ezgi CAN, Büşra EMİR, Ferhan ELMALİ, Can ATEŞ, Mustafa Agah TEKİNDAL
09.45-10.00	Boyut İndirgeme Uygulanan ve Uygulanmayan Verilerin Analizinde Makine Öğrenimi Yöntemlerinin Karşılaştırılması Feyza İNCEOĞLU, Cemil ÇOLAK	Propensity Skora Dayalı Eşleştirme Analizi İçin Yeni Bir R Shiny Web Uygulaması Emre DEMİR, Gülçin AYDOĞDU, Muhammed Ali YILMAZ, Yasemin YAVUZ	Kanıtsal Güç ve Literatür Bütünlüğünün p-Eğrisi ile İncelenmesi Alev BAKIR KAYI, Arzu BAYGÜL EDEN

10.00-10.15	Yüksek Boyutlu Verilerde Eksik Veri Değer Atama Yöntemlerinin Tahmin Performanslarının Aşırı Öğrenme Makineleriyle İncelenmesi Buğra VAROL, İmran ÖMÜRLÜ, Mevlüt TÜRE	R Shiny ile Zaman Serileri Analizi için İnteraktif Arayüz Tasarımı Esra GÜLTÜRK, Abdolvahap PINAR, Ahmet Kadir ARSLAN, Pinar YILMAZ	Taguchi ve Faktöriyel Deney Tasarım Metotlarının Karşılaştırılması: Lactobacillus spp. ERIC-PCR Optimizasyonu Örneği Selen YILMAZ IŞIKHAN, Ceren ÖZKUL
10.15-10.30	Sınıflamada Kullanılan Veri Madenciliği Yöntemlerinin Performanslarının D Vitamini Veri Seti Üzerinde Karşılaştırılması Özlem ARIK, İnci ARIKAN, Erdem KARABULUT	Twowaytests: İki Yönlü Bağımsız Grup Tasarımı İçin Bir R Paketi Muhammed Ali YILMAZ, Osman DAĞ, Samaradasa WEERAHANDI, Malwane M.A. ANANDA	Eğilim Skoru Eşleştirmesinde Greedy ve Optimal Yöntemlerin Karşılaştırılması Güliden HAKVERDİ, Timur KÖSE, Cemil ÇOLAK
10.30-10.45	Bulanık Kümeleme Yöntemi ile Biyoistatistik Dersi Alan Öğrencilerin Değerlendirilmesi Senem GÖNENÇ, Kamber KAŞALI, Özge PAŞIN	Kantil Regresyon Yaklaşımı için İnteraktif Bir Arayüz Tasarımı Abdolvahap PINAR, Ahmet Kadir ARSLAN	KOAH Hastalarında Fonksiyonel Kapasite Üzerindeki Etkilerin İstatistiksel Yöntemlerle Analizi Sinan SARAÇLI, Aydın BALCI, Berkay TUNCA
10.45-11.00	Çay/Kahve Arası		
Oturum Başkanları	Prof. Dr. Ferhan ELMALI Doç. Dr. İlker ÜNAL Doç. Dr. Selen YILMAZ IŞIKHAN	Prof. Dr. Ünal ERKORKMAZ Doç. Dr. Selçuk KORKMAZ Doç. Dr. M. Agah TEKİNDAL	Prof. Dr. Erdem KARABULUT Doç. Dr. Selcen YÜKSEL Doç. Dr. Can ATEŞ
11.00-11.15	Covid-19 Tanısında Konvolüsyonel Sinir Ağları Mimarilerinin Karşılaştırılması Işıl ÜNALDI, Leman TOMAK	Performance Analysis of Simulated Data Sets by Machine Learning Algorithms Senem KOC, Leman TOMAK, Erdem KARABULUT	Holstein Irkı İneklerde Güç Doğum Üzerinde Etkili Olan Faktörlerin Belirlenmesinde FP Growth Algoritmasının Kullanımı Elif ÇELİK GÜRBULAK, Murat ABAY, Aytaç AKÇAY, Funda İPEKTEN, Gökmen ZARARSIZ, Esmâ Gamze AKSEL
11.15-11.30	Görüntü İşleme Yöntemlerinin COVID-19 Pozitif Tespitinde Klasik Veri Madenciliği Yöntemleri ile Sınıflandırmaya Etkisi Fatma Gül KURT, Jale KARAKAYA	Propensity Skora Dayalı Eşleştirme Analizinde Optimum Caliper Belirlenmesi İçin Bir Benzetim Çalışması Gülçin AYDOĞDU, Muhammed Ali YILMAZ, Emre DEMİR, Yasemin YAVUZ	Kardiyoverter Defibrilatör (ICD) İmplantasyonu Yapılan Hastalarda Yaşam Durumu Tahmin Modellemesi Hatice AKTAŞ GÖKÇE, Amine BAYRAKLI, Emre YAŞAR
11.30-11.45	Diagnosis and Prognosis of COVID-19 Disease Using Routine Blood Values and LogNet Neural Network Mehmet Tahir HUYUT, Andrei VELICHKO	Assessing and comparison of various biomarker scenarios for guiding treatment selection: Simulation results and recommendations Hülya KOÇYİĞİT	Erken Evre Meme Kanseri Hastalarına ait Hazard Faktörlerinin Uygun Çeşitli Model Sonuçlarının Karşılaştırılması Nural BEKİROĞLU, Ayşe ÜLGİN, Emrah Gökay ÖZGÜR, Sinan UZUN, Wentian LI
11.45-12.00	Önerilen Derin Öğrenme Mimarisi ile MR Görüntülerine Dayalı Alzheimer Hastalığının Sınıflandırılması Mehmet KIVRAK, Cemil ÇOLAK	BKSY: Bilgi Keşfi Süreci Yazılımı Ahmet Kadir ARSLAN, Cemil ÇOLAK	MEVARYA Mutlu UMAROĞLU
12.00-12.15	Çay/Kahve Arası		
12.15-13.00	Kapanış Konuşmaları (Ord. Prof. Dr. Abdülkadir Noyan Konferans Salonu)		

SB - 1	Yapısal Eşitlik Modellemesi Uygulamalarının Raporlanması: Bir Kontrol Listesi	1
	Zeliha AYDIN KASAP, Burçin KURT, Kemal TURHAN	
SB - 2	Kadaverik Böbrek Naklinde Yaşam Süresini Belirleyen Nakil Sırasında ve Sonrasındaki Önemli Değişkenler	2
	Pembe KESKİNOĞLU, Kemal OLÇA, Duygu KARAKAYA, Ebru ÖZTÜRK, Mehmet ORMAN, Timur KÖSE, Erdem KARABULUT, Ahmet DIRİCAN	
SB - 3	Assessing and Comparison of Various Biomarker Scenarios for Guiding Treatment Selection: Simulation Results and Recommendations	3
	Hülya KOÇYİĞİT	
SB - 4	Boyut İndirgeme Uygulanan ve Uygulanmayan Verilerin Analizinde Makine Öğrenimi Yöntemlerinin Karşılaştırılması	4
	Feyza INCEOĞLU, Cemil ÇOLAK	
SB - 5	Holstein Irkı İneklerde Güç Doğum Üzerinde Etkili Olan Faktörlerin Belirlenmesinde FP Growth Algoritmasının Kullanımı	5
	Elif ÇELİK GÜRBÜLAK, Murat ABAY, Aytaç AKÇAY, Funda İPEKTEN, Gökmen ZARARSIZ, Esmâ Gamze AKSEL	
SB - 6	Mamografi Görüntülerinde Farklı Ön İşleme Kombinasyonlarının Sınıflandırma Performansları Üzerine Etkisi	6
	Hanife AVCI, Jale KARAKAYA	
SB - 7	SHINY ile Zaman Serileri Analizi İçin İnteraktif Arayüz Tasarımı	7
	Esra GÜLTÜRK, Abdolvahap PINAR, Ahmet Kadir ARSLAN, Pınar YILMAZ	
SB - 8	Kardiyoverter Defibrilatör (Icd) İmplantasyonu Yapılan Hastalarda Yaşam Durumu Tahmin Modellemesi	8
	Hatice AKTAŞ GÖKÇE, Amine BAYRAKLI, Emre YAŞAR	
SB - 9	Sağkalım Analizi için Veri Simülasyonu	9
	İrem KAR, Erdal COŞGUN, Atilla Halil ELHAN	
SB - 10	R Shiny ile Zaman Serisi Analizi İçin İnteraktif Arayüz Tasarımı	10
	Esra GÜLTÜRK, Abdolvahap PINAR, Ahmet Kadir ARSLAN, Pınar YILMAZ	
SB - 11	Periton Diyaliz Hastalarında Sağkalım ve Uzunlmasına Verilerin Birlikte Modellemesi: Kişiyi Özgü Dinamik Risk Kestirimleri	11
	Merve Başol GÖKSÜLÜK, Dinçer GÖKSÜLÜK, Ergun KARAAĞAOĞLU	
SB - 12	Boylamsal ve Yaşam Verilerinin Parametrik Bileşik Modellemesinde Parametre Tahmin Yaklaşımlarının Karşılaştırılması	12
	Elif YILDIRIM, Duru KARASOY	
SB - 13	Önerilen Derin Öğrenme Mimarisi ile MR Görüntülerine Dayalı Alzheimer Hastalığının Sınıflandırılması	13
	Mehmet KIVRAK, Cemil ÇOLAK	
SB - 14	Yüksek Boyutlu Verilerde Eksik Veri Değer Atama Yöntemlerinin Tahmin Performanslarının Aşırı Öğrenme Makineleriyle İncelenmesi	14
	Buğra VAROL, İmran ÖMÜRLÜ, Mevlüt TÜRE	

SB - 15	Biocrates Mxp® Quant 500 Kiti Kullanılarak Elde Edilen Metabolomik Verilerinin Laboratuvarlar Arası Karşılaştırması	15
	Gözde ERTÜRK ZARARSIZ, Gözde ERTÜRK ZARARSIZ, Alexander CECIL, Jutta LINTELMANN, Gernot POSCHET, Jennifer KIRWAN, Sven SCHUCHARDT, Xue L. GUAN, Daisuke SAIGUSA, David WISHART, Jiamin ZHENG, Rupasri MANDAL, Lisa ST. JOHN-WILLIAMS, Kendra ADAMS, J. Will THOMPSON, Michael P. SNYDER, Kevin CONTREPOIS, S Songjie CHEN, Nadia ASHRAFI, Sumeyya AKYOL, Ali YILMAZ, Stewart GRAHAM, Thomas M. O'CONNELL, Karel KALECKÝ, Karel KALECKÝ, T. Hai PHAM, Therese KOAL, Teodoro BOTTIGLIERI, Teodoro BOTTIGLIERI, Jerzy ADAMSKI, Gabi KASTENMÜLLER	
SB - 16	Sağkalım Analizinde Kopula Entropi Tabanlı Değişken Seçimi	17
	Aslıhan ALHAN	
SB - 17	Propensity Skora Dayalı Eşleştirme Analizinde Optimum Caliper Belirlenmesi İçin Bir Benzetim Çalışması	18
	Gülçin AYDOĞDU, Muhammed Ali YILMAZ, Emre DEMİR, Yasemin YAVUZ	
SB - 18	Görüntü İşleme Yöntemlerinin COVID-19 Pozitif Tespitinde Klasik Veri Madenciliği Yöntemleri ile Sınıflandırmaya Etkisi	20
	Fatma Gül KURT, Jale KARAKAYA	
SB - 19	Covid-19 Tanısında Konvolüsyonel Sinir Ağları Mimarilerinin Karşılaştırılması	21
	İşıl ÜNALDI, Leman TOMAK	
SB - 20	Taguchi ve Faktöriyel Deney Tasarım Metotlarının Karşılaştırması: Lactobacillus spp. ERIC-PCR Optimizasyonu Örneği	22
	Selen YILMAZ İŞIKHAN, Ceren ÖZKUL	
SB - 21	Eğilim Skoru Eşleştirmesinde Greedy ve Optimal Yöntemlerin Karşılaştırılması	23
	Güliden HAKVERDİ, Timur KÖSE, Cemil ÇOLAK	
SB - 22	Veri Madenciliğinde Hibrit Model Yaklaşımı	24
	Batuhan BAKIRARAR, Atilla Halil ELHAN	
SB - 24	Küçük Küme ve Periyotlu Boylamsal Çalışmalarda Değişken ve Model Seçimi	25
	Merve TURKEGÜN, Bahar TAŞDELEN, Saim YOLOĞLU	
SB - 25	MEVARYA	26
	Mutlu UMAROĞLU	
SB - 26	Makine Öğrenmesi Algoritmalarının Performanslarının Farklı Koşullar Altında Karşılaştırılması: Bir Monte Carlo Benzetim Çalışması	27
	Müge COŞKUN, Osman DAĞ	
SB - 27	Web Tabanlı Açıklanabilir Yapay Zeka (XAI) arayüzü: Akut Apandisit Yorumlanabilir Tahmini İçin Bir Uygulama	28
	Fatma Hilal YAĞIN, İlhami SEL, İpek BALIKÇI ÇİÇEK, Davut HANBAY, Cemil ÇOLAK, Sami AKBULUT	
SB - 28	Kantil Regresyon Yaklaşımı İçin İnteraktif Bir Arayüz Tasarımı	29
	Abdulvahap PINAR, Ahmet Kadir ARSLAN	
SB - 29	RNA-Dizileme Verilerinde Sağkalım Algoritmalarının Kapsamlı Bir Karşılaştırması	30
	Ahu CEPHE, Ahmet SEZGIN, Necla KOÇHAN, Gözde ERTÜRK ZARARSIZ, Vahap ELDEM, Erdem KARABULUT, Gökmen ZARARSIZ	

SB - 30	Sıralı Bir Yanıt Değişkenini Sınıflamada Sayısal Bir Ölçüm İçin Optimal Kesim Noktalarını Belirleme	31
	İlker UNAL, Yaşar SERTDEMİR, H. Esin ÜNAL	
SB - 33	Efor Kullanımlı ve Üç Parametrelili Lojistik Madde Yanıt Teorisi Modellerinin Bilgisayar Uyarlamalı Test Performanslarının Karşılaştırılması	32
	Yusuf Kemal ARSLAN, Afra ALKAN, Atilla Halil ELHAN	
SB - 34	dtComb: İki Sürekli Tanı Testinin Birleştirilmesine Yönelik Geliştirilmiş Kapsamlı Bir R Paketi	33
	Serra İlayda YERLİTAŞ, Serra İlayda YERLİTAŞ, Serra Bersan GENGEÇ, Gözde ERTÜRK ZARARSIZ, Gözde ERTÜRK ZARARSIZ, Selçuk KORKMAZ, Gökmen ZARARSIZ, Gökmen ZARARSIZ, Gökmen ZARARSIZ	
SB - 35	RNA-Dizi Verilerinde İntron Tutulumunun Belirlenmesinde Kullanılan Biyoinformatik Algoritmaların Değerlendirilmesi	34
	Esma Gamze Akse, Vahap ELDEM, Gökmen ZARARSIZ	
SB - 36	Biyoistatistik Dersine Yönelik İstatistik Öz Yeterlilik İnanç Ölçeği, İstatistik Kaygı Ölçeği ve İstatistik Tutum Ölçeği Arasındaki Aracılık İlişkisinin Araştırılması: Bir Yapısal Eşitlik Modellemesi Çalışması (Path Analizi)	35
	Kamber KAŞALI, Senem GÖNENCİ	
SB - 37	Bulanık Kümeleme Yöntemi ile Biyoistatistik Dersi Alan Öğrencilerin Değerlendirilmesi	36
	Senem GÖNENCİ, Kamber KAŞALI, Özge PAŞIN	
SB - 38	Erken Evre Meme Kanseri Hastalarına Ait Hazard Faktörlerinin Uygulanan Çeşitli Model Sonuçlarının Karşılaştırılması	37
	Nural BEKİROĞLU, Ayse ULGEN, Emrah Gökay ÖZGÜR, Sinan UZUN, Wentian LI	
SB - 39	Makine Öğrenme Yöntemlerinin Başarılarını Değerlendirirken Hangi Performans Ölçütleri Kullanılmalıdır.	38
	Yaşar SERTDEMİR, İlker ÜNAL, Hülya BİNOKAY	
SB - 40	Propensity Skora Dayalı Eşleştirme Analizi İçin Yeni Bir R Shiny Web Uygulaması (A New R Shiny Web Application for Propensity Score Matching Analysis)	39
	Emre DEMİR, Gülçin AYDOĞDU, Muhammed Ali YILMAZ, Yasemin YAVUZ	
SB - 41	Kantil Regresyon Yöntemi ile Oluşturulan Pediatrik Büyüme Eğrilerinin Model Uyumunun Worm Grafiği ile Değerlendirilmesi	40
	Eda ÇAKMAK, Ergun KARAAĞAOĞLU, Serhat KILIÇ, Pınar ÖZDEMİR	
SB - 42	gw-Exp-Tool: A Snakemake Workflow for Genome-wide Characterization and Expression Profiling of Transcription Factors or Gene Families	41
	Vahap ELDEM	
SB - 43	Performance Analysis of Simulated Data Sets by Machine Learning Algorithms	42
	Senem KOC, Leman TOMAK, Erdem KARABULUT	
SB - 44	Makine Öğrenmesinde Sınıf Dengesizliği Sorunu İçin Bazı Veri Ön İşleme Yöntemlerinin Karşılaştırılması: Bir Uygulama	43
	Hülya ÖZEN, Doğukan ÖZEN	
SB - 45	Twowaytests: İki Yönlü Bağımsız Grup Tasarımı İçin Bir R Paketi	44
	Muhammed Ali YILMAZ, Osman DAĞ, Samaradasa WEERAHANDI, Malwane M.A. ANANDA	

SB - 46	Referans Aralığı Belirlemede Kullanılan Tek Değişkenli ve Çok Değişkenli Yöntemlerin Karşılaştırılması	45
	Esra Kutsal MERGEN, Sevilay KARAHAN	
SB - 47	Yayın Yanıllığını Önlemede P-Eğrisi Yöntemi	46
	Arzu BAYGÜL EDEN, Alev Bakır KAYI	
SB - 48	Kanıtsal Güç ve Literatür Bütünlüğünün p-Eğrisi ile İncelenmesi	47
	Alev BAKIR KAYI, Arzu BAYGÜL EDEN	
SB - 49	The Effectiveness of Blood Routine Parameters and Some Biomarkers as a Potential Diagnostic Tool in the Diagnosis and Prognosis of COVID-19 Disease	48
	Mehmet Tahir HUYUT, Sıddık KESKİN	
SB - 50	Çok Değişkenli Varyans Analizinde Kullanılan Test İstatistiklerinin Tip I Hata Oranları Bakımından Karşılaştırılması	50
	Fatma Ezgi CAN, Büşra EMİR, Ferhan ELMALI, Can ATEŞ, Mustafa Ağah TEKİNDAL	
SB - 51	Obuchowski Parametrik olmayan ROC Eğrisi Analizi ve Küme Bootstrap Yaklaşımları ile İlişkili Verilerin Değerlendirilmesi: Multipl Skleroz Hastalarında Optik Nöropati Teşhisi Üzerine Bir Uygulama	51
	Büşra EMİR, Fatma Ezgi CAN, Mustafa Ağah TEKİNDAL, Ferhan ELMALI, Ertuğrul ÇOLAK	
SB - 52	Genom-boyu İlişki Çalışmalarında Poligenik Risk Skorunun Makine Öğrenimi ve Derin Öğrenme Yöntemleri ile Tahmin Edilmesi	52
	Ragıp Onur ÖZTORNACI, Erdal COŞGUN, Bahar TAŞDELEN, Cemil ÇOLAK	
SB - 53	Doğrusal Olmayan Yapısal Eşitlik Modellemesi Yaklaşımlarının İncelenmesi ve Bir Uygulama	53
	Nur Efsan TIĞLI, Şengül CANGÜR	
SB - 54	Machine Learning Sensors for Diagnosis of COVID-19 Disease Using Routine Blood Values for Internet of Things Application	54
	Mehmet Tahir HUYUT, Andrei VELICHKO, Belyaev MAKSİM	
SB - 55	Sınıflamada Kullanılan Veri Madenciliği Yöntemlerinin Performanslarının D Vitamini Veri Seti Üzerinde Karşılaştırılması	56
	Özlem ARIK, İnci ARIKAN, Erdem KARABULUT	
SB - 56	Psikiyatrik Hastalıklarda, Psikopatoloji ve Çevresel Etken İlişkisinin Özyinelemesiz Model ile Değerlendirilmesi	57
	Selen Begüm UZUN, Derya GÖKMEN, Meram Can SAKA	
SB - 57	Diagnosis and Prognosis of COVID-19 Disease Using Routine Blood Values and LogNet Neural Network	58
	Mehmet Tahir HUYUT, Andrei VELICHKO	
SB - 58	Ölçme Araçlarının Boyutluluk Yapısının Belirlenmesinde Yapısal Eşitlik Modellemesine Dayalı Yöntemler	59
	Serhat HAYME, Derya GÖKMEN, Şehim KUTLAY, Ayşe Adile KÜÇÜKDEVECİ	
SB - 59	The Interaction Term and Its Graphical Interpretation In Regression Models	60
	Handan ANKARALI, Özge PAŞIN, Senem GÖNENÇ	

SB - 60	
BKSY: Bilgi Keşfi Süreci Yazılımı	61
Ahmet Kadir ARSLAN, Cemil ÇOLAK	
SB - 61	
KOAH Hastalarında Fonksiyonel Kapasite Üzerindeki Etkilerin İstatistiksel Yöntemlerle Analizi	62
Sinan SARAÇLI, Aydın BALCI, Berkalp TUNCA	
SB - 62	
Türkiye Covid-19 Verileri için Parçalı Fonksiyona Dayalı Büyüme Modelleri	63
Leyla BAKACAK KARABENLİ, Serpil AKTAŞ ALTUNAY	
SB - 63	
Recurrent Neural Network for Complex Survival Problems	64
Pius MARTIN, Nihal ATA TUTKUN	
KS - 1	
Önceden Eğitilmiş Evrimsel Sinir Ağlarından Elde Edilen Öznitelikler ile Sperm Morfolojisinin Julia Tabanlı Web Uygulaması ile Sınıflandırılması	65
Hüseyin KUTLU, Emek GÜLDOĞAN, Cemil ÇOLAK	
KS - 2	
Lökosit Antijen Molekülleri Toplum Doku Tipi Sıklığı	66
Kemal OLÇA, Ahmet KESKİNOĞLU, Duygu KARAKAYA, Ebru ÖZTÜRK, Pembe KESKİNOĞLU, Mehmet ORMAN, Timur KÖSE, Ahmet DİRİCAN, Erdem KARABULUT	
KS - 3	
Mamografi görüntülerinde K-ortalamlar kümeleme ve Canny kenar algılama bölütleme yöntemlerinin sınıflandırma performanslarının karşılaştırılması	67
Jale KARAKAYA, Hanife AVCİ	
KS - 4	
Log Multinomial Regresyon Modeline Hosmer-Lemeshow Uyum İyiliği Testinin Uyarlanması	68
Yasemin ÖZTÜRK, Erdem KARABULUT	
KS - 5	
Sobel ve Baron-Kenny Yöntemlerinin Performanslarının Karşılaştırılması	69
Hülya BİNOĞLU, Yaşar SERTDEMİR	
KS - 6	
Lojistik Regresyon Modeli ile Kardiyovasküler Kalp Hastalığı Sınıflandırılması ve Tahmini	70
Mehmet KIVRAK	
KS - 7	
Çok Boyutlu Ölçekleme Yöntemi ile Trakya'da Tarımsal Üretim Bölgelerinde Ağır Metal Yoğunluğunun Gösterimi	71
Büşra AYDIN, Özgecan KORKMAZ AĞAOĞLU, İsmayıl SAFA GURCAN	
KS - 8	
Diyabetik Gözlerde Multifokal Elektoretinografide Halka Amplitüdünün ve Halka Amplitüdünü Etkileyen Faktörlerin Değerlendirilmesi	72
Ayşe Büşra GÜNAY ŞENER, Hidayet SENER	

Yapısal Eşitlik Modellemesi Uygulamalarının Raporlanması: Bir Kontrol Listesi

¹Zeliha AYDIN KASAP, ²Burçin KURT, ³Kemal TURHAN

¹Karadeniz Teknik Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı

²Karadeniz Teknik Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı

³Karadeniz Teknik Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı

Amaç: Bu çalışmada, Yapısal Eşitlik Modellemesi (YEM) uygulamalarında içerik, örneklem boyutu, eksik verilerin yönetimi, modellerin özellikleri ve tanımlanması, parametre tahmin yöntemi seçimleri ile model performanslarında uygun uyum indeksleri ve etkilendiği durumların göz önüne alınarak raporlamada yapılan ve sık karşılaşılan hataların azaltılmasına yönelik araştırmacı/okuyucu/değerlendiricilere rehberlik edecek bir kontrol listesinin geliştirilmesi amaçlanmıştır.

Gereç ve Yöntem: YEM uygulamalarında araştırma sorularının açık ve yalın bir dil ile belirlenmesi ve geliştirilecek ölçüm/yapısal modelin kavramsal çerçevesinin teorik temele dayandırılması oldukça önemlidir. Çalışmanın başında sürece hakim olunması, çalışmanın kalite standartlarını arttırmada ve süreç yönetiminin hızlandırılmasında etkin rol oynamaktadır. Literatürde, YEM için ihtiyaç duyulan örneklem sayısının, 5 ile 20 gözlem arasında değiştiği görülmektedir. Bunun yanında çalışmalarda eksik verilerin üstesinden nasıl geldiği konusu da kapsamlı bir tartışmayı içermelidir. Bu çalışmada, literatürde gözlenen kontrol listelerindeki uygulama basamakları detaylıca incelenmiş ve geliştirilen kontrol listesinin ana hatları modelin tasarımı, modelin değerlendirilmesi, modelin modifikasyonu ve YEM tabanlı çalışmaların yorumlanması/raporlanması olmak üzere 4 temel başlık altında listelenerek güncellenmiştir.

Bulgular: YEM uygulamalarında, varsayımların kontrol edilmemesi, önsel modelin teorik temele dayandırılmaması, ölçüm sonuçlarının geçerli ve güvenilir ölçeklerle elde edilmemesi, modelin teorisinde girdi matrisine alınacak elemanlar ile çıktı matrisinde oluşacak duruma dikkat edilmemesi ve sonucunda ilgili metriklerin uygun şekilde hesaplanamaması (underidentified/overidentified) sık görülen hatalardan bazılarıdır. Ayrıca, YEM modeli için yapılan açıklayıcı ve doğrulayıcı faktör analizlerinde aynı veri setinin kullanılması sonucu uyum indekslerinin yanıltıcı sonuçlar vermesi ve modelde uzman görüşüne başvurulmadan modifikasyon indekslerinin iyileştirilmeye çalışılması YEM uygulamalarında göz ardı edilen durumlar arasında gelmektedir. İstatistiksel bir model oluşturma sürecinin amacının karmaşık veri yapısını basitleştirmek olduğu düşünüldüğünde, değişken sayısını azaltmaya çalışmamış bir modelin parsimoni ilkesi açısından daha az değerli olduğu unutulmamalıdır.

Sonuç: YEM kavramı, ölçüm, regresyon, yol analizi ve faktör modelleri dahil olmak üzere çok geniş bir model yelpazesini kapsar. Bu nedenle, araştırmacıların yöntemi anlayabilmeleri ve uygun şekilde raporlayabilmeleri için önemli düzeyde istatistiksel bilgiye sahip olmaları gerekmektedir. Bu çalışma, karmaşıklığı nedeniyle YEM modellerinin her yönünü kapsayamasa da, bir YEM çalışmasının; araştırmacılar tarafından ele alınmasında, gerek çalışmanın zaman yönetimine gerekse raporlamaların standart ve etkin bir şekilde gerçekleştirilmesine katkı sağlayacağı düşünülmektedir. Böylece, bu çalışma, YEM tabanlı uygulamaların araştırma kalitesinin iyileştirilmesine ve raporlama sürecinde modelin bir bütün halinde, kontrol listesindeki adımlar ile değerlendirmesinde araştırmacılara yardımcı olacaktır.

Anahtar Kelimeler: Faktör Analizi, Yapısal Eşitlik Modellemesi, Kontrol Listesi

Kadaverik Böbrek Naklinde Yaşam Süresini Belirleyen Nakil Sırasında ve Sonrasındaki Önemli Değişkenler

¹Pembe KESKİNOĞLU, ²Kemal OLÇA, ³Duygu KARAKAYA, ⁴Ebru ÖZTÜRK, ⁵Mehmet ORMAN, ⁶Timur KÖSE, ⁷Erdem KARABULUT, ⁸Ahmet DİRİCAN

¹Dokuz Eylül Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD

²İstanbul Üniversitesi Cerrahpaşa Lisansüstü Eğitim Enstitüsü, Biyoistatistik ve Tıbbi Bilişim AD

³Ege Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD

⁴Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik AD

⁵Ege Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD

⁶Ege Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD

⁷Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik AD

⁸İstanbul Üniversitesi Cerrahpaşa Lisansüstü Eğitim Enstitüsü, Biyoistatistik ve Tıbbi Bilişim AD

Amaç: Son dönem böbrek yetmezliği gelişmiş hastalarda temel hedef tedavi, böbrek naklidir. Ülkemizde ve dünyada organ bekleme listelerinin çoğunu böbrek nakli bekleyen hastalar oluşturmaktadır. Ülkemizde organ nakli bekleyen hastaların %86.4'ü böbrek beklemektedir ve kadavra donör sayısı diğer ülkelerden çok daha sınırlıdır. Optimal eşleştirme için öncelikle klinik gözlem ve kararlar ön plana çıkmıştır. Sonradan doku-organ reddi gibi başarısızlık sonuçları biyoistatistik yaklaşımlarla değerlendirilmeye başlanmıştır. Konu ile ilgili biyoistatistik değerlendirmeler, nakilli böbreğin yaşam süresini etkili değişkenlerle birlikte inceleyen cox regresyon analizi ile yapılır. Böbrek naklinde red veya olumsuz sonuçların tahminlenmesi, alıcı, verici ve doku tipi ve klinik özelliklerle gerçekleştirilir. Bu çalışmada kadaverik organ nakli yapılmış çok merkez verileri retrospektif değerlendirilerek organ veya hasta ölümüne neden olan eşleştirme ile ilgili değişkenlerin etkisi araştırılmıştır.

Gereç ve Yöntem: Üç büyük ilde (İzmir, İstanbul ve Ankara) kadaverik nakil sayısı yüksek olan 9 organ nakli merkezinin verileri (hardcopy hasta dosyaları, elektronik ortamda ulaşılan veriler) kaydedilmiştir. Merkezler arası farklı kayıt özelliği olabilecek verilerin girişleri için standardizasyonu eğitimler yapılmıştır. Amerika ve Avrupa organ nakil sistemleri (UNOS ve Eurotransplant) incelenmiş, klinik alandaki deneyimler ve ulaşılabilen böbrek nakline ait veriler ile cinsiyet, yaş, birden fazla transplantasyon geçirme durumu, nakil öncesi hipertansiyon ve diyabet varlığı, nakil sonrası gecikmiş greft fonksiyonu yaşanması durumu ve HLA doku uyum değişkenlerinin nakil edilen organın yaşam süresine etkileri araştırılmıştır. Açıklayıcı değişkenler stepwise (adım adım) seçme yöntemi ile modele dahil edilip çıkarılarak yaşam süresine en anlamlı etki eden en anlamlı açıklayıcı değişkenler araştırılmıştır. Bu çalışmada sunulan veriler proje verilerinin tamamını içermemektedir.

Bulgular: Üç ilde dokuz organ nakli merkez verilerinden elde edilen 645 hastanın verisi analiz edilmiştir. Kadaverik böbrek nakli yapılan bu hastaların yaş ortalaması 40,3 yıldır ve 74'ü 18 yaş altında (çocuk), 73'ü 35-50 yaş aralığında ve 413'ü 50 yaş üzerinde olduğu saptandı. Alıcı hastaların 201'inde nakil öncesi hipertansiyon ve 48'inde nakil öncesi diyabet tanısı bulunmaktadır. Nakil anı ve hemen sonrası gelişen gecikmiş greft fonksiyonu 137 hastada gerçekleşmiştir, 29 alıcının birden fazla nakil geçirmiş olduğu belirlenmiştir. Parametrik modeller ve Cox oransal tehlikeler modelleri uygulanarak, alıcının gecikmiş greft fonksiyonu yaşanması ve yaş artışının organ yaşam süresini açıklamada diğer açıklayıcı değişkenlere göre daha etkili olduğu bulunmuştur.

Sonuç: Yalnızca kadaverik böbrek nakilli hastaların verilerinden eşleştirme için önemli olan değişkenlerin dahil edildiği biyoistatistik modellerle analiz edildiğinde, yaş ve gecikmiş greft fonksiyon varlığı değişkenleri böbrek yaşam süresi için önemli bulunmuştur. Ancak bu değişkenler klinik bilgi ile de birlikte değerlendirilmelidir. Gecikmiş greft fonksiyonu ara sonuç ise neden-sonuç ilişkileri tekrar değerlendirilerek analiz edilmelidir. Teşekkür: Bu çalışma TÜBİTAK-SBAG- 318S164 Nolu proje verilerinden gerçekleştirilmiştir.

Anahtar Kelimeler: Kadaverik Böbrek Nakli, Sağkalım analizi, Cox regresyon

SB – 3

Assessing and Comparison of Various Biomarker Scenarios for Guiding Treatment Selection: Simulation Results and Recommendations

Hülya KOÇYİĞİT

Karamanoglu Mehmetbey University

Aim: The use of treatment selection biomarkers is essential for assessing whether or not a patient improves after receiving a certain treatment. Oncology has performed a lot of research on tumor markers throughout the years, however, few of these markers have been shown to be useful medically. Therefore, it is crucial to identify the appropriate therapy biomarker to verify the patient's benefits. This study suggests a set of descriptive and inferential techniques for assessing individual markers and evaluating potential markers.

Materials and Methods: It is frequently required to take into account factors related to the target marker when assessing these markers. This study determines the best decision rule based on the marker to categorize patients following the benefits of their treatments, and we characterize the marker's performance using the associated classification accuracy as well as the classifier's overall distribution.

Results: Across this paper, Monte Carlo simulation, which imitates breast cancer datasets, has been performed to illustrate treatment selection based on the biomarkers.

Conclusion: This paper is a rare study in observational studies investigating and presenting summary measures for MC simulation based on the cancer markers. Few studies in the understanding of biomarker rely on observational studies instead of randomized control trial. But, in a non-randomized observational study, researchers have no control over the treatment. Thus, this paper performs various simulation scenario to illustrate confounding effects when used in observational data sets. Thus, generating different treatment, outcome or both have been performed significantly different impact on treatment selection.

Keywords: Treatment selection, biomarker, observational study

Boyut İndirgeme Uygulanan ve Uygulanmayan Verilerin Analizinde Makine Öğrenimi Yöntemlerinin Karşılaştırılması

¹Feyza INCEOĞLU, ²Cemil ÇOLAK

¹İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Malatya, Türkiye

²İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Malatya, Türkiye

Amaç: Çok boyutlu verilerde araştırmacılar veri seti hakkında bilgi sahibi olmadığı durumlarda veriyi yorumlayabilmek için boyut indirgeme yöntemlerini kullanmaktadır. Boyut indirgeme yöntemleri ile çok boyutlu verilerde daha küçük yapılar elde edilerek veri hakkında daha kolay yorumlar yapılmaktadır. Temel amaç veri kaybını minimum düzeyde tutarak veriden yüksek düzeyde bilgi sağlamaktır. Boyut indirgeme yöntemi için en çok tercih edilen yöntemlerden biri de “Temel Bileşenler Analizi”dir. Gerçek veri ile izdüşümleri arasındaki ortalama uzaklık minimum yapılarak analiz gerçekleştirilmekte ve çok boyutlu veriler küçük boyutlu homojen yapılara dönüştürülmektedir. Bu çalışmanın amacı açık kaynak kodlu bir veri setinde boyut indirgeme yöntemlerinden olan temel bileşenler analiz yöntemi uygulanan ve uygulanmayan verilerde Random Forest ve Destek Vektör Makinesi yöntem sonuçlarının doğruluk oranlarını karşılaştırmaktır.

Gereç ve Yöntem: Çalışmada Kaggle kardiyovasküler veri seti kullanılmıştır. 462 bireyden oluşan veri setine ilk olarak boyut indirgeme analizi uygulanmadan önce Random Forest (RF) ve Destek Vektör Makinesi (SVM) doğruluk oranları hesaplanmıştır. Daha sonra temel bileşenler analizi yöntemi (PCA) uygulanarak boyut indirgeme analizi uygulanmıştır. PCA sonrası Random Forest (RF) ve Destek Vektör Makinesi (SVM) doğruluk oranları hesaplanmıştır. Elde edilen sonuçlar karşılaştırılmıştır. Bütün hesaplamalarda Python programlama dili ve paketleri kullanılmıştır.

Bulgular: Çalışmada kullanılan veri setinde yer alan katılımcılara ait yaş ortalaması 42.82 ± 14.61 standart sapmadır. Katılımcıların 302’sinde (%64.4) kardiyovasküler hastalık bulunmaz iken 160’ında (%34.6) bulunmaktadır. Veri setinde kayıp gözlem bulunmamaktadır. Temel bileşenler analizi (PCA) sonucunda Bartlett Küresellik test sonucu 936.247 ($p=0.001<0.05$) olarak ve Kaiser-Meyer-Olkin (KMO) test değeri ise 0.69 olarak hesaplanmıştır. PCA %86.6 açıklanan varyans oranına sahip 5 faktörlü yapı elde edilmiştir. 5 faktörlü yapının elde edilen doğruluk oranı RF ile 0.60, SVM ile 0.71 olarak hesaplanmıştır. Gözetimli makine öğrenmesi ile gerçekleştirilen analiz sonucunda doğruluk oranı RF ile 0.70, SVM ile 0.73 olarak hesaplanmıştır.

Sonuç: Çok sınıflı yapılarda elde edilen doğruluk oranları iki sınıflı yapılara göre daha düşük bulunmuştur. Gözetimsiz makine öğrenme yöntemleri veri yapısı bilinmediği ve dağılımlarının orantılılığı olmadığı durumlarda dezavantajlıdır. Veri yapısında orantısız dağılımlar sağlandığı zaman sonuçların doğruluk oranı da artacaktır. Yüksek boyutlu veriler ile çalışmak hem zaman hem de maddi açıdan kayıplara neden olmaktadır. Bu nedenle boyut indirgeme yöntemleri avantaj sağlamaktadır. Bununla birlikte boyut indirgeme analizleri veri görselleştirmede de araştırmacılara sunum açısından kolaylık sağlamaktadır.

Anahtar Kelimeler: Boyut İndirgeme, Temel bileşenler Analizi, Random Forest, Destek Vektör Makinesi.

Holstein Irkı İneklerde Güç Doğum Üzerinde Etkili Olan Faktörlerin Belirlenmesinde FP Growth Algoritmasının Kullanımı

¹Elif ÇELİK GÜRBULAK, ²Murat ABAY, ³Aytaç AKÇAY, ⁴Funda İPEKTEN, ⁵Gökmen ZARARSIZ, ⁶Esmâ Gamze AKSEL

¹Erciyes Üniversitesi Veteriner Fakültesi Biyometri Anabilim Dalı Kayseri

²Erciyes Üniversitesi Veteriner Fakültesi, Doğum ve Jinekoloji Anabilim Dalı, Kayseri

³Ankara Üniversitesi Veteriner Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

⁴Erciyes Üniversitesi Sağlık Bilimleri Enstitüsü, Kayseri

⁵Erciyes Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Kayseri

⁶Erciyes Üniversitesi Veteriner Fakültesi, Genetik Anabilim Dalı, Kayseri

Amaç: Güç doğum, anne ve yavru kaynaklı olup, gerek anne ve yavrunun sağlığı bakımından gerekse ekonomik açıdan önemli bir problemdir. Bu nedenle, güç doğum özellikle süt verimi yüksek olan işletmelerde önemli bir sorun olarak ortaya çıkmaktadır. Literatürlere göre güç doğum üzerinde etkili faktörlerin başında hayvanın ırkı, cinsiyeti, beslenme, buzağılama mevsimi, yaş değişkenleri ve yönetsel faktörler gelmektedir. Güç doğumla ilişkili nedenler çoğunlukla bilinmekle beraber özellikle büyük işletmelerde klasik ilişki analizleri ile tespit edilemeyen örüntüleri bulmak önem arz etmektedir. Bu çalışmada, güç doğum üzerinde etkili olan bazı faktörlerin birliktelik kuralları algoritmalarından biri olan FrequentPatternGrowth (FP Growth) algoritması ile belirlenmesi amaçlanmıştır.

Gereç ve Yöntem: Birliktelik kuralları alışılmış istatistiksel yöntemler yerine kullanılabilecek analitik bir yöntemdir. Birliktelik kuralları ile tüketicilerin satın aldığı ürünler arasındaki ilişkilerin ortaya çıkarılarak satın alma eğilimleri belirlenmektedir. FP Growth ve Apriori algoritmaları birliktelik kuralları algoritmaları arasında en sık kullanılan algoritmalar. Algoritmanın performans açısından diğer algoritmalarından daha işlevsel olduğu daha önceki çalışmalarda gösterilmiştir. FP Growth Algoritmasını diğer algoritmalarından ayıran en temel karakteristikleri yaygın nesne kümelerini dolaysız olarak test edebilmesi ve büyük veri kümelerinde kısa sürede sonuç verebilmesidir. Bu çalışmada güç doğum, metritis, mastitis, ayak hastalığı, buzağılama mevsimi ve buzağılama yaşı değişkenlerinden oluşan 108 baş Holstein ineğe ait veriler kullanılmıştır. Minimum destek değeri ve minimum güven değeri 0.03 olarak seçilmiştir. Veriler R 4.2.1 (<https://cran.r-project.org>) yazılımında ve “rCBA” ve “arules” paketleri kullanılarak analiz edilmiştir.

Bulgular: Uygulanan FP Growth algoritması ile 176 kural 0.03 saniyede oluşturulmuştur. Elde edilen sonuçlara göre Holstein ineklerde 3-4 yaşlı olma ve normal doğumun birlikte görülme olasılığı %10 iken, 3-4 yaşlı Holstein ineklerinde normal doğum görülme olasılığı %65’tir. 9-10 yaşlı ve mastitisli olma durumu ile güç doğumun birlikte görülme olasılığı %11 iken; 9-10 yaşlı ve mastitisli ineklerde güç doğumun da görülme olasılığı %71 olarak belirlenmiştir.

Sonuç: FP Growth algoritması ile elde edilen sonuçların Apriori algoritması ile elde edilen sonuçlara benzerlik gösterdiği görülmüştür. Ayrıca, birliktelik kurallarının veteriner hekimlik alanında işlevsel bir yöntem olduğu sonucuna varılmıştır.

Anahtar Kelimeler: Holstein, güç doğum, FP Growth, apriori, birliktelik kuralları

Mamografi Görüntülerinde Farklı Ön İşleme Kombinasyonlarının Sınıflandırma Performansları Üzerine Etkisi

¹Hanife AVCI, ²Jale KARAKAYA

¹Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

²Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Amaç: Bir görüntü içerisinde çevresel faktörler, nem, elektromanyetik dalgalar ya da görüntü yakalama aygıtlarından kaynaklanan görüntüde istenmeyen parazitler oluşur. Görüntüler üzerinde ortaya çıkan istenmeyen işaretler (gürültü, noise), bölütleme ve özellik çıkarımı adımlarından çıkarılacak sonuçları etkiler. Bu nedenle görüntülerin bu gürültülerden temizlenmesi gerekmektedir. Alanyazında görüntülerdeki gürültülerin yok edilmesi ve görüntülerin iyileştirilmesi için farklı ön işleme yöntemleri önerilmiştir. Bu nedenle çalışmanın amacı, görüntü ön işleme yöntemleri ile, mamografi görüntüleri üzerinde netleştirme, belirli ayrıntıları ortaya çıkarma, görüntüyü yumuşatma ve kenar keskinleştirme gibi işlemler gerçekleştirmek ve ön işleme kombinasyonlarının sınıflandırma performansına katkısını incelemektedir.

Gereç ve Yöntem: Mamografi, meme kanseri taraması ve anormalliklerin tespiti için en etkili görüntüleme yöntemidir. Mamografi görüntülerinin okunmasının zor olması ve görüntü kalitesinin düşük olması nedeniyle bilgisayar destekli tanı (CAD) modelleri ile ilgili çalışmalar son yıllarda artmıştır. Bu çalışmada, mamografi görüntülerinde şüpheli alanların tespiti için bilgisayar destekli tanı sistemleri kullanılmıştır. Bu çalışmadaki CAD sistemi beş ana adımdan oluşmaktadır: ön işleme, bölütleme, ilgi bölgesi (ROI), özellik çıkarma ve sınıflandırma. Ön işleme algoritmaları, bölütleme ve sınıflandırma yöntemlerinin doğruluğunu önemli ölçüde etkiler. Bu çalışmada, mini-MIAS veri tabanından elde edilen 322 mamografi görüntüsü kullanılarak kontrast sınırlı uyarlamalı histogram eşitleme (CLAHE), Medyan filtresi ve Keskinliği Azaltma maskeleme gibi farklı ön işleme yöntemlerinin farklı kombinasyonlarının sınıflandırma performansı üzerindeki etkisi değerlendirilmiştir. ROI örnekleri, k-ortalamlar kümeleme bölütleme yöntemiyle elde edilmiştir. Gri seviye eş oluşum matrisi (GLCM) ve Gri seviye çalışma uzunluğu matrisi (GRLM) teknikleri kullanılarak ROI örneklerinden özellikler çıkarılmıştır. Korelasyon matrisi ile öznelik seçimi yapıldıktan sonra bu özneliklerin sınıflandırma başarısına etkileri 6 farklı klasik veri madenciliği yöntemi ile incelenmiştir. Görüntü işleme sonucu elde edilen sayısal veriler sınıflandırma yöntemlerinde giriş değişkeni olarak kullanılmıştır.

Bulgular: Çalışmada yapılan sınıflandırmada CLAHE algoritması ön işleme yöntemi olarak tek başına kullanıldığında diğer ön işleme kombinasyonlarına göre daha düşük sınıflandırma performans değerleri elde edilmiştir. CLAHE algoritmasına Medyan filtresi ve Unsharp maskeleme algoritmaları eklendiğinde, sınıflandırma yöntemlerinin performansı artmıştır. Sınıflandırma başarısı açısından en iyi algoritmaya baktığımızda Destek Vektör Makineleri (DVM), Rastgele Orman (RO), Yapay Sinir Ağları (YSA), Naive Bayes (NB) ve Karar Ağaçları (KA) yöntemleri k-En Yakın Komşuluk (k-NN) algoritmasına göre her iki sınıflandırmada da yüksek sınıflandırma sonuçları vermiştir.

Sonuç: Üçlü kombinasyonun ikili ve tekli kombinasyonlardan daha iyi performans göstermediği gözlemlenmiştir. Sınıflandırma yöntemleri açısından en yüksek performansı DVM, RO ve YSA göstermiştir.

Anahtar Kelimeler: Mamografi görüntüleri, ön işleme algoritmaları, sınıflandırma performansı.

SHINY ile Zaman Serileri Analizi İçin İnteraktif Arayüz Tasarımı

¹Esra GÜLTÜRK, ²Abdolvahap PINAR, ³Ahmet Kadir ARSLAN, ⁴Pınar YILMAZ

¹Cumhuriyet Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Sivas

²İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Malatya

³Cumhuriyet Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Sivas

⁴İnönü Üniversitesi Tıp Fakültesi Tıbbi Biyoloji ve Genetik Anabilim Dalı, Malatya

Amaç: Günümüzde ileriye dönük güvenilir karar alabilmek için geleceği doğru tahmin etmek son derece önemlidir. Zaman serileri analizi zamanla değişen gözlemlerin geçmişteki değerleri kullanılarak gelecekte alacağı değerleri tahmin edebilmek için kullanılan yöntemlerdir. Bu çalışmada R programlama dili kullanılarak, zaman serileri analizinde kullanılabilecek bir program geliştirmek ve ayrıca bu programa web üzerinden erişebilen bir arayüz oluşturmak amaçlanmıştır.

Gereç ve Yöntem: Bu çalışmada shiny paketi kullanılarak bir arayüz oluşturulmuştur. Geliştirilen arayüz xls/.xlsx, .arff, .sav ve .csv/.txt uzantılı veri dosyalarını desteklemektedir. Tek değişkenli zaman serileri yöntemleri ile ilgili "Karma otoregresif-hareketli ortalamalar (ARMA), Filtre (süzme) modelleri, Düzleştirme yöntemleri, Zaman serisi ayrıştırma yöntemleri ve yaygın olarak kullanılan zaman serisi teknikleri için geçmiş gözlemlerin ağırlıklandırılmasından oluşan kısa dönemli dalgalanmaları önleyen istatistiksel tekniklerdir. Bu tekniklerin temelinde bazı düzgün eğrilerden rasgele ayrılmaları gösteren zaman serisine ait geçmiş dönemlerde yaşanan dalgalanmalar gelecek için bir öngörü niteliğinde olacaktır. Buna göre en iyi öngörüye karar vermek için geliştirilen yazılım örnek veri setleri kullanılarak test edilmiştir.

Bulgular: En uygun modeli belirlemek için zaman yolu grafiği çizilmiştir. Durağanlığı bozan unsurlar var ise (trend ve mevsimsellik gibi) durağanlaştırma yapılmıştır. Veri setinin normal dağılıma uygunluğu test edilmiştir. En son uygun öngörüye göre model oluşturulmuştur.

Sonuç: Yazılım, sağlık araştırmalarında, ekonomide, tarım gibi birçok alanda kolaylıkla tek değişkenli zaman serisi analizini uygulayabilmek için geliştirilmiştir. İlerleyen çalışmalarda ilgili yazılımın çok değişkenli zaman serilerini analiz etme imkanı sağlayan sürümü de geliştirilecektir.

Anahtar Kelimeler: Arayüz, Shiny, Zaman serileri

Kardiyoverter Defibrilatör (Icd) İmplantasyonu Yapılan Hastalarda Yaşam Durumu Tahmin Modellemesi

¹Hatice AKTAŞ GÖKÇE, ²Amine BAYRAKLI, ³Emre YAŞAR

¹TC Sağlık Bakanlığı

²Karadeniz Teknik Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı

³Ankara Yıldırım Beyazıt Üniversitesi Tıp Fakültesi, Biyoistatistik Ana Bilim Dalı, Ankara, Türkiye

Amaç: Geçici kalp pili implante edilen hastaların klinik özellikleri, komorbiditeleri, altta yatan hastalıkları saptanıp, hastane içi mortalitelerinin tahminlemesi amaçlanmıştır.

Gereç ve Yöntem: Çalışmaya farklı sebeplerle kardiyoverter defibrilatör (ICD) implantasyonu işlemi yapılan 2019-2021 tarihleri arasında ülkemizde bulunan 2. Ve 3. Basamak sağlık tesislerinde takip edilen 25233 hasta dahil edilmiştir. Çalışmada Lojistik Regresyon yöntemi ile ölüm durumunu tahminleme amaçlanmıştır.

Bulgular: Çalışmaya geriye dönük olarak taranan 18768 erkek birey, 6464 kadın birey katılım sağlamıştır. Bireylerin tamamı kardiyoverter defibrilatör (ICD) implantasyonu işlemi gerçekleştirmiş hastalar olup, kadınların yaş ortalaması 63.22 ± 14.83 , erkeklerin yaş ortalaması 62.72 ± 12.93 olarak saptanmıştır. En sık kalp pili işlemi Implantable Cardioverter Defibrillator takılması, tek elektrot olarak belirlenmiştir. Ölüm durumunu bağımlı değişken olarak belirlediğimizde modele eklediğimiz yaş, cinsiyet, işlem tipi, ölüm zamanı (ölüm zamanı-işlem zamanı) gibi bağımsız değişkenler ile bağımlı değişkenin %78.7'sinin lojistik regresyon yöntemi ile açıklandığı saptandı. Çalışma Python dilinde NumPy, Matplotlib kütüphaneleri ve Scikit-learn paketleri de kullanılarak ilave bağımsız değişkenler ile lojistik regresyon yöntemi ve Karar Ağaçları yöntemi uygulanarak tekrar değerlendirilip karşılaştırılacaktır.

Sonuç: Kardiyoverter defibrilatör (ICD) implantasyonu işlemi gerçekleştirmiş bireylerin yaşa, cinsiyete, işlem tipine bağlı olarak yaşama süreleri farklılık göstermektedir. Farklılığın sonucu ise ölümcül seyirdir.

Anahtar Kelimeler: Tahmin modelleme, yaşam durumu, kalp pili

Sağkalım Analizi için Veri Simülasyonu

¹İrem KAR, ²Erdal COŞGUN, ³Atilla Halil ELHAN

¹Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

²GenomicsTeam, Microsoft Research&AI, Redmond, WA, USA

³Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Amaç: Sağkalım analizi çalışmalarında gerçek veriye ulaşmak oldukça zordur. Bu nedenle, veri analistleri simülasyon çalışmalarına yönelmişlerdir. Bu çalışmada, sağkalım analizinde derin öğrenme için veri simüle ederken dikkat edilmesi gereken parametreleri vurgulamak ve bu parametrelerin DeepSurv yönteminin tahminleme başarısına etkisini ortaya koymak amaçlanmıştır.

Gereç ve Yöntem: Burada, sağkalım analizleri için veri simülasyonunda en çok tercih edilen R programlama dilindeki “coxed” paketinde üzerinde durulmuş olup, simüle edilmiş veri setleri kullanılarak bir benzetim çalışması gerçekleştirilmiştir. Bu kapsamda örneklem büyüklüğünün, sansür oranının, bağımsız değişkenlere ilişkin katsayıların ve çeşitli istatistiksel dağılımlardan bağımsız değişkenlerin simüle edildiği pek çok senaryo oluşturulmuştur. Oluşturulan her senaryo ise 100 kez tekrarlanmıştır. Ayrıca, simüle verilere ek olarak gerçek veri üzerinde de DeepSurv yöntemi uygulanmıştır. Yöntemlerin karşılaştırılmasında Harrell’in C-indeks değeri kullanılmıştır. Verilerin türetilmesi R programlama dili (ver. 3.6.3) ile derin öğrenmeye ilişkin analizler ise Python programlama dilinde gerçekleştirilmiştir.

Bulgular: Yapılan benzetim çalışması sonucunda, örneklem büyüklüğünün, sansür oranının, bağımsız değişkenlere ilişkin katsayıların ve bağımsız değişkenlerin simülasyonunda kullanılan istatistiksel dağılımların tahminleme başarısını etkilediği bulunmuştur. Örneklem büyüklüğü arttıkça yöntemlerin performansı artmıştır. Sansür oranının artması tahminleme başarısını azaltmıştır. Bağımsız değişkenlere ilişkin katsayıların artması yöntemin sınıflama performansını arttırmıştır. Bağımsız değişken katsayılarının “coxed” paketindeki mevcut haliyle kullanıldığı durumda ise 0,50-0,60 aralığında bir C-indeks değeri elde edilmiştir.

Sonuç: Bu çalışmada, sağkalım verisi simülasyonunda kullanılan parametrelerin önemi ve tahminleme başarısına etkisi açık bir şekilde ortaya konmuştur. Sonuç olarak, sağkalım analizinde derin öğrenme için kullanılan DeepSurv algoritmasında çok sayıda hiperparametre kullanılması model başarısını arttırmak için avantaj olmasına rağmen, veri simüle ederken göz ardı edilen ya da yanlış kullanılan parametrelerin tahminleme başarısı üzerindeki etkisinin önemi anlaşılmıştır.

Anahtar Kelimeler: DeepSurv, harrell’in C-indeksi, sağkalım analizi

R Shiny İle Zaman Serisi Analizini İçin İnteraktif Arayüz Tasarımı

¹Esra GÜLTÜRK, ²Abdolvahap PINAR, ³Ahmet Kadir ARSLAN, ⁴Pınar YILMAZ

¹Cumhuriyet Üniversitesi Tıp Fakültesi, Biyoistatistik Ana Bilim Dalı, Sivas

²İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Malatya Türkiye

³İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Malatya Türkiye

⁴İnönü Üniversitesi Tıp Fakültesi Tıbbi Biyoloji ve Genetik Anabilim Dalı, Malatya

Amaç: Günümüzde ileriye dönük güvenilir karar alabilmek için geleceği doğru tahmin etmek son derece önemlidir. Zaman serileri analizi zamanla değişen gözlemlerin geçmişteki değerleri kullanılarak gelecekte alacağı değerleri tahmin edebilmek için kullanılan yöntemlerdir. Bu çalışma R programlama dili kullanılarak, zaman serileri analizinde kullanılabilecek bir program geliştirmek ve ayrıca bu programa web üzerinden erişebilen bir arayüz oluşturmak amaçlanmıştır.

Gereç ve Yöntem: Bu çalışmada R-Studio ile shiny paketi kullanılarak bir arayüz oluşturulmuştur. Geliştirilen arayüz xls/.xlsx, .arff, .sav ve .csv/.txt uzantılı veri dosyalarını desteklemektedir. En çok kullanılan öngörülerden tek değişkenli zaman serileri yöntemleri ile ilgili "Karma otoregresif-hareketli ortalamalar (ARMA), Filtre (süzme) modelleri, Düzleştirme yöntemleri, Zaman serisi ayrıştırma yöntemleri ve yaygın olarak kullanılan zaman serisi teknikleri için geçmiş gözlemlerin ağırlıklandırılmasından oluşan kısa dönemli dalgalanmaları önleyen istatistiksel metotlardır. Bu metotların temelinde bazı düzgün eğrilerden tesadüfi ayrılmaları gösteren zaman serisine ait geçmiş dönemlerde yaşanan dalgalanmalar gelecek için bir öngörü niteliğinde olacaktır. Buna göre en iyi öngörüye karar vermek için hazırlanan program ve arayüz örnek veri setleri kullanılarak test edilmiştir.

Bulgular: En uygun modeli belirlemek için zaman yolu grafiği çizilmiştir. Durağanlığı bozan unsurlar var ise (trend ve mevsimsellik gibi) durağanlaştırma yapılmıştır. Veri setinin normal olup olmadığı test edilmiştir. En son uygun öngörüye göre model oluşturulmuştur.

Sonuç: Yazılım, sağlık araştırmalarında, ekonomide, tarım gibi birçok alanda kolaylıkla tek değişkenli zaman serisi analizini uygulayabilmek için geliştirilmiştir. İlerleyen çalışmalarda çok değişkenli zaman serileri ile ilgili yazılım geliştirilecektir.

Anahtar Kelimeler: Arayüz, Shiny, Zaman serileri

Periton Diyaliz Hastalarında Sağkalım ve Uzunlmasına Verilerin Birlikte Modellemesi: Kişiyi Özgü Dinamik Risk Kestirimleri

¹Merve Başol GÖKSÜLÜK, ²Đinçer GÖKSÜLÜK, ³Ergun KARAAĞAOĞLU

¹Erciyes Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi ABD, Kayseri, Türkiye

²Erciyes Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi ABD, Kayseri, Türkiye

³Erciyes Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi ABD, Kayseri, Türkiye

Amaç: Periton diyaliz (PD) tedavisi, geç dönem böbrek yetmezliği olan hastalarda sıklıkla uygulanmaktadır ve nefrolji alanında PD hastalarının sağkalım durumları güncel olarak çalışılmaya devam etmektedir. Hastalığın seyrini değerlendirebilmek için böbrek fonksiyonu ile ilişkili olan biyobelirteçler ve çeşitli biyokimyasal parametreler izlem süresinde tekrarlı olarak ölçülmektedir. Literatürdeki çalışmalar incelendiğinde sağkalım analizlerinde tekrarlı alınan bu ölçümlerin başlangıç ve/veya izlem süresi içindeki ortalama değerlerinin çoğunlukla dikkate alındığı görülmektedir. Bu yaklaşım yanlış olmamakla birlikte tekrarlı ölçümlerin zaman içerisindeki değişimini gözardı etmekte ve yanlış sonuçlar elde edilmesine yol açabilmektedir. Bu çalışma, PD hastalarında çeşitli biyobelirteçlerin mortalite ile olan ilişkisini tekrarlı ölçüm yapısını dikkate alarak ortaya koymayı amaçlamaktadır.

Gereç ve Yöntem: Bu çalışmada 1995 – 2007 yılları arasında Erciyes Üniversitesi Nefroloji bölümünde periton diyaliz (PD) tedavisine başlayan 511 hastaya ait veriler geriye dönük olarak değerlendirilmiştir [1]. Hastalar, PD başlangıcından itibaren ölüm, böbrek nakli, hemodiyalize geçiş, izlemiden çıkma, PD tedavisinin sonlandırılması gibi sebeplerden herhangi birisi gözleninceye kadar takip edilmiştir. Mortalite yanıt değişkeni PD sürecinde gözlenen tüm sebeplerden kaynaklı ölümdür. Biyobelirteçlerin zaman içerisindeki değişiminin sağkalım üzerindeki etkisi birleşik modelleme (joint modelin) yaklaşımı kullanılarak değerlendirilmiştir [2, 3]. Sağkalım modeli için Cox regresyon yöntemi ve uzunlmasına veriler için karma etkili doğrusal modelden yararlanılmıştır. Çalışmada tekrarlı ölçüm olarak serum albumin değerleri dikkate alınmıştır.

Bulgular: Elde edilen sonuçlara göre ortalama serum albumin değerlerinin mortalite ile ilişki olmadığı görülmüştür. Ancak, tekrarlı olarak ölçülen serum albumin değerlerinin mortalite ile anlamlı ve ters yönlü bir ilişkisi olduğu gözlenmiştir (HR: 2.43, %95 G.A.: 1.48 – 4.16). Peritonit hızı, hemodiyaliz geçmişi ve toplam komorbid hastalık sayısı ile mortalite arasında pozitif yönlü bir ilişkinin olduğu ve ölüm oranlarının sırasıyla 1.74, 3.21 ve 1.41 kat artış gösterdiği belirlenmiştir. Sağkalım durumları hastalar arasında anlamlı bir değişkenlik göstermektedir. Kişiselleştirilmiş dinamik risk kestirimleri dikkate alındığında Cox regresyon modelinden elde edilen genel sağkalım kestirimlerinin tüm hastaları temsil edemediği, hastaların karama etkili modelden elde edilen bireysel sağkalım kestirimlerinin genel tahmin değerlerinden önemli ölçüde farklılaştığı görülmüştür.

Sonuç: Serum albumin değerlerinin sağkalım ile olan ilişkisi tekrarlı ölçümler üzerinden uygun istatistiksel modelleme teknikleri ile değerlendirildiğinde daha doğru ve yansız kestirimler elde edilmektedir. Kişi bazlı elde edilen dinamik risk kestirimleri hastalığın seyrini tahmin etmede ve hastaların takibinde önemli bir rol oynamaktadır. Ayrıca, izlem süresince yeni veriler elde edildikçe risk kestirimleri dinamik olarak güncellenebilmekte ve kişiyi özel kararlar alınabilmektedir.

Anahtar Kelimeler: Mortalite, kişiselleştirilmiş tıp, sağkalım, birleşik model, periton diyaliz

Boylamsal ve Yaşam Verilerinin Parametrik Bileşik Modellemesinde Parametre Tahmin Yaklaşımlarının Karşılaştırılması

¹Elif YILDIRIM, ²Duru KARASOY

¹Konya Teknik Üniversitesi, Rektörlük, Kalite Koordinatörlüğü, Konya

²Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Ankara

Amaç: Boylamsal ve yaşam verilerinin farklı yöntemler ile modellenmesi üzerine yapılan çalışmaların sayısı son yirmi yılda artmıştır. Farklı modellemelerin yanı sıra parametre tahminlerini daha yansız olarak elde edebilmek için farklı parametre tahmin yöntemlerinin geliştirildiği görülmektedir. Bileşik modelleme ve parametre tahmin yaklaşımları üzerine yapılan çalışmaların çoğu boylamsal veriler için doğrusal karma etkili model ve yaşam verilerinin Cox regresyon modelinin paylaşılmış rastgele etkiler ile birleştirilmesinden elde edilmektedir. Yaşam verilerinin orantılı tehlikeler varsayımını sağlamadığı durumlarda ise parametrik bileşik modellemeler kullanılabilir. Bu çalışmanın amacı, yaşam verilerinin orantılı tehlikeler varsayımını sağlamadığı durumlarda da kullanılabilen parametrik bileşik modellemede daha yansız tahminler elde etmek için parametre tahmin yöntemlerini farklı simülasyon senaryoları ile incelemek ve en uygun parametre tahmin yöntemini belirlemektir.

Gereç ve Yöntem: Parametrik bileşik modellemede parametre tahmin yaklaşımlarını incelemek amacıyla, yaşam verileri için Weibull hızlandırılmış başarısızlık süresi (AFT) modeli ile boylamsal veriler için doğrusal karma etkili modelin parametrik bileşik modellemesi ele alınmıştır. Elde edilen modelin parametrelerini elde etmek için kullanılan standart Gauss Hermite ve Uyarlanmış (adaptive) Gauss Hermite yaklaşımları farklı örneklem büyüklüklerinde simülasyon çalışmasıyla incelenmiş ve iki yöntem karşılaştırılmıştır.

Bulgular: Çalışmada, Weibull AFT modeli ve doğrusal karma etkili modelin rastgele etkiler ile birleştirilmesinden elde edilen bileşik modelin parametre tahminlerini elde etmek için kullanılan ve integrasyon yaklaşımı olan standart Gauss Hermite ve uyarlanmış Gauss Hermite yaklaşımlarının performanslarını karşılaştırmak için Monte Carlo simülasyonu çalıştırılmıştır. Simülasyon kodları R programında yazılmış olup, model tahminlerini elde etmek için "JM" paketi kullanılmıştır. Çalışmada durdurma oranı %50 olarak belirlenmiş ve $n=50$, 100 ve 500 için iki parametre tahmin yaklaşımının performans karşılaştırması yapılmıştır. Boylamsal alt modeldeki intercept, zaman, grup*zaman ve grup değişkenlerinin gerçek değerleri sırasıyla 2, 1.5, 0.5 ve 0.8 olarak; yaşam alt modeldeki grup değişkeninin gerçek değeri -0.3 ve birleştirme katsayısı 1.6 olarak belirlenmiştir. Aynı zamanda boylamsal gözlemin zaman boyutu 0, 2, 4, 6 ve 8 olmak üzere 5 boyuttan oluşacak şekilde üretilmiştir. Yöntemlerin performanslarının karşılaştırmak için parametre tahminlerinin yan ve hata kareler ortalamalarının karekök (RMSE) değerleri kullanılmıştır.

Sonuç: Parametrik bileşik modellemede parametre yaklaşım yöntemlerini incelemek için yapılan simülasyon sonuçlarına göre; örneklem genişliği küçük olduğunda uyarlanmış Gauss Hermite yaklaşımının RMSE ve yan değerlerinin standart Gauss Hermite yaklaşımına göre daha küçük olduğu gözlemlenirken, örneklem genişliği arttıkça Gauss Hermite yaklaşımının, uyarlanmış Gauss Hermite yaklaşımına göre daha az yanlı tahminler verdiği görülmüştür.

Anahtar Kelimeler: Parametrik Bileşik Model, Parametre Tahmin Yaklaşımı, Gauss Hermite

Önerilen Derin Öğrenme Mimarisi ile MR Görüntülerine Dayalı Alzheimer Hastalığının Sınıflandırılması

¹Mehmet KIVRAK, ²Cemil ÇOLAK

¹Recep Tayyip Erdoğan Üniversitesi Biyoistatistik ve Tıp Bilişimi

²İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Malatya

Amaç: Bu çalışma, keras kütüphanesi ve evrimsel sinir ağları (CNN'ler) kullanarak derin öğrenme yaklaşımı ile MRI görüntülerinden oluşan çok hafif/hafif/orta/demansız hastaları yüksek doğrulukla sınıflandırmayı amaçlamaktadır.

Gereç ve Yöntem: İncelenen görüntü veri seti, 179 hafif demanslı, 12 orta demans, 640 demanssız ve 448 çok hafif demanslı olmak üzere toplam 1279 MRI hasta taramasından oluşmaktadır. Araştırmada geriye dönük vaka-kontrol tasarımı kullanılmıştır. Keras kitaplığı (TensorFlow), kullanıcının evrimsel sinir ağı modellerini daha basit bir şekilde oluşturmaları sağlar ve kullanıcıyı bu düşük seviyeli kitaplıkların karmaşıklığından kurtarır. GPU donanımı olarak CUDA destekli yapay zeka ve analitik taleplerini karşılamak üzere geliştirilen NVIDIA DGX™ Sistemleri kullanıldı. Modellerin sınıflandırma başarımları çeşitli metrikler ile değerlendirilmiştir.

Bulgular: CNNs modelini geliştirmek için kullanılan veri seti, eğitim (1024) ve test (255) veri setlerine ayrılmıştır. Modelin eğitim aşamasındaki en yüksek doğruluk oranı 0.96 olarak tahmin edilmiştir. Test verilerinin doğruluk, duyarlılık, özgüllük, pozitif ve negatif tahmin değerlerine göre en yüksek sınıflandırma performans oranı 0.93 ile pozitif tahmin değeri olmuştur.

Sonuç: Deneysel sonuçlar, MRI görüntülerine dayalı Alzheimer sınıflandırılmasında derin öğrenme yöntemlerinin (CNN'ler) oldukça başarılı olduğunu ve benzer çalışmaların diğer hastalıklar için de yapılabileceğini göstermiştir.

Anahtar Kelimeler: Alzheimer, Görüntü İşleme, Derin Öğrenme, Evrimsel Sinir Ağları.

Yüksek Boyutlu Verilerde Eksik Veri Değer Atama Yöntemlerinin Tahmin Performanslarının Aşırı Öğrenme Makineleriyle İncelenmesi

¹Buğra VAROL, ²İmran ÖMÜRLÜ, ³Mevlüt TÜRE

¹Adnan Menderes Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı, Aydın

²Adnan Menderes Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Aydın

³Adnan Menderes Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Aydın

Amaç: Eksik değerli veri setleri yüksek boyutlu verilerin analizi ve sınıflandırılması açısından temel bir sorundur. Bu çalışmanın amacı, yüksek boyutlu verilerde farklı eksik veri değer atama yöntemlerinin tahmin performanslarının türetilmiş veriler üzerinde aşırı öğrenme makinalarıyla (ELM) karşılaştırılmasıdır.

Gereç ve Yöntem: Çalışmada, öncelikle $n=150$ gözlemden oluşan iki kategorili bağımlı değişken ve karma ilişki yapılı $p=500$ değişkenden oluşan rastgele veriler türetilmiştir. Daha sonra bağımsız değişkenlerden %10, %20, %30, %40, %50 oranında eksik değerler oluşturulmuştur. Eksik veri değer atama yöntemlerinden; ortalama, medyan, rastgele, k en yakın komşu (KNN), sınıflandırma ve regresyon ağaçları (CART), rasgele orman (RF) değer atama yöntemlerinin yanı sıra yüksek boyutlu veriler için geliştirilen düzenlenleştirilmiş regresyonun doğrudan kullanımı (DURR) ve düzenlenleştirilmiş regresyonun dolaylı kullanımı (IURR) yöntemleri ile eksik değerler tahmin edilmiştir. 1000 döngü ile yapılan simülasyon sonunda yöntemlerin ELM ile sınıflandırma skorları elde edilmiş ve bu skorların referansa olan yakınlığına göre eksik değer tahmin performansları değerlendirilmiştir.

Bulgular: Simülasyon bulguları incelendiğinde düşük eksik oranlarında tüm yöntemler arasından CART, KNN, RF, DURR ve IURR yöntemlerinin birbirine ve referansa yakın iyi performans gösteren yöntemler olduğu belirlenmiş; bunun yanı sıra CART, KNN ve RF yöntemlerinin eksik oranı arttıkça performanslarının genel olarak düşüş eğiliminde olduğu; DURR ve IURR yöntemlerinin ise yüksek eksik oranlarında bile istikrarlı bir şekilde iyi performans göstermeye devam ettiği gözlenmiştir.

Sonuç: Eksik değer oranı arttıkça ortalama, medyan, rastgele, CART, RF ve KNN yöntemlerinin tahmin performansında düşüşler gözlenirken, DURR ve IURR yöntemlerinin eksik değer oranından etkilenmediği belirlenmiştir. Eksik değer oranındaki artışa karşı daha kararlı sonuçlar vermeleri nedeniyle DURR ve IURR yöntemleri, yüksek boyutlu verilerde daha avantajlıdır.

Anahtar Kelimeler: Aşırı öğrenme makineleri, Eksik veri, Değer atama, Sınıflandırma, Simülasyon.

Biocrates Mxp® Quant 500 Kiti Kullanılarak Elde Edilen Metabolomik Verilerinin Laboratuvarlar Arası Karşılaştırması

¹Gözde ERTÜRK ZARARSIZ, ²Gözde ERTÜRK ZARARSIZ, ³Alexander CECIL, ⁴Jutta LINTELMANN, ⁵Gernot POSCHET, ⁶Jennifer KIRWAN⁷Sven SCHUCHARDT, ⁸Xue L. GUAN, ⁹Daisuke SAIGUSA, ¹⁰David WISHART, ¹¹Jiamin ZHENG, ¹²Rupasri MANDAL, ¹³Lisa ST. JOHN-WILLIAMS, ¹⁴Kendra ADAMS, ¹⁵J. Will THOMPSON, ¹⁶Michael P. SNYDER, ¹⁷Kevin CONTREPOIS, ¹⁸S Songjie CHEN, ¹⁹Nadia ASHRAFI, ²⁰Sumeyya AKYOL, ²¹Ali YILMAZ, ²²Stewart GRAHAM, ²³Thomas M. O'CONNELL, ²⁴Karel KALECKÝ, ²⁵Karel KALECKÝ, ²⁶T. Hai PHAM, ²⁷Therese KOAL, ²⁸Teodoro BOTTIGLIERI, ²⁹Teodoro BOTTIGLIERI, ³⁰Jerzy ADAMSKI, ³¹Gabi KASTENMÜLLER

¹Erciyes Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi ABD, Kayseri, Türkiye

²Institute of Computational Biology, Helmholtz Zentrum München, Almanya,

³Metabolomics and Proteomics Core, Helmholtz Zentrum München, Almanya

⁴Metabolomics and Proteomics Core, Helmholtz Zentrum München, Almanya

⁵Metabolomics Core Technology Platform, Heidelberg University, Almanya

⁶BIH Metabolomics Platform, Berlin Institute of Health@Charité, Almanya

⁷Fraunhofer-Institute for Toxicology and Experimental Medicine, Hannover, Almanya

⁸Lee Kong Chian School of Medicine, Nanyang Technological University, Singapur

⁹Laboratory of Biomedical and Analytical Sciences, Faculty of Pharma-Science, Teikyo University, Japonya

¹⁰Departments of Biological Sciences and Computing Science

¹¹University of Alberta, Alberta, Kanada

¹²University of Alberta, Alberta, Kanada

¹³Duke University, Durham, NC, Amerika

¹⁴Duke University, Durham, NC, Amerika

¹⁵Duke University, Durham, NC, Amerika

¹⁶Stanford University, Stanford, CA, Amerika

¹⁷Stanford University, Stanford, CA, Amerika

¹⁸Stanford University, Stanford, CA, Amerika

¹⁹Beaumont Health System, Royal Oak, MI, Amerika

²⁰Beaumont Health System, Royal Oak, MI, Amerika

²¹Beaumont Health System, Royal Oak, MI, Amerika

²²Beaumont Health System, Royal Oak, MI, Amerika

²³Indiana University, Indianapolis, IN, Amerika

²⁴Center of Metabolomics, Institute of Metabolic Disease, Baylor Scott & White Research Institute, Dallas, TX, Amerika

²⁵Institute of Biomedical Studies, Baylor University, Waco, TX, Amerika

²⁶Biocrates life sciences ag, Innsbruck, Avustralya

²⁷Biocrates life sciences ag, Innsbruck, Avustralya

²⁸Center of Metabolomics, Institute of Metabolic Disease, Baylor Scott & White Research Institute, Dallas, TX, Amerika

²⁹Institute of Biomedical Studies, Baylor University, Waco, TX, Amerika

³⁰Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Almanya

³¹Institute of Computational Biology, Helmholtz Zentrum München, Almanya,

Amaç: Amaç: Metabolomik, yüzlerce düşük ağırlıklı organik molekülü (metabolit) tek bir biyo numunede paralel olarak ölçerek, belirli hücre içi biyokimyasal süreçlerin geride bıraktığı kimyasal parmak izlerini inceler. Bu projede, 14 uluslararası laboratuvar 634'e yakın metabolitin kantifikasyonunda kullanılan MxP® Quant 500 (Biocrates Life Science AG, Innsbruck, Avusturya) metabolomik kitinin analitik performansının değerlendirilmesituvarlar arasındaki analitik değişimin değerlendirilmesi hedeflendi.

Gereç ve Yöntem: Yöntem: Laboratuvarlarda aynı insan serumu ve plazma setleri (kadın ve erkek), NIST 1950 insan plazması, lipemik insan plazması ve fare ve rat plazma örnekleri ile kalite kontrol örnekleri (QC2) analiz edildi. MxP® Quant 500 kiti bu laboratuvarlarda bulunan dört farklı kütle spektrometrisi ve dört farklı sıvı kromatografisi sisteminde kullanıldı. Çalışmaya katılan her laboratuvara A'dan N'ye kadar tek bir harf etiketi verildi. Deneylerin gerçekleştirildiği laboratuvarlardan veriler toplandı. Laboratuvarlar arası tekrar üretilebilirlik varyasyon katsayıları ile değerlendirildi. Her bir laboratuvar arasındaki farklılık yöntem karşılaştırma ile değerlendirildi. Analitik ölçüm hata oranlarının laboratuvarlar arasında sabit olduğu varsayıldı. Laboratuvar çiftleri arasındaki karşılaştırmalarda Deming regres-

23. Ulusal ve 6. Uluslararası Biyoistatistik Kongresi

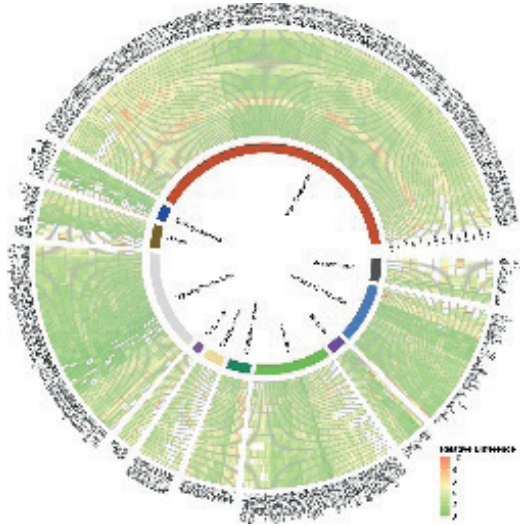
26-29 Ekim 2022, Ankara Üniversitesi Tıp Fakültesi

yon yöntemi kullanıldı. Ölçüm hataları her metabolitin üç tekrarlı ölçümlerinden tahmin edildi. Güven aralıklarının tespitinde jackknife yöntemi uygulandı. Doğrusal regresyon analizleri ile birlikte mutlak ve göreceli farklılıklar hesaplandı. Sonuçlar her bir metabolitin en küçük, ortanca ve en büyük değerleri için ayrı ayrı değerlendirildi.

Bulgular: Bulgular: Çalışma kayıp verilerin temizlenmesi ile 634'ten 561 metabolite analize hazır hale getirildi. Metabolite düzeylerine göre laboratuvarlar arasındaki sabit hata miktarı 574 iken oransal hata miktarı 595'tir. Sistemik hata miktarı 600 olarak bulunmuştur. En yüksek hata miktarı PC.aa.C36.0 metabolitinde bulunmuştur. Yapılan detaylı karşılaştırmalar sonucunda metabolit konsantrasyonlarına göre laboratuvarlar arasında sistematik farklılıklar bulundu. Metabolitlerin en küçük değerleri için bağıl farklılığın daha fazla olduğu tespit edildi. Laboratuvar J için triağılgiserit metabolit sınıfında yüksek bağıl farklılıklar elde edildi. Laboratuvar sonuçları Varyasyon katsayısı %5'in altında 6, %10'un altında 122, %20'nin altında ise 392 metabolit tespit edildi.

Sonuç: Sonuç: Laboratuvarlar arasında belirli metabolit ve metabolit grupları için sistematik farklılıkların olduğu görülmüştür. Bu doğrultuda kullanılan MxP® Quant 500 kitinin hangi laboratuvarlarda farklı sonuçlar verdiği ve bu farklılığın ne kadar olduğu belirlenmiştir. Çalışmada kullanılan bu yaklaşım ile birbirlerinden uzak ama aynı işlemi yapan laboratuvarlarda MxP® Quant 500 kiti kullanılarak metabolomik analizleri gerçekleştirilebilir. Gözlem sayısının fazla olduğu ve örneklerin tekrarlı çalışıldığı farklı laboratuvarlarda veriler birlikte değerlendirilerek analiz edilebilir.

Anahtar Kelimeler: Anahtar kelimeler: Bağıl farklılık, Deming regresyon, metabolomik, sistematik farklılık, yöntem karşılaştırma



Şekil-1. Dairesel ısı grafiği ile laboratuvarlar arasında metabolitlerin göreceli sistematik farklılıklarının gösterilmesi

SB – 16

Sağkalım Analizinde Kopula Entropi Tabanlı Değişken Seçimi

Aslıhan ALHAN

Ufuk Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Amaç: Veri analizlerinde modele seçilen değişkenlerle, oluşturulan son modelin yorumlanabilir olması kritik öneme sahiptir. Bir modelin yorumlanabilir olması, modelin sonucunu tutarlı bir şekilde tahmin edebilme ve güvenilirlik derecesidir. Yüksek boyutlu değişkenlerin modellenmesinde, modelin yorumlanabilir olması amacıyla daha az değişken kullanılarak modelleme yapılmaktadır. Bu durum ise modele alınabilecek aday değişkenlerin seçiminde önemli bir problem oluşturmaktadır. Son zamanlarda değişken seçiminde uygulanan kopula entropisi, değişkenler ile olaya kadar geçen süre arasındaki korelasyonu elde eden ve istatistiksel bağımsızlığı göstermek amacıyla kullanılan bir matematiksel kavramdır. Sağkalım analizinde kullanılan Cox regresyon, Random Survival Forest (RSF) ve Lasso-Cox değişken seçim yöntemlerine alternatif olacak bir yaklaşım olanağı sunan kopula entropi tabanlı değişken seçim yöntemini tanıtmak bu çalışmanın amacıdır.

Gereç ve Yöntem: Kopula entropi ile bağımlı ve bağımsız değişkenler arasındaki korelasyon ölçülerek değişken seçimi yapılmıştır. Sağkalım analizinde değişken seçimini uygulamak amacıyla Jian Ma'nın (2022) ABD Kuzey Merkez Kanseri Tedavi Grubu'nda ileri evre akciğer kanserli hastalardan elde edilen 10 değişkenli 167 hastanın verilerinden yararlanılmıştır. İstatistiksel analizler için "copent", "survival", "randomForestSRC" ve "fastcox" R paket programı kullanılmıştır.

Bulgular: Model elde etmek amacıyla kullanılan 10 değişken (hastane türü, hayatta kalma süresi, sansürleme durumu, yaş, cinsiyet, ECOG performans puanı ve doktorlar tarafından değerlendirilen Karnofsky performans puanı, hastaların kendi Karnofsky performans puanı, öğünlerde tüketilen kaloriler ve son altı ayda kilo kaybı) için, seçilen değişkenler Cox regresyon, Random Survival Forest (RSF), Lasso-Cox ve Kopula Entropi modellerinde farklı bulunmuştur. Tüm modeller tarafından seçilen cinsiyet, akciğer kanserinin iyi bilinen bir prognostik faktörüdür. Performans skorları, akciğer kanseri için ortak faktörlerdir ve diğer faktörlerden daha önemli olarak kabul edilmektedir. Kopula entropi modeli, klinik kullanım için etkinliğini gösteren üç performans faktörünün tümünü seçerken RSF ve Lasso-Cox ise iki performans faktörünü seçmiştir (hastalara göre ECOG performans puanı ve Karnofsky performans puanı). Risk faktörü için bir gösterge olarak kabul edilen kilo kaybı ise sadece RSF modeli tarafından seçilmiştir.

Sonuç: Kopula entropi yönteminde, değişkenler ve olaya kadar geçen süre arasındaki korelasyon ölçülür ve ardından daha küçük kopula entropi değerleriyle hayatta kalma süresi ile ilişkilendirilen değişkenler seçilir. Kopula entropi, istatistiksel bağımsızlık için modelsiz bir ölçüdür ve parametrik olmayan bir tahmin yöntemine sahiptir. Bu nedenle, herhangi bir varsayım olmaksızın sağkalım analizine uygulanabilir. Kopula entropiyi tahmin etmek için kullanılan yöntemin hiper-parametre seçimlerine karşı duyarsız olması nedeniyle ayar gerektirmemesi diğer bir avantajdır. Yapılan analizler sonucunda, kopula entropi ile tahmin performansından ödün vermeden yorumlanabilir değişkenlerin seçilebileceği görülmektedir.

Anahtar Kelimeler: Sağkalım analizi, kopula entropisi, değişken seçimi, yorumlanabilirlik.

Propensity Skora Dayalı Eşleştirme Analizinde Optimum Caliper Belirlenmesi İçin Bir Benzetim Çalışması

¹Gülçin AYDOĞDU, ²Muhammed Ali YILMAZ, ³Emre DEMİR, ⁴Yasemin YAVUZ

¹Hitit Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Çorum

²Sağlık Bakanlığı, Çağrı Hizmetleri Dairesi Başkanlığı, Ankara

³Hitit Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Çorum

⁴Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Amaç: Randomizasyonun yapılamadığı gözlemsel araştırmalarda vaka ve kontrol grupları arasında karıştırıcı değişken/faktörlerin benzer dağılmasını sağlamak amacıyla vaka-kontrol eşleştirmesi için literatürde yaygın olarak lojistik regresyon ile hesaplanan Propensity Skor (PS) tahminine dayalı eşleştirme analizi kullanılmaktadır. Eşleştirme analizinde caliper seçimi eşleştirmenin başarısını doğrudan etkileyen önemli bir parametredir. Bu çalışmanın amacı, Propensity skora dayalı eşleştirme analizinde farklı senaryolar oluşturarak optimum caliper değerinin belirlenmesi amacıyla bir benzetim çalışması gerçekleştirmektir.

Gereç ve Yöntem: Eşleştirme analizini ve PS tahminini gerçekleştirmek için “matchIt” kütüphanesi kullanıldı. Ortak değişkenlere ilişkin eşleştirme dengesinin değerlendirilmesi için Standardize Ortalama Farklar (SOF), Varyans Oranları (VO), Kolmogorov Smirnov (KS) test istatistikleri ve çeşitli grafiksel yaklaşımlar kullanıldı. Farklı senaryolar ile üretilen veri setleri ile eşleştirme analizi yapıldı. Propensity skorlar lojistik regresyon kullanılarak hesaplandı. Eşleştirme analizlerinde en yakın komşuluk algoritması yöntemi 1:1 eşleştirme oranında gerçekleştirildi. Benzetim senaryolarında sırasıyla 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.40, 0.45, 0.5 ve propensity skorların standart sapmasından hesaplanan caliper değerleri kullanıldı.

Bulgular: Tüm benzetim senaryolarında SOF, KS ve VO denge değerlendirmesi bulguları Şekil 1 ve Şekil 2’de gösterildi.

Sonuç: Sadece kategorik değişkenlerden oluşan ve hem kategorik hem sayısal değişkenlerden oluşan her iki modelde de eşleştirme sonrasında denge değerlendirmelerine göre en başarılı caliper değeri 0.1 olarak belirlendi.

Anahtar Kelimeler: Propensity skor, eşleştirme, caliper, gözlemsel çalışmalar

Parametreler		Standardleştirilmiş Ortalamalar Arası Farklar							Kolmogorov-Smirnov Testi				
Senaryo	Örneklem Büyüklüğü	X1	X2	X3	X4	X5	X6	X7	X1	X2	X3	X4	X5
A	200	0.1	0.1	0.15	0.1	0.25	0.1	0.2	0.35	0.25	0.25	0.4	0.2
B	500	0.1	0.1	0.1	0.1	0.1	0.1	Sd of Ps	Sd of Ps	0.3	Sd of Ps	0.5	0.35
A	500	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.25	0.15	0.1	0.15	0.25
B	500	0.1	0.15	0.15	0.1	0.3	0.3	0.1	0.1	0.2	0.2	0.5	0.1
A	1000	0.1	0.3	0.1	0.1	0.45	0.35	0.15	0.1	0.1	0.1	0.5	0.1
B	1000	0.1	0.1	0.1	0.1	0.45	0.5	0.1	0.1	0.3	0.1	0.5	0.1

Şekil-1. Sadece kategorik değişkenlerden oluşan modele ilişkin farklı senaryolarda en başarılı eşleştirme dengesine sahip caliper değerleri (Sd of Ps: propensity skorların standart sapması)

23. Ulusal ve 6. Uluslararası Biyoistatistik Kongresi

26-29 Ekim 2022, Ankara Üniversitesi Tıp Fakültesi

Parametreler		Standartlaştırılmış Ortalamalar Arası Fark									
Senaryo	Örneklem Büyüklüğü	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
A	200	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,25	0,1	0,1
B	200	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2
A	500	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,3
B	500	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2
A	1000	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,5
B	1000	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,4
		Kolmogorov-Smirnov İstatistikleri									
Senaryo	Örneklem Büyüklüğü	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
A	200	0,15	0,25	0,15	Sd of Ps	0,15	0,2	0,15	0,4	0,2	0,4
B	200	0,2	Sd of Ps	0,2	0,2	Sd of Ps	0,15	0,15	0,4	0,25	0,5
A	500	0,1	0,1	0,1	0,15	0,1	0,1	0,1	0,4	0,1	0,5
B	500	0,1	0,15	0,1	0,1	0,1	0,1	0,1	0,3	0,2	0,35
A	1000	0,1	0,1	0,1	0,1	0,1	0,1	0,1	Sd of Ps	0,1	0,5
B	1000	0,1	0,15	0,1	0,1	0,1	0,1	0,1	Sd of Ps	0,1	0,5
		Varyans Oranları									
Senaryo	Örneklem Büyüklüğü	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
A	200		Sd of Ps		0,15			0,1			0,4
B	200		0,25		0,15			0,1			0,5
A	500		0,15		0,1			0,1			0,15
B	500		0,1		0,15			0,1			0,35
A	1000		0,1		0,1			0,1			0,3
B	1000		0,1		0,1			0,1			0,3

Şekil-2. Hem kategorik hem sayısal değişkenlerden oluşan modele ilişkin farklı senaryolarda en başarılı eşleştirme dengesine sahip caliper değerleri (Sd of Ps: propensity skorların standart sapması)

Görüntü İşleme Yöntemlerinin COVID-19 Pozitif Tespitinde Klasik Veri Madenciliği Yöntemleri ile Sınıflandırmaya Etkisi

¹Fatma Gül KURT, ²Jale KARAKAYA

¹Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

²Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Amaç: Dünya Sağlık Örgütü (DSÖ), Mart 2020’de SARS-CoV-2’nin neden olduğu bulaşıcı bir salgın olan yeni koronavirüsü (COVID-19) pandemi olarak ilan etmiştir. Erken tanı virüsün yayılmasını azaltabileceğinden, akıllı tahmin ve tanı araçlarına sahip olmak önemlidir. Günümüzde tıbbi görüntüler yardımı ile birçok sağlık problemi tespitinde (beyin tümörü tespiti, meme kanseri tespiti, zatürre tespiti vb.) yapay zeka tabanlı çözümlerden yararlanılmaktadır. İçerisinde bulunduğumuz bu pandemi sürecinde de COVID-19’un tanı süresini kısaltmak ve doğru tanı oranını arttırmak için yapay zeka tabanlı birçok çalışma yapılmıştır. Görüntülerdeki yüksek frekanslı nesne kenarlarının geliştirilmesi görüntü kalitesinde iyileşmeye yol açar. Böylece öznetelik çıkarımı aşamasında nesnelerin belli sınıflara atanması kolaylaşır bu da sınıflandırma başarısını artırır. Çalışmanın amacı görüntü işleme tekniklerinin COVID-19 ve Non-COVID akciğer grafi gruplarını sınıflandırma performansına etkisini incelemektir.

Gereç ve Yöntem: Çalışmada, açık erişimli Kaggle platformundan elde edilen, akciğer grafilerinin bulunduğu COVID-19 Radiography veri tabanından yararlanılmıştır. COVID-19 ve Non-COVID akciğer grafi gruplarının her birinden rastgele 2784 akciğer grafisi seçilmiştir. Akciğer grafileri tıbbi görüntü işleme ile sayısallaştırılmıştır. Sayısal görüntülere Bulanık Maskeleme (USM) görüntü işleme yöntemi uygulanarak görüntü kalitesinin iyileştirilmesi hedeflenmiştir. Bir derin öğrenme yöntemi olan VGGNet-16 (VGG16) evrişimli sinir ağı mimarisi ile görüntülerden öznetelik çıkarımı gerçekleştirilmiştir. Destek Vektör Makineleri (DVM), Rastgele Orman (RO) ve Yapay Sinir Ağları (YSA) algoritmaları ile görüntüler sınıflandırılırken uygulanan görüntü işleme tekniklerinin sınıflandırma performansına etkisi incelenmiştir. Algoritmaların sınıflandırma performansları doğruluk, kesinlik, duyarlılık ve F1 ölçüleri ile değerlendirilmiştir.

Bulgular: Klasik veri madenciliği algoritmalarında sınıflandırma performansı başarıları sırasıyla RO>DVM>YSA iken görüntülerden öznetelik çıkarımı VGGNet-16 evrişimli sinir ağı mimarisi ile gerçekleştirildiğinde algoritmaların sınıflandırma başarıları sırasıyla YSA>DVM>RO şeklindedir. Orijinal görüntüler sınıflandırıldığında en yüksek sınıflandırma performansı VGG16+YSA algoritması uygulandığında %95,617 doğruluk ile elde edilmiştir. Görüntülere USM görüntü işleme yöntemi uygulandığında ise en yüksek sınıflandırma performansı %97,09 doğruluk ile VGG16+YSA algoritması sonucunda elde edilmiştir.

Sonuç: VGG16 evrişimli sinir ağı mimarisi ile öznetelikler çıkarıldığında elde edilen sınıflandırma performansları sadece klasik veri madenciliği algoritmaları ile elde edilen sonuçlara göre oldukça yüksek bulunmuştur. Uygulanan USM görüntü işleme tekniği, orijinal görüntülere göre bütün sınıflandırma algoritmalarında sınıflandırma performansını arttırmıştır.

Anahtar Kelimeler: Görüntü İşleme, Veri Madenciliği, Evrişimli Sinir Ağları, COVID-19

SB – 19

Covid-19 Tanısında Konvolüsyonel Sinir Ağları Mimarilerinin Karşılaştırılması

¹İşıl ÜNALDI, ²Leman TOMAK

¹Ondokuz Mayıs Üniversitesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Samsun

²Ondokuz Mayıs Üniversitesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Samsun

Amaç: COVID-19'un bulaş hızının önüne geçebilmek için erken tanının yüksek doğrulukla yapılması esastır. Bu çalışma ile, göğüs BT (bilgisayarlı tomografi) görüntüleri üzerinden, COVID-19'un doğru ve hızlı bir şekilde tespit edilebilmesine olanak sağlayan, önceden eğitilmiş KSA (konvolüsyonel sinir ağları) mimarilerinin çeşitli performans ölçülerıyla sınıflandırma başarılarının karşılaştırılarak değerlendirilmesi amaçlanmaktadır.

Gereç ve Yöntem: Çalışmada, Kaggle üzerinden erişilen açık kaynak SARS-COV-2 Ct-Scan Dataset kullanıldı. Veride, toplam 2482 hastanın BT görüntüsü olup, bunlardan 1252 hasta Covid-19 tanısı almış, 1230 hasta Covid-19 tanısı almamıştır. Hastalığın sınıflandırması KSA mimarilerinden InceptionResNetV2, DenseNet169, InceptionV3, MobileNetV2, ResNet50, VGG16 ve VGG19 ile yapıldı. Mimarilerin performansları doğruluk, duyarlık, seçicilik ve F1 skoru ile değerlendirildi. Verilerin analizinde Python programlama dili ile Keras kütüphanesi kullanıldı.

Bulgular: Bu çalışmada en yüksek performansa sahip mimariler; %99,0 doğruluk, %99,1 duyarlık, %98,9 seçicilik ve %98,9 F1 skoru değerleri ile DenseNet169; %98,8 doğruluk, %99,1 duyarlık, %98,5 seçicilik ve %98,7 F1 skoru ile InceptionResNetV2 olurken, en düşük performans değerine sahip mimari %93,0 doğruluk, %96,0 duyarlık, %90,5 seçicilik ve %92,4 F1 skoru ile VGG19 oldu. Diğer mimarilerin performans değerleri; MobileNetV2 için %98,0 doğruluk, %96,4 duyarlık, %99,3 seçicilik ve %97,7 F1 skoru; InceptionV3 için %98,0 doğruluk, %98,7 duyarlık, %97,4 seçicilik ve %97,8 F1 skoru; ResNet50 için %98,4 doğruluk, %97,3 duyarlık, %99,3 seçicilik ve %98,2 F1 skoru; VGG16 için %97,4 doğruluk, %99,1 duyarlık, %96,0 seçicilik ve %97,1 F1 skoru olarak belirlendi.

Sonuç: Elde edilen sonuçlar değerlendirildiğinde, COVID-19 tanısında kullanılan KSA mimarileri için her bir performans ölçüsünün %90'ın üzerinde değere sahip olduğu görüldü. Bu çalışmada, COVID-19'un BT görüntüleri kullanılarak tespit edilmesinde KSA yönteminin, karar alıcılara yardımcı olacak otomatik bir karar mekanizması olduğu, bu sayede hastalığa en kısa sürede tanı konulabileceği ve böylece tedavinin etkinliği artırılarak ölümlerin önüne geçilebileceği ortaya konulmuştur.

Anahtar Kelimeler: Covid-19, derin öğrenme, konvolüsyonel sinir ağları, görüntü sınıflandırma, bilgisayarlı tomografi

Taguchi ve Faktöriyel Deney Tasarım Metotlarının Karşılaştırması: *Lactobacillus* spp. ERIC-PCR Optimizasyonu Örneği

¹Selen YILMAZ İŞIKHAN, ²Ceren ÖZKUL

¹Hacettepe Üniversitesi Sosyal Bilimler Meslek Yüksekokulu

²Hacettepe Üniversitesi Eczacılık Fakültesi, Farmasötik Mikrobiyoloji Anabilim Dalı

Amaç: Faktöriyel tasarımlar çok güçlü istatistiksel araçlardır çünkü bir araştırmacının sonuç değişkeni üzerindeki çoklu faktör düzeyindeki kombinasyonların etkilerini aynı anda test etmesine izin verir. Ancak bazen faktöriyel deneylerde kullanılan faktör sayıları ve seviyeleri zor ve pahalıdır. Taguchi yöntemi, ortogonal diziler kullanılarak, faktöriyel tasarıma göre deney zamanını ve maliyetini en aza indiren bir tasarımdır. Ayrıca, deney birimlerinin sayısının azaltılması, kontrol edilemeyen faktörlerin etkilerinin azaltılmasına izin verir. Bu çalışmada deney tasarım metotlarından Taguchi ve faktöriyel tasarımların performansını *Lactobacillus* spp.'nin ERIC-PCR ile parmak izi analizleri üzerinde karşılaştırmak hedeflenmiştir. Probiyotik mikroorganizmaları içeren *Lactobacillus* genusu içerisindeki türlerin, DNA parmak izi analizleri ile profilendirilmesi ve farklı türlerin tespiti özellikle anne sütü gibi klinik örneklerde kullanılan yaklaşımlardır.

Gereç ve Yöntem: Uygulamada *L. plantarum* 299v suşu için ERIC-PCR optimizasyon verileri kullanılmıştır. Saf *L. plantarum* 299v kültüründen DNA izolasyonu sonrası, DNA miktar ve saflığı ölçülmüştür. PCR basamakları için 4 faktörün varyasyonları kullanılmıştır (MgCl₂, dNTP, primer, DNA miktarı). ERIC primer dizileri kullanılarak gerçekleştirilen PCR sonrası, ampikonlar %1.8 agaroz jelde yürütülerek Etidium bromür varlığında bantlar görüntülenmiş ve bant sayıları belirlenmiştir. Bu deneyde temel amaç en yüksek bant sayısını, dolayısıyla ayrımı sağlayan faktör düzeylerini tespit etmektir. Verilerin deneysel tasarımı 3x3x3x3 (L9, 34) faktöriyel tasarımıdır. Tam faktöriyel deney tasarımı ile 81 deney yapılması gerekirken, Taguchi L9 deney tasarım tablosu kullanılarak 9 farklı değer içeren deney seti ile çalışma yapılmıştır. Taguchi yöntemi ve Faktöriyel deney sonuçları karşılaştırmalı olarak incelenmiştir.

Bulgular: Taguchi yöntemi, daha az sayıda deney ünitesi ile *Lactobacillus* spp. bant sayısına etkisi olan dNTP(μM), MgCl₂ (mM), DNA (ng) ve Primer (μM)'ün en iyi kombinasyonunu belirlemiştir. Bant sayısı üzerinde en etkili ilk faktör Primer (μM), ikincisi dNTP (μM) olarak belirlenmiştir. MgCl₂ (mM) ile DNA (ng) ise aynı Delta değerini (rank) verdiklerinden istatistiksel olarak anlamlı olmayan eşit ve daha az bir etkiye sahiptirler. Taguchi modeli sonucunda her bir primer konsantrasyonun 0.9 μM olduğu koşulda en yüksek bant sayısı 12 olarak elde edilmiştir (p=0.033).

Sonuç: Taguchi deney tasarım metodu faktöriyel tasarıma kıyasla, hangi faktörlerin sonuç değişkeni üzerinde daha etkili olduğunu, ayrıca hangi alt düzeylerin minimum maliyet ve optimum performansı sağladığını tasarımın başında hızlıca anlamamızı sağlamaktadır. Günümüzde yaygın olarak kullanılmayan Taguchi yöntemi, ERIC-PCR gibi DNA parmak izi analizlerinde PCR reaktiflerinin konsantrasyonunu belirlemek için maliyet ve süreyi azaltan bir yaklaşım olarak kullanılabilir.

Anahtar Kelimeler: Taguchi deney tasarımı, faktöriyel tasarım, ortogonal dizi optimizasyonu, *Lactobacillus*, ERIC-PCR

Eğilim Skoru Eşleştirmesinde Greedy ve Optimal Yöntemlerin Karşılaştırılması

¹Gülden HAKVERDİ, ²Timur KÖSE, ³Cemil ÇOLAK

¹Ege Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, İzmir

²Ege Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, İzmir

³İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Malatya

Amaç: Eğilim skoru, randomizasyonun sağlanamadığı gözlemsel çalışmalarda vaka ve kontrol gruplarında ortak değişkenler bakımından var olan dengesizliği gidermek için kullanılan bir ölçüdür. Ortak değişkenler açısından vaka-kontrol gruplarında dengesizlik olduğunda sonuç hakkında yorum yapmak yanlılığa neden olmaktadır. Bu yanlılığı gidermek için sıklıkla eğilim skoru eşleştirmesi kullanılmaktadır. Bu çalışmanın amacı açık kaynak (hipotiroid) veri setinde eğilim skoru eşleştirme yöntemlerinden greedy ve optimal eşleştirme yöntemlerini kullanarak çalışma gruplarını dengelemek ve sonuçları karşılaştırmaktır.

Gereç ve Yöntem: Mevcut araştırmada UCI hipotiroid veri seti kullanılmış olup, veri seti 135 vaka ve 2138 kontrol olmak üzere toplam 2273 bireyden oluşmaktadır. Yaş, cinsiyet, hipotiroid, tiroksin, t4u ortak değişkenleri için çoklu lojistik regresyon analizi ile eğilim skoru tahmin edildikten sonra gruplar greedy ve optimal eşleştirmesi ile sırasıyla 1:1, 1:2 ve 1:3 oranında eşleştirilmiştir. Greedy ve optimal eşleştirme sonuçları denge bakımından karşılaştırılmıştır. Bütün sonuçlar için standartlaştırılmış ortalama farkları, varyans oranı, vb. hesaplanarak değerlendirilmiştir. Bütün hesaplamalarda R programlama dili ve paketleri kullanılmıştır.

Bulgular: Çalışmada kullanılan yöntemlerin bulguları incelendiğinde; optimal ve greedy eğilim skoru eşleştirme yöntemi kullanılarak 1:1 (135 vaka, 135 kontrol), 1:2 (135 vaka, 270 kontrol) ve 1:3 (135 vaka, 405 kontrol) oranında eşleştirme yapılmıştır. Eşleştirme sonuçlarına göre, yaş değişkeni dışındaki diğer değişkenlerdeki ortalama farklar ve varyans oranları, optimal eşleştirmede greedy eşleştirmeye göre daha küçük olarak elde edilmiştir. Tüm değişkenler için standartlaştırılmış ortalama farkları <0.1 ve varyans oranı <2 elde edilerek gruplar arasında denge sağlanmıştır.

Sonuç: Vaka ve kontrol grupları arasında ortak değişkenler açısından dengesizlik olması yanlı sonuçlar elde edilmesine sebep olur. Bu durum sağlık alanında yapılan tedavi veya müdahalelerin sonuçlarını olumsuz etkilemektedir. Çalışmada elde edilen bulgulara göre optimal eşleştirme greedy eşleştirme yöntemine göre ortalama farklar daha küçük elde edilmiştir. Dolayısıyla optimal eşleştirmede eşleşen çiftler uzaklık olarak birbirine daha yakın bulunmuştur. Bu çalışmadaki veriler açık kaynaklı olmakla birlikte elde edilen sonuçların literatürle benzer olduğu belirlenmiştir.

Anahtar Kelimeler: Eğilim skoru, eşleştirme, ortak değişken, vaka-kontrol

Veri Madenciliğinde Hibrit Model Yaklaşımı

¹Batuhan BAKIRARAR, ²Atilla Halil ELHAN

¹Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara, Türkiye

²Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara, Türkiye

Amaç: Çalışmanın amacı, son teknoloji veri madenciliği algoritmalarını ve uygulamalarını sunmak ve tıbbi verilerin kümelenmesi ve sınıflandırılması için yeni bir hibrit veri madenciliği yaklaşımı önermektir. Ayrıca çalışmada, dengeli ve dengesiz dağılıma sahip veri setlerinde, farklı örneklem büyüklüklerinde ve farklı değişkenler arası ilişkiler olması durumunda performans ölçütlerinin hesaplanması ve bu ölçütlerin hibrit modelden elde edilen ölçütler ile karşılaştırılması amaçlanmıştır.

Gereç ve Yöntem: Çalışmada UCI veri tabanından alınan hepatit veri seti kullanılmıştır. Sık kullanılan ve veri setlerinde en iyi performansa sahip denetimli öğrenme algoritmalarından Karar Ağaçları, Destek Vektör Makinesi, Random Forest, Naive Bayes ve K-en yakın komşu'nun yanı sıra Lojistik Regresyon ve Yapay Sinir Ağları algoritmaları, denetimsiz öğrenme algoritmalarından ise K-ortalama kullanılmıştır. Ayrıca kullanılan denetimli ve denetimsiz öğrenme algoritmaları birleştirilerek hibrit modeller oluşturulmuştur. Performans ölçütü olarak dengesiz veri setlerinde daha yansız sonuçlar verdiği için MCC belirlenmiştir. MCC değeri benzer olan modellerde ise bu ölçüte ek olarak Accuracy ve F-measure değerlerine bakılması uygun görülmüştür. Veri setinde ilk aşamada gerçek değerler kullanılarak değerlendirme yapılmış, daha sonra bu verinin tüm özellikleri korunarak örneklem büyüklüğü 250 ve 500'e yükseltilmiştir.

Bulgular: Gerçek veri setinde ve 250, 500 örnekleme sahip veri setlerinde performans ölçütleri hesaplanıp karşılaştırmalar yapıldığında ise örneklem büyüklüğü arttıkça performans ölçütlerinin de arttığı görülmüştür. Gerçek veri seti için MCC değeri 0,730 olarak bulunurken, bu değer 250 örnekleme sahip veri setinde 0,798 ve 500 örnekleme sahip veri setinde 0,920 olarak bulunmuştur.

Sonuç: Tüm veri setlerinde hibrit model daha iyi sonuçlar vermiştir. Uygun veri madenciliği yöntemlerinin birlikte kullanımı durumunda, yapılacak tüm çalışmalarda hibrit model yaklaşımının daha iyi modeller elde etmemizi sağlayacağı sonucuna ulaşılmaktadır.

Anahtar Kelimeler: Hibrit model, veri madenciliği, performans ölçütleri

Küçük Küme ve Periyotlu Boylamsal Çalışmalarda Değişken ve Model Seçimi

¹Merve TURKEGÜN, ²Bahar TAŞDELEN, ³Saim YOLOĞLU

¹Mersin Üniversitesi Biyoistatistik Bölümü, Mersin, Türkiye

²Mersin Üniversitesi Biyoistatistik Bölümü, Mersin, Türkiye

³Mersin Üniversitesi Biyoistatistik Bölümü, Mersin, Türkiye

Amaç: Boylamsal verilerde, küçük küme sayısının sebep olduğu en önemli problemlerden biri bir ya da daha fazla açıklayıcı değişkenin ikili sonuç değişkenini mükemmel ya da mükemmele yakın olarak tahminlemesi ile ortaya çıkan tam ayrılma ya da aşırı uyum durumudur. Aşırı uyumun gözlemlendiği durumlarda adimsal değişken seçim yöntemlerinin kullanımı katsayıların standart hatalarını şişirerek yanlış tahminlere sebep olur. Aşırı uyum probleminin çözümü için tüm açıklayıcı değişkenleri aynı anda değerlendirerek, regresyon katsayı tahminlerini bir ceza fonksiyonu yardımıyla sıfıra doğru küçültme cezalandırma yöntemlerinin kullanımı önerilmektedir. Çeşitli ceza fonksiyonları ile Genelleştirilmiş Tahmin Eşitlikleri (GEE) yöntemini birleştiren Cezalı Genelleştirilmiş Tahmin Eşitlikleri (PGEE) yöntemi, aşırı uyum durumunda ve çok yüksek boyutlu boylamsal verilerde değişken seçimi için sıklıkla kullanılan bir yöntemdir. Amacımız; aşırı uyum problemine sebep olan açıklayıcı değişken sayısının küme sayısında

Gereç ve Yöntem: 500 tekrarlı simülasyon verisinin üretimi ve analizler için R programlama dili kullanılmıştır. Simülasyon çalışmasında, popülasyonda nadir görülen hastalıkların araştırıldığı klinik çalışmalarda sıklıkla karşılaşılan veri yapılarına ve çalışma kısıtlılıklarına benzerlik göstermesi amaçlanmıştır. Bu benzerlikler, ölçümlerin tekrar ve küme sayıları dikkate alınarak yakalanmaya çalışılmıştır. Tekrar sayısı sayısının minimumda olduğu ($k=2$), küme sayısının (N); 20, 30, 50, 100 ve 200 olduğu durumlar için açıklayıcı değişken sayısı (P); 10, 20, 50 olacak şekilde ve açıklayıcı değişkenler arası korelasyonun $r=0.20$, 0.50 ve 0.80 olan çok değişkenli normal dağılıma uygun veri üretilmiştir. İkili yapıda sonuç değişkeni binom dağılımından üretilmiştir. Sonuç değişkeninin tekrarlanan ölçümü için gözlemler arasındaki korelasyon 0.40 alınmış bu sebeple çalışan korelasyon yapısı sadece “değiştirilebilir” olarak seçilebilmiştir. GEE ve PGEE model performansları, katsayılar ait Medyan Kare Hatası (MEDSE), Medyan Mutlak Hatası (MDAE) ve Bias istatistikleri ile değerlendirilmiştir. PGEE modelinde değişken seçiminde, katsayı değeri $>0,001$ olanlar sıfırdan farklı kabul edilmiştir. Değişken seçim performansları ise doğru seçim (Correct) ve hatalı seçim (Incorrect), kesinlik (Precision), hassasiyet (Recall) ve F1 skor kriterleri kullanılmıştır.

Bulgular: Değişken ve küme sayısının birbirine yakın olduğu durumlarda ($P=10, 20, 50$; $N=20, 30, 50$) GEE katsayılarının MEDSE, MDAE ve Bias değerleri çok büyük hesaplanmıştır. PGEE ise tüm senaryolarda 0’a yakın MEDSE ve MDAE değerlerine sahip olmuştur. GEE etkisiz sayılabilecek tüm katsayıları (sıfıra eşit olan) sıfırdan farklı tahminlemiştir. Bu sebeple kesinlik ve F1 skoru hesaplanamamıştır. Hassasiyet değeri ise tüm senaryolarda 0 olarak hesaplanmıştır. Değişken sayısı ile küme sayısının artırılması PGEE’nin tahmin gücünü arttırmıştır ($P=50, N=200, F1 \text{ skor}=0,78$).

Sonuç: PGEE, $P>n$ durumunda hem küme içindeki hem de değişkenler arasındaki korelasyonlu yapıyı açıklamada GEE’ye göre daha başarılı olmuştur. PGEE, hem küme içindeki hem de değişkenler arasındaki ilişkileri kontrol altında tutarak daha tutarlı sonuçlar üretmiştir. PGEE’nin, değişken seçiminden ziyade değişken ayıklama mantığıyla çalıştığı görülmüştür. Doğrudan etkili değişkenleri seçmek yerine dolaylı olarak seçerek, modele etkisi olmayan değişkenleri tespit etmede daha başarılı olduğu tespit edilmiştir.

Anahtar Kelimeler: GEE, cezalı GEE, değişken seçimi, model seçimi

SB – 25

MEVARYA

Mutlu UMAROĞLU

Orta Doğu Teknik Üniversitesi

Amaç: Tıp, sağlık bilimleri, sosyal bilimler ve eğitim bilimleri gibi alanlarda ölçek geliştirme çalışmaları sıklıkla yapılmaktadır. Yapılan bu ölçek geliştirme çalışmaları farklı ülkelerde kullanılan ölçeklerin anadile uyarlaması olabileceği gibi belirli bir konuya açıklık getirmek amacı ile yeni bir ölçek geliştirme şeklinde de olabilmektedir. Ölçek geliştirme çalışmalarında izlenmesi gereken bazı adımlar ve değerlendirilmesi gereken bazı istatistikler bulunmaktadır. Bu adımlar ölçme aracının ilgili özelliği doğru bir şekilde ölçüp ölçmediğinin değerlendirildiği geçerlik ve ölçeğin aynı durumdaki kişilere uygulandığında alınan sonuçların benzerliğinin değerlendirildiği güvenilirlik olarak adlandırılmaktadır. Ancak ölçek geliştirme çalışmalarında bazı geçerlik-güvenirlik adımlarının uygulanmadığı ve bazı istatistikler sunulmadığı görülmektedir.

Gereç ve Yöntem: MEVARYA ölçek geliştirme çalışmalarında sunulması gereken istatistiklerin hesaplanmasını sağlayan kapsamlı bir web uygulamasıdır. MEVARYA geçerlik-güvenirlik çalışmalarında kodlarla ve karmaşık menülerle uğraşmadan sistematik bir şekilde uygulanması gereken tüm analizlerin gerçekleştirilmesine olanak sağlar.

Bulgular: MEVARYA tanımlayıcı istatistikler, açıklayıcı ve doğrulayıcı faktör analizi, Cronbach alfa katsayısı, ICC, Krippendorff alfa katsayısı, Bland-Altman grafiği gibi araçları bünyesinde barındırır. Bu web uygulamasının diğer uygulamalara göre avantajları ise web tabanlı olması nedeni ile bilgisayardan ve işletim sisteminden bağımsız olarak çalışabilmesidir.

Sonuç: Sonuç olarak piyasada bulunan ücretsiz ve ticari pek çok yazılım ile geçerlik-güvenirlik çalışmalarında kullanılan istatistikleri hesaplamak mümkündür. Kullanılacak istatistiklerin seçimi ve yorumlanması konusunda rehber içermeyen bu yazılımlara alternatif olarak MEVARYA geliştirilmiştir.

Anahtar Kelimeler: Ölçek geliştirme, geçerlik, güvenilirlik, web uygulaması

Makine Öğrenmesi Algoritmalarının Performanslarının Farklı Koşullar Altında Karşılaştırılması: Bir Monte Carlo Benzetim Çalışması

¹Müge COŞKUN, ²Osman DAĞ

¹Biyoistatistik Uygulama ve Araştırma Birimi, Lokman Hekim Üniversitesi, Ankara, Türkiye

²Biyoistatistik Anabilim Dalı, Hacettepe Üniversitesi, Ankara, Türkiye.

Amaç: Bu çalışmada, makine öğrenmesi algoritmalarını karşılaştırmak için bir Monte Carlo benzetim çalışması yapılarak ilgili yöntemlerin performanslarının karşılaştırılması amaçlanmıştır.

Gereç ve Yöntem: Araştırmada farklı koşulların makine öğrenmesi algoritmaları üzerine etkisini araştırmak için bir Monte Carlo benzetim çalışması yapılmıştır. Benzetim çalışması 162 farklı senaryo altında oluşturulmuştur. BinNor R paketi altında jointly. generate.binary. normal fonksiyonu kullanılarak veri türetilmiştir. Bağımsız değişkenler iki sınıflı (%40) ve sürekli (%60) değişken olarak türetilmiştir. Benzetim çalışması aşağıdakilerin tüm olası kombinasyonlarını içerir (Dağ ve ark., 2019). • Bağımsız değişkenlerin sayısı: 5, 10, 15; • Gözlem sayısı: 50, 100, 500; • Bağımlı ve bağımsız değişkenler arasındaki korelasyon: 0.2 (düşük), 0.5 (orta), 0.8 (yüksek); • Bağımsız değişkenler arasındaki korelasyon: 0- 0.1 (düşük), 0.4- 0.5 (orta), 0.8- 0.9 (yüksek); • Pozitiflerin oranı: 0.3, 0.5. Model performanslarını değerlendirebilmek için veri, eğitim (%80) ve test (%20) olarak ikiye ayrılmıştır. Test kümeleri için gerçek ve kestirilen sınıflar kullanılarak elde edilen karışıklık matrislerine dayalı olarak doğru sınıflama oranı, F1, duyarlılık ve seçicilik ile yöntemlerin performansı araştırılmaktadır. Karışıklık matrisi için GMDH2 paketin altında confMat fonksiyonu kullanılmıştır. Benzetim senaryoları 10,000 kez tekrarlanmıştır.

Bulgular: Gözlem sayısı ile bağımlı ve bağımsız değişkenler arasındaki korelasyon arttıkça doğru sınıflama oranı, duyarlılık ve seçicilik artış göstermektedir. Pozitiflerin oranı 0.5 iken çoğu senaryo altında doğru sınıflama oranı, duyarlılık ve seçicilik bakımından ridge regresyon daha iyi performans vermiştir. Bağımsız değişkenler arası korelasyon orta iken gözlem sayısı düşük olduğunda rastgele orman daha iyi performans vermiştir. Bağımsız değişkenler arası korelasyon 0- 0.1 ve pozitiflerin oranı 0.3 iken çoğu senaryo altında doğru sınıflama oranı ve duyarlılık bakımından ridge regresyon daha iyi performans verirken, seçicilik bakımından rastgele orman daha iyi performans vermiştir. Bağımsız değişkenler arası korelasyon 0.4- 0.5 ve pozitiflerin oranı 0.3 iken çoğu senaryo altında doğru sınıflama oranı ve duyarlılık bakımından ridge regresyon daha iyi performans verirken, seçicilik bakımından rastgele orman daha iyi performans vermiştir. Ancak aynı koşullarda gözlem sayısı küçükken rasgele orman daha iyi performans vermiştir. Bağımsız değişkenler arası korelasyon 0.8- 0.9 ve pozitiflerin oranı 0.3 iken çoğu senaryo altında doğru sınıflama oranı ve seçicilik bakımından rastgele orman, duyarlılık bakımından lightgbm daha iyi performans vermiştir.

Sonuç: Bu çalışmada; xgboost, lightgbm, rastgele orman, destek vektör makineleri, ridge regresyon ve GMDH yöntemleri Monte Carlo benzetim çalışmasıyla karşılaştırılmıştır. Bu benzetim çalışması ışığında ridge regresyon algoritmasının çoğu senaryo altında daha iyi performans gösterdiği saptanmıştır.

Anahtar Kelimeler: Makine öğrenmesi algoritmaları, veri madenciliği, monte carlo benzetim çalışması

Web Tabanlı Açıklanabilir Yapay Zeka (XAI) arayüzü: Akut Apandisit'in Yorumlanabilir Tahmini İçin Bir Uygulama

¹Fatma Hilal YAĞIN, ²İlhami SEL, ³İpek BALIKÇI ÇİÇEK, ⁴Davut HANBAY, ⁵Cemil ÇOLAK, ⁶Sami AKBULUT

¹İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Malatya Türkiye

²İnönü Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü

³İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Malatya Türkiye

⁴İnönü Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü

⁵İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Malatya Türkiye

⁶İnönü Üniversitesi Tıp Fakültesi Genel Cerrahi Anabilim Dalı, Malatya

Amaç: Bu çalışma, veri ön işleme analizleri de dahil olmak üzere sınıflandırma görevlerini gerçekleştiren ve makine öğrenimi (ML) modellerine dayalı açıklanabilir yapay zeka (XAI) yöntemlerini içeren açık erişimli web tabanlı bir arayüz geliştirmeyi amaçlamaktadır. Ayrıca bu çalışmanın diğer bir amacı ise geliştirilen web arayüzünün klinik uygulaması için, özellikle komorbiditesi olan hastalarda perforasyon, serbest peritonit, apse, pyleflebit plastron, morbidite ve ölümlerle sonuçlanabilecek akut apandisit (AAP) yorumlanabilir tahminini gerçekleştirmektir.

Gereç ve Yöntem: XAI web arayüzünü kullanarak, AAP'nin yorumlanabilir tahmini için apendektomi uygulanan 6094 hastanın biyokimyasal kan parametre ölçümleri geriye dönük olarak analiz edildi. Olası eksik değerler Rastgele Orman yöntemi kullanılarak atanmıştır. Veri setindeki sınıf dengesizliği problemi SMOTETomek yaklaşımı ile çözülmüştür. AAP tahmini için XAI'de farklı ML algoritmaları test edildi. ML modellerini değerlendirmek için iç içe çapraz doğrulama uygulandı. Modellerin performansını değerlendirmek için işlem karakteristik eğrisi (AUC), doğruluk, F1-puanı, kesinlik, duyarlılık, yanlış pozitif oran (FPR) ve negatif tahmin değeri (NPV) hesaplandı. En uygun modeli seçtikten sonra, sonuca katkısına (AAP) dayalı olarak her bir tahmincinin önemini incelemek ve tahminleri açıklamak için Shapley'nin katkı değerleri hesaplandı. Python tabanlı Dash paketi aracılığıyla geliştirilen XAI web arayüzü, bir algoritma oluşturmak için gereken süreyi azaltma, kodun yeniden kullanımını teşvik etme ve kodsuz bir kullanıcı arayüzü kullanarak teknik olmayan kullanıcılar için XAI görevlerini yaygınlaştırma potansiyeli ile geliştirildi.

Bulgular: XAI arayüzünde AAP'yi tahmin etmede, XGBoost algoritması en iyi performansı gösterdi (AUC = 0.977, doğruluk = 0.931, F1 skoru = 0.932, kesinlik = 0.908, hatırlama = 0.958, FPR = 0.044, NPV = 0.905). Modelin ana belirleyicileri, önem sırasına göre Total Bilirubin, Beyaz Kan Hücreleri (WBC) ve Ortalama Trombosit Hacmi (MPV) idi. SHAP sonuçları, yüksek Total Bilirubin, Lenfosit, Trombosit ve Kırmızı Hücre Dağılım Genişliği (RDW) değerlerine ek olarak, düşük WBC ve MPV değerlerinin AAP hastalarını ayırt etmede öncül olduğu gösterildi. Önerilen arayüzün, birden fazla veri türünü işlemek ve XAI prosedürünü uygulamak için etkili olduğu gösterilmiştir.

Sonuç: Önerilen ML modeli (XGBoost) AAP'yi tahmin etmek için başarılı olduğu görülmüştür. Çalışmada, kara kutu ML yöntemlerinden yorumlanabilir tahminler elde etmek için "XAI: Açıklanabilir Yapay Zeka Arayüzü v.1.0" geliştirilmiştir. Açık erişimli web arayüzüne (<http://biostatapps.inonu.edu.tr/XAI/>) adresinden erişilebilir. Geliştirilen XAI web arayüzü, araştırmacıların kapsamlı veri ön işleme, ML, doğrulama ve XAI yöntemleriyle analizleri kolay ve hızlı bir şekilde gerçekleştirmesini sağlayabilecektir.

Anahtar Kelimeler: Akut apandisit, açıklanabilir yapay zeka, web arayüzü, LIME, SHAP.

Kantil Regresyon Yaklaşımı İçin İnteraktif Bir Arayüz Tasarımı

¹Abdulvahap PINAR, ²Ahmet Kadir ARSLAN

¹İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Malatya

²İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Malatya

Amaç: Regresyon analizi, bağımlı değişken ile bağımsız değişken(ler) arasındaki ilişkinin matematiksel yorumu olarak adlandırılması yönünde En Küçük Kareler (EKK) yöntemine dayalı istatistiksel bir yöntemdir. Kantil regresyon, klasik regresyon varsayımı olan hata teriminin normal dağılmaması ve uç değerlere karşı etkin olması bakımından sağlam bir tekniktir. Klasik regresyon yöntemlerinin uygulanmadığı durumlarda kuantil regresyon sıklıkla kullanılan yöntemlerin başında gelir. Tıpta referans grafikleri uygun kantillerden oluşan normalden farklı bazı özel ölçüm değeri olan kişileri belirlemek için kullanılır. Yaşam analizleri için uygulamalarda ise bağımsız değişkenler orta, düşük ve yüksek riskli bireyler üzerinde farklı bir etkiye sahip olabilir. Bu etkileri yaşam süresi için farklı kantil değerleri ele alınarak anlaşılabilir. Sağlık çalışmalarında (uzunluk, kilo, hastalıkların artışı, uyuşturucu kullanım ile ilgili) tedavi yöntemlerin belirlenmesi gibi konular için web üzerinde çalışabil

Gereç ve Yöntem: Detaylı Kantil Regresyon Analiz Yazılımı, Shiny paketi yardımı ile web tabanlı arayüzü oluşturulmuştur. Geliştirilen yazılım, xls/.xlsx, .arff, .sav ve .csv/.txt uzantılı veri dosyalarını desteklemektedir. Geliştirilen yazılım, bağımsız değişkenler tek tek ya da istenirse etkileşimli olarak modele eklenebilmesine imkân tanımaktadır. Ayrıca, model performansını artırmaya yönelik olarak Çapraz Doğrulama (Cross-Validation) ve Düzenleştirme (Lasso, Elastic-net) yöntemleri de geliştirilen yazılımda uygulanabilmektedir.

Bulgular: Geliştirilen yazılımda bulgular, modele ilişkin genel bulgular, varyans analizi tablosu, ortalama karesel hata (MSE), modele ilişkin grafikler, model katsayıları başlıkları altında raporlanmaktadır. Ayrıca bulgular, PDF ve HTML uzantılı raporlar olarak kişisel bilgisayarlara kaydedilebilmektedir.

Sonuç: Yazılım, sağlık araştırmalarında, özellikle uç değerlerin varlığı durumunda klasik doğrusal regresyonun kullanılamadığı durumlar için kolaylıkla kantil regresyon analizini uygulayabilmek amacıyla geliştirilmiştir. İlerleyen çalışmalarda, kantil regresyonunun farklı varyantlarının da (Derin Kuantil Regresyon vb.) yazılıma eklenmesi hedeflenmektedir.

Anahtar Kelimeler: Kantil regresyon, web-tabanlı yazılım, shiny

RNA-Dizileme Verilerinde Sağkalım Algoritmalarının Kapsamlı Bir Karşılaştırması

¹Ahu CEPHE, ²Ahmet SEZGIN, ³Necla KOÇHAN, ⁴Gözde ERTÜRK ZARARSIZ, ⁵Vahap ELDEM, ⁶Erdem KARABULUT, ⁷Gökmen ZARARSIZ

¹Erciyes Üniversitesi, Rektörlük, Kurumsal Veri Yönetimi ve Analitiği Koordinatörlüğü, Kayseri, Türkiye

²Abdullah Gül Üniversitesi, Bilgisayar Mühendisliği

³İzmir Ekonomi Üniversitesi, Fen-Edebiyat Fakültesi, Matematik Bölümü

⁴Erciyes Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi ABD, Kayseri, Türkiye

⁵İstanbul Üniversitesi, Fen Fakültesi, Biyoloji Bölümü, Zooloji Anabilim Dalı, İstanbul

⁶Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara

⁷Erciyes Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi ABD, Kayseri, Türkiye

Amaç: Son yıllarda Rna-dizileme gibi yüksek-boyutlu gen ifade verileri kullanılarak çeşitli hastalıklarda sağkalım tahminleme çalışmaları yapılmaktadır (1-4). Belirli bir olay olana kadar geçen süre ve gen ifade profili arasındaki ilişki bulunup daha kesin prognoz tahminleri yapılarak daha iyi tedavi stratejileri geliştirilebilmektedir. Düşük-boyutlu sağkalım verilerinde zaman ve sansür değişkeni dikkate alınarak birçok geleneksel sağkalım yöntemi ile analizler yapılabilirken, RNA-dizileme verilerinde kullanılacak olan sağkalım algoritmalarının bu değişkenlere ek olarak yüksek-boyutluluk, heterojenlik ve yüksek-korelasyonlu genlerle de başa çıkabilmesi gerekmektedir (1). Bu nedenle, bu verilerde sağkalım analizleri yapılırken düzenlenleştirilmiş Cox ve sağkalım verilerine uyarlanmış makine öğrenme yaklaşımları kullanılmaktadır. Bu çalışmada, çeşitli kanser hastalarına ait 20 farklı RNA-dizileme verisinde mevcut sağkalım algoritmalarının performanslarının kapsamlı bir karşılaştırılması yapılmıştır.

Gereç ve Yöntem: Karşılaştırmalar yapılırken cezalı Cox, rasgele sağkalım ormanı, destek vektör makineleri, yapay sınır ağırları yöntemlerini içeren 25 farklı algoritma kullanılmıştır. Sağkalım zamanı ve sansür değişkeni içeren RNA-dizileme veri setleri Kanser Genom Atlas Portalı (TCGA, <https://portal.gdc.cancer.gov>) veri tabanından TCGAbiolinks R paketi kullanılarak indirilmiştir. Sağkalım analizlerinde performans karşılaştırmaları için Harrell'in C-indeksi kullanılmıştır. Tüm analizler R programlama dilinin 4.2.1 versiyonu ile mlr3, mlr3proba, survival, survminer, edgeR, DESeq2 paketleri kullanılarak yapılmıştır.

Bulgular: Çalışma sonuçları incelendiğinde rasgele orman (5-7) algoritmalarının genellikle yüksek performans gösterdiği görülmüştür. Bu algoritmalarından sonra lasso, ridge ve elastic-net regresyonlarının (6-8) RNA-dizileme verilerinin sağkalım analizlerinde yüksek C-indeksi sonuçları verdiği görülmüştür.

Sonuç: Yüksek-boyutlu, heterojen ve yüksek-korelasyonlu genler içeren RNA-dizileme verilerinin sağkalım analizlerinde rasgele ormanlar ile lasso, ridge ve elastic-net regresyonu algoritmaları yüksek performans göstermektedir.

Anahtar Kelimeler: Cox, gen ifade verileri, makine öğrenmesi, RNA-dizileme, sağkalım analizi

Sıralı Bir Yanıt Değişkenini Sınıflamada Sayısal Bir Ölçüm İçin Optimal Kesim Noktalarını Belirleme

¹İlker UNAL, ²Yaşar SERTDEMİR, ³H. Esin ÜNAL

¹Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana, Türkiye

²Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana, Türkiye

³Çukurova Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü, Adana, Türkiye

Amaç: Sınıflama grupların uygun kriterler ile ayrıştırılması olarak tanımlanabilir. Her ne kadar sınıflama denince iki grup akla gelse de ikiden fazla grup için de aynı terminoloji kullanır. Bu çalışmada üç grubun sınıflamasında bir sayısal ölçüm için uygun kesim noktası belirlenme yöntemleri karşılaştırılmıştır.

Gereç ve Yöntem: Literatürde sıralı bir yanıt değişkenin sınıflamasında sayısal ölçümler için kesim noktası belirleme işlemi, iki sınıflı durum için kullanılan Youden endeksi, sol üst köşeye en yakın nokta, uyum olasılığı gibi metotların genelleştirilmesi ile yapılmaktadır. Bu çalışmada aynı amaçla kullanılan IU metodunun üçlü bir yanıt değişkeni için genelleştirilmesi yapılarak kesim noktaları belirlenmiştir. Çalışmada farklı senaryolarda üretilmiş veri ve literatürden alınan gerçek veri seti kullanılmıştır. Veri üretiminde örneklem büyüklüğünün etkisi, sayısal ölçümün dağılımının etkisi, gruplardaki gözlem sayısının dengesi gibi özellikler incelenmiştir.

Bulgular: Yapılan analizler sonucunda IU metodunun genelleştirilmiş halinin uygun kesim noktasını belirlemede diğer yöntemlerle benzer ve kimi durumlarda da daha iyi sonuç verdiği görülmüştür.

Sonuç: IU metodu kullanım kolaylığı ve hesap gerektirmeyen yapısı nedeniyle sınıflama için uygun kesim noktası belirlemede oldukça etkili ve kullanışlı bir yöntemdir.

Anahtar Kelimeler: Sınıflama, sıralı yanıt, IU metodu, Youden indeksi

Efor Kullanımlı ve Üç Parametrelili Lojistik Madde Yanıt Teorisi Modellerinin Bilgisayar Uyarlamalı Test Performanslarının Karşılaştırılması

¹Yusuf Kemal ARSLAN, ²Afra ALKAN, ³Atila Halil ELHAN

¹Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana, Türkiye

²Ankara Yıldırım Beyazıt Üniversitesi Tıp Fakültesi, Biyoistatistik Ana Bilim Dalı, Ankara, Türkiye

³Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara, Türkiye

Amaç: Bu çalışmanın asıl amacı, üç parametrelili lojistik (3PL) madde yanıt teorisi modeli ve Efor kullanımlı (EM) madde yanıt teorisi modelinin bilgisayar uyarlamalı test (BUT) performanslarını karşılaştırmaktır. Primer amaca ek olarak klasik kağıt kalem testleri ve BUT performansları da karşılaştırılacaktır.

Gereç ve Yöntem: Çalışma bir benzetim çalışmasıdır. Kişi sayısı 1000; soru bankasındaki madde sayısı 50, 100 ve 250; BUT durdurma kriteri için standart hata <0.3 ve <0.5 olacak şekilde toplam 6 senaryo için veri setleri üretilmiştir. Her senaryo için 1000 tekrar uygulanmıştır. Benzetim çalışmasında 3PL model ve EM model için düzenlenen kodlar kullanılmıştır. Bilgi fonksiyonları için uyarlanan kodların doğruluğu kontrol edilmiştir. Her bir senaryoda, BUT 3PL ve BUT EM'den elde edilen θ kestirimleri kaydedilmiştir. BUT performansını belirlemek üzere BUT kestirimlerinin yanlılıkları, standart hataları, uygulanan madde sayısı ortalaması ve standart sapması ile soru bankasındaki maddelerin uygulanma frekansları hesaplanmıştır. Ayrıca kestirimlerin gerçek θ değerleri ile uyumları değerlendirilmiştir. Oluşturulan soru bankasındaki madde parametreleri uygun dağılımlardan üretilmiştir. Sonsal beklenti kestirimleri için $[-4, 4]$ aralığında 0.1 artışla 81 adet θ oluşturulmuştur. Yanıt süresi için her bir senaryoda kişilerin %10'u hızlı yanıt davranışı, %90'u çözümsel davranış gösterecek şekilde b'lerle ilişkili olarak üretilmiştir. BUT uygulaması, tüm kişilere orta düzeyli madde sorularak başlatılmıştır, modeller için sonraki madde seçimi en yüksek bilgiyi veren maddeyi seçerek ilerlenmesi şeklinde gerçekleştirilmiş ve süreç durdurma kriteri sağlanana kadar devam ettirilmiştir. Yetenek düzeyi kestirimlerinde en çok olabilirlik (ML) ve sonsal beklenti (EAP) kestirimleri kullanılmıştır.

Bulgular: Benzetim çalışması sonucunda yöntemden bağımsız olarak soru bankasındaki madde sayısı artarken durdurma kriteri olarak belirlenen her iki standart hata düzeyi için de uygulanan ortalama soru sayısı azalmıştır. Durdurma kriteri olarak alınan standart hatanın 0,3'ten 0,5'e çıkması uygulanan soru sayısının artmasına sebep olmuştur. BUT EM'de uygulanan ortalama soru sayılarının BUT 3PL'ye göre daha düşük olduğu görülmüştür. Ortalama standart hata için elde edilen bulgular incelendiğinde madde havuzundaki madde sayısının artmasının her iki yöntemde de ortalama standart hatayı azalttığı sonucuna varıldı. Senaryolardan 50 madde içeren ve standart hata 0,3 olan durumda her iki yöntemde de ortalama standart hatanın minimum değeri beklenen standart hata düzeyine ulaşamadı. Madde sayısındaki artışın gerçek yetenek düzeyi ve BUT yöntemleri ile kestirilen yetenek düzeyleri arasındaki SKK'nın artmasına sebep olduğu görüldü. Benzer şekilde standart hatadaki azalışın SKK'nın artmasına sebep olduğu sonucuna varıldı. Gerçek yetenek düzeyleri ve kestirilen yetenek düzeyleri arasındaki uyum incelendiğinde, standart hatanın 0,3 olduğu durumda, BUT 3PL yönteminde madde sayısı arttıkça uyum sınırları daralmış ve uyum sınırları dışında kalan kişilerin yüzdesi azalmıştır. BUT 3PL'ye benzer şekilde BUT EM yönteminde de durum benzerdi.

Sonuç: BUT EM yöntemi yetenek düzeyini BUT 3PL yöntemine göre önemsiz düzeyde düşük kestirme eğilimindedir. Her iki yöntemde de standart hata düzeyinin düşmesi sorulan soru sayısını arttırmıştır ve gerçek yetenek düzeyi ile daha uyumlu kestirimler yapılmasını sağlamıştır. Klasik kağıt-kalem yöntemleri ile BUT uygulamaları benzer sonuçlar vermiştir. Dijitalleşmenin hızlandığı günümüzde, kişi yetenek düzeyini daha belirgin şekilde belirlemek ve kişilerin sınav stratejilerini çözümlmek adına BUT EM'ler kullanılabilir.

Anahtar Kelimeler: Efor kullanımlı model, üç parametrelili lojistik model, yanıt süresi, bilgisayar uyarlamalı test

dtComb: İki Sürekli Tanı Testinin Birleştirilmesine Yönelik Geliştirilmiş Kapsamlı Bir R Paketi

¹Serra İlayda YERLİTAŞ, ²Serra İlayda YERLİTAŞ, ³Serra Bersan GENGEÇ, ⁴Gözde ERTÜRK ZARARSIZ, KORKMAZ, ⁷Gökmen ZARARSIZ, ⁸Gökmen ZARARSIZ, ⁹Gökmen ZARARSIZ

⁵Gözde ERTÜRK ZARARSIZ, ⁶Selçuk

¹Erciyes Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi ABD, Kayseri, Türkiye
²Erciyes Üniversitesi, İlaç Uygulama ve Araştırma Merkezi (ERFARMA), Kayseri, Türkiye
³Erciyes Üniversitesi, İlaç Uygulama ve Araştırma Merkezi (ERFARMA), Kayseri, Türkiye
⁴Erciyes Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi ABD, Kayseri, Türkiye
⁵Erciyes Üniversitesi, İlaç Uygulama ve Araştırma Merkezi (ERFARMA), Kayseri, Türkiye
⁶Trakya Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Edirne, Türkiye
⁷Erciyes Üniversitesi, Rektörlük, Kurumsal Veri Yönetimi ve Analitiği Koordinatörlüğü, Kayseri, Türkiye
⁸Erciyes Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi ABD, Kayseri, Türkiye
⁹Erciyes Üniversitesi, İlaç Uygulama ve Araştırma Merkezi (ERFARMA), Kayseri, Türkiye

Amaç: Hastalıkların teşhisi ve birbirlerinden ayırt edilmesi için tanı testlerinin geliştirilmesi sağlık alanında önemli bir araştırma alanıdır. Birden fazla tanı testi sonuçlarının birleştirilmesi ile daha yüksek tanı doğruluğu performansı ve güvenilirlik elde edilebilir. Bu çalışmada, dağınık biçimde yer alan çok sayıda kombinasyon yönteminin bir araya getirilmesi, kolay erişim ve kullanım ile araştırmacılara sunulması amacıyla dtComb R paketi ve web yazılımının geliştirilmesi amaçlanmıştır.

Gereç ve Yöntem: Tanı testlerinin performansını artırmaya yönelik birleştirilmesi için literatürde çok sayıda yöntem mevcuttur. dtComb paketi kapsamında bu kombinasyon yöntemleri doğrusal yöntemler, doğrusal olmayan yöntemler, matematiksel operatörler ve makine öğrenmesi algoritmaları olarak dört gruba ayrılmıştır. Bu dört grup içerisinde 143 adet tanı testi birleştirme yöntemine ait kodlamalar, 20 farklı R fonksiyonu içerisinde gerçekleştirilmiştir. Paketin oluşturulma aşamaları devtools R paketi ile gerçekleştirilmiştir. Geliştirilen fonksiyonların her biriminin beklendiği gibi çalıştığını doğrulamakta kullanılan birim testler için testthat R paketinden yararlanılmıştır. Oluşturulan fonksiyonların yazılım testleri 270 adet birim test ile tamamlanmıştır. Ayrıca, R kullanıcı olmayanlar için shiny R paketi kullanılarak dtComb paketinin web yazılımı geliştirilmiştir.

Bulgular: Tanı testlerini birleştirmek için oluşturulan dört ana fonksiyon altın standart durumun ve iki tanı testinin sürekli değerlerini içeren bir veriyi girdi olarak kullanmaktadır. Bu fonksiyonlar yardımıyla kullanıcı seçtiği kombinasyon yöntemini çeşitli standartlaştırma, yeniden örnekleme ve kesim noktası yöntemleriyle birlikte kullanarak tanı testlerini birleştirebilir. Oluşturulan model ile birleştirilmiş tanı testi skorları, tekli ve birleştirilmiş tanı testi istatistikleri, ROC eğrisi ve eğri altında kalan alan istatistikleri, birleştirilmiş tanı testinin tekli tanı testi skorları ile karşılaştırmasına ilişkin test sonuçları, birleştirilmiş ve tekli tanı testi dağılım ve duyarlılık&seçicilik grafikleri raporlanabilmektedir. Ayrıca kurulan modelden alınan parametreler ile yeni veriler için tahminleme yapabilmektedir.

Sonuç: Çalışmanın kapsamında iki sürekli tanı testini birleştirmek amacıyla mevcut kombinasyon yöntemlerinin uygulanabilmesi için bir R paketi ve web yazılımı geliştirilmiştir. Geliştirilen dtComb R paketine izleyen adresten erişilebilir: <http://biosoft.erciyes.edu.tr/app/dtComb>.

Anahtar Kelimeler: Makine öğrenmesi, R programlama dili, ROC analizi, tanı testlerinin birleştirilmesi

RNA-Dizi Verilerinde İntron Tutulumunun Belirlenmesinde Kullanılan Biyoinformatik Algoritmaların Değerlendirilmesi

¹Esmâ Gamze Aksel, ²Vahap ELDEM, ³Gökmen ZARARSIZ

¹Erciyes Üniversitesi, Veteriner Fakültesi, Genetik Anabilim Dalı, Kayseri

²İstanbul Üniversitesi, Fen Fakültesi, Biyoloji Bölümü, Zooloji Anabilim Dalı, İstanbul

³Erciyes Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Kayseri

Amaç: Alternatif ekleme tiplerinden biri olan intron tutulumu (intron retention, IR), bir intron pre-mRNA'ya kopyalandığında ve son mRNA'da kaldığında meydana gelir. İntron tutulumunun güvenilir bir şekilde ölçülmesi zor olduğu için genellikle ihmal edilen bir alternatif ekleme sınıfını oluşturur. Biyoinformatik algoritma araçları ile intron tutulumunun tespitinde karşılaşılabilecek birçok karıştırıcı faktör bulunmaktadır. Dolayısıyla IR'nin belirlenmesi için bu faktörlerin tespiti önemlidir. Bu çalışmada RNA-dizileme verilerinde intron tutulumunun belirlenmesinde kullanılan algoritmalar sunulacaktır. Bunlardan bazıları; Mixture-of-isoforms model (MISO), Keep Me Around (KMA), IRFinder, Vast-tools, MAJIQ-SPEL, Whippet, Junction usage model (JUM), IntERESt, IREAD, IRFinder-S algoritmalarıdır. Ayrıca NCBI veri tabanında bildirilen kanserli ve kontrol meme dokusuna ait 3'er örnek RNA-dizi verilerinin IRFinder-S algoritmasında yapılacak intron tutulumu analizi ile örneklendirilmesi planlanmıştır.

Gereç ve Yöntem: Meme kanseri ve kontrol dokularına ait RNA-dizileme verileri NCBI veri tabanından paired-end, fastqc formatında elde edilmiştir. Analizler için Linux tabanlı sanal sunucuda Python 3.9 programlama dili ile çalışılmıştır. Referans genom olarak insan genomuna ait GRCh38, Hisat2 programı ile indekslenmiştir. Genomik indexleme sonrası IRFinder-S programı kullanılarak RNA-dizileme verileri referans genoma hizalanarak tutulan intronların sayısal verileri elde edilmiştir. Kontrol ve kanser dokularına ait intron tutulumuna ilişkin elde edilen sayısal verilerin istatistiksel anlamlılığı DESeq2 paketi ile belirlenmiştir.

Bulgular: Elde edilen bulgulara göre meme kanserine ait belirli genlerde intron tutulumu oranının istatistiksel olarak anlamlı bulunduğu belirlenmiştir.

Sonuç: Özellikle kanser türlerinin oluşum ve gelişim safhalarında intron tutulumu rolünün belirlenmesinde bahsi geçen algoritmaların tanınması, kullanımı ve yaygınlaştırılması kanser oluşum yolağının belirlenmesinde önem arz etmektedir. Teşekkür: Bu çalışma Erciyes Üniversitesi Bilimsel Araştırmalar Koordinasyon Birimi tarafından TYL-2022-12231 proje numarası ile desteklenmiştir.

Anahtar Kelimeler: Gen ekspresyonu, intron tutulumu, RNA-dizileme

Biyoistatistik Dersine Yönelik İstatistik Öz Yeterlilik İnanç Ölçeği, İstatistik Kaygı Ölçeği ve İstatistik Tutum Ölçeği Arasındaki Aracılık İlişkisinin Araştırılması: Bir Yapısal Eşitlik Modellemesi Çalışması (Path Analizi)

¹Kamber KAŞALI, ²Senem GÖNENÇ

¹Atatürk Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Erzurum

²Atatürk Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Erzurum

Amaç: Araştırmacıların istatistik konularına karşı korkuları akademik hayatlarını doğrudan etkilemektedir. Çünkü akademik çalışmalar en başından en sonuna kadar istatistiksel bir metodoloji ile hazırlanmalıdır. Bu sebeple, istatistik konularına karşı tutum, kaygı ve yeterlilikleri belirlemek için ölçekler geliştirilmiştir. Çalışmanın amacı; istatistik öz yeterlilik inanç ölçeği, istatistik kaygı ölçeği ve istatistik tutum ölçeği arasındaki aracılık ilişkisinin belirlenmesidir.

Gereç ve Yöntem: Çalışmamızda; istatistik öz yeterlilik inanç ölçeği, istatistik kaygı ölçeği (Alt boyutlar; Endişe, Kaçınma ve Duygusallık) ve istatistik tutum ölçeği (Alt boyutlar; İlgi duyma, Kaygı ve Motivasyon) kullanılarak öğrencilerin biyoistatistik dersine karşı durumları belirlenmiştir. Veriler dersin öncesinde (T0), ortasında (T1) ve sonunda (T2) olmak üzere 3 farklı zamanda toplanmıştır. Çalışmaya biyoistatistik dersini alan lisans, yüksek lisans ve doktora öğrencilerinden oluşan toplam 74 kişi katılmıştır. Nicel değişkenler arasındaki ilişki için korelasyon testleri uygulanmış olup gruplar arası kıyaslamalarda 2 grup için Independent Samples t testi, Mann Whitney u testi; 3 grup için ANOVA ve Kruskal Wallis testi kullanılmıştır. Oluşturulan modelde yer alan dolaylı etkilerin anlamlılık düzeyi yapısal eşitlik modellemesi Bootstrapping yöntemi ile test edilmiş ve modele ilişkin yol katsayıları (β) hesaplanmıştır. Analizlerde, IBM SPSS 20 ve JAMOVI 2.2.2 programları kullanılmıştır.

Bulgular: Araştırmamızda ders öncesinde, ders ortasında ve ders sonunda ölçek alt boyutları arasındaki aracılık etkisi incelenmiştir. Path analizi sonuçlarına bakıldığında zaman ders öncesinde istatistik kaygı ile istatistik öz yeterlilik inanç durumu arasında istatistiksel endişenin aracılık etkisi olduğu belirlenmiştir ($\beta=0,20$, $p=0,019$). Ders ortasında istatistik kaygı ile istatistik öz yeterlilik inanç durumu arasında istatistiksel duygusallığın kısmen aracılık etkisi olduğu belirlenmiştir ($\beta=0,19$, $p=0,057$). Ders sonunda istatistik kaygı ile istatistik öz yeterlilik inanç durumu arasında istatistiksel duygusallığın aracılık etkisi olduğu belirlenmiştir ($\beta=0,23$, $p=0,049$).

Sonuç: Çalışmamızda sonuç olarak Biyoistatistik dersini alan öğrencilerin istatistik öz yeterlilik ve inançları üzerinde kaygının dolaylı bir etkisinin olduğu ve bu etkide endişe ve duygusallığın aracı bir etkisi olduğu belirlenmiştir.

Anahtar Kelimeler: Yapısal Eşitlik Modellemesi, Aracılık Etkisi, Biyoistatistik

Bulanık Kümeleme Yöntemi ile Biyoistatistik Dersi Alan Öğrencilerin Değerlendirilmesi

¹Senem GÖNENCİ, ²Kamber KAŞALI, ³Özge PASİN

¹Atatürk Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Erzurum

²Atatürk Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Erzurum

³Bezmîâlem Vakıf Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, İstanbul

Amaç: İstatistik dersi akademik hayatımızdaki başarıyı doğrudan etkilemektedir. Ancak sağlık alanında çalışma yürüten araştırmacılar genelde istatistik konularına karşı olumsuz bir tutum, kaygı ve yetersizlik içerisine girebilmektedir. Bu gibi hususların belirlenmesi ve ortaya çıkarılması akademik başarı açısından oldukça önemli bir konudur. Bu çalışmada biyoistatistik dersi alan lisans, yüksek lisans ve doktora öğrencilerinden oluşan araştırma grubu bulanık kümeleme analizi ile incelenmiştir.

Gereç ve Yöntem: Araştırma grubuna ilişkin eğitim gördükleri bölümler (beslenme ve diyetetik, hemşirelik, diğer) ve daha önce istatistik dersi alma durumlarına ait bilgilere ulaşılarak öğrencilerin biyoistatistik dersine karşı durumları; istatistik öz yeterlilik inanç ölçeği, istatistik kaygı ölçeği ve istatistik tutum ölçeği kullanılarak değerlendirilmeye çalışılmıştır. Veriler dersin öncesinde ve sonrasında olmak üzere 2 farklı zamanda toplanmıştır. Söz konusu bu iki farklı zaman grubunda elde edilen veriler, bulanık kümeleme yöntemleri içerisinde en çok tercih edilen yöntem olan Bulanık k-Ortalamalar algoritması (FkM) kullanılarak analiz edilmiştir. İstatistiksel olarak bulanık kümeleme sonuçları R programında 'cluster', 'fclust' ve 'factoextra' gibi paketler vasıtasıyla elde edilmiş olup her bir farklı zaman grubunda ortaya çıkan kümeleme yapısı, öğrencilerin okuduğu bölümler ve daha önce istatistik dersi alıp almama durumlarına göre SPSS programında karşılaştırılmıştır.

Bulgular: Araştırma sonucunda bulanık modelden elde edilen kümeleme sonuçları ile çalışma öncesindeki gruplar tahmin edilmeye çalışılmıştır. Ders öncesi ve sonrası elde edilen kümeler ile bölümler arasında istatistiksel olarak bir uyum tespit edilemezken daha önce dersi alma durumunun oluşan kümeleme yapısı ile uyumlu olduğu görülmüştür (Bölüm için KappaP1=0.122; KappaP2=0.792; Daha önce istatistik dersi alma durumu için McNemarP1=0.036; McNemarP2=0.002). Öğrencilerin okuduğu bölümler açısından doğru sınıflandırma yüzdelerine bakıldığında ders öncesi grupta bulanık yöntemden elde edilen 3 küme; bölümlerdeki bireyleri sırasıyla %47, %61 ve %20 oranında doğru tespit ederken ders sonrası grupta bu oranlar; %30, %43 ve %0 olarak bulunmuştur. Daha önce istatistik dersi alma durumuna göre ders öncesi grup için bulanık kümeleme sonucu oluşan 2 küme; bireyleri sırasıyla %29 ve %43 şeklinde, ders sonrası ise %56 ve %46 gibi genel olarak artan doğru sınıflandırma oranları ile tespit etmiştir.

Sonuç: Bulgular kısmı incelendiğinde, bölüm bakımından biyoistatistik dersine başlamadan önce ölçeklerdeki sorulara verilen cevaplar arasında ders sonrasına göre daha belirgin bir fark olduğu belirlenmiştir. Biyoistatistik dersinin bitimine doğru bölümlerdeki öğrencilerin vermiş oldukları cevaplar birbirine benzemeye başladığından bulanık kümeleme sonuçlarına ait doğru sınıflandırma oranları da giderek azalmıştır. İstatistik dersini daha önce alıp almama durumunda ise ders öncesinde ölçeklerdeki sorulara verilen cevapların ders sonrasına göre daha fazla benzer oldukları (istatistik dersi almış ama konuları unutmuş) belirlenmiş ve dolayısıyla bulanık kümeleme sonuçları daha düşük doğruluk oranlarında sınıfları tahmin etmiştir. Biyoistatistik dersi devam ederken daha önce aldığı istatistik dersi konularını hatırlayıp üzerine yeni bilgiler eklediklerinden ölçek sorularına verilen cevaplar arasında ders sonuna doğru daha belirgin bir fark oluşmuştur. Bununla birlikte bulanık kümeleme sonuçları da daha yüksek doğruluk oranı gösteren tahminlerde bulunmuştur. Çalışma sonucunda, bulanık kümelemenin istatistiksel olarak anlamlı ve yorumlanabilir sonuçlar ortaya koyduğu bununla birlikte bu tarz çalışmalarda kullanılabilir bir yöntem olarak tercih edilmesinin uygun olacağı düşünülmektedir.

Anahtar Kelimeler: Kümeleme yöntemleri, bulanık kümeleme, biyoistatistik, istatistiksel analiz, sınıflandırma

Erken Evre Meme Kanseri Hastalarına Ait Hazard Faktörlerinin Uygulanan Çeşitli Model Sonuçlarının Karşılaştırılması

¹Nural BEKİROĞLU, ²Ayşe ULGEN, ³Emrah Gökay ÖZGÜR, ⁴Sinan UZUN, ⁵Wentian LI

¹Marmara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, İstanbul

²Girne Amerikan Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Karmi, KKTC

³Marmara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, İstanbul

⁴Marmara Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı, İstanbul

⁵Robert S Boas Center for Genomics and Human Genetics, The Feinstein Institute for Medical Research, Northwell Health, Manhassat, New York, ABD

Amaç: Cox Orantılı Hazard (COH) sağkalım analizinde kullanılan çok değişkenli regresyon modelidir. Orantılı hazard (OH) varsayımını sağlamak her zaman mümkün olmamaktadır. Bu çalışmada, erken evre meme kanseri hastaların klinik bilgilerine göre, tedavi sonrası sağkalım olasılıklarına ilişkin R yazılımlı COH regresyon, makine öğrenmesi teknikleri, parametrik Hızlandırılmış Başarısızlık (ölüm) Zamanı (HBZ) modeli ile hazard faktörler açısından tahminlenmesi ve karşılaştırılması amaçlanmıştır.

Gereç ve Yöntem: Çalışmaya Marmara Üniversitesi Tıbbi Onkoloji polikliniğinde erken evre meme kanseri tanısı konulan 697 hasta dâhil edilmiştir. Erken evre meme kanseri hastaların 5 ve 10 yıllık sağkalım olasılıkları Kaplan-Meier yöntemi ile hesaplanmıştır. Sağkalıma etkili olan anlamlı hazard faktörlerinin belirlenmesinde; tek değişkenli analizlerde anlamlı hazard faktörlerin tespiti için Log-Rank testi, çok değişkenli analizlerde ise COH regresyon çözümlenmesi yapılmıştır. Aynı veri setine makine öğrenmesi algortimaları olan Rastgele Orman, Xgboost ve Gbm modelleri uygulanarak, 5 ve 10 yıllık sağkalım olasılıkları hesaplanmış olup önem derecesine göre sağkalıma etkili olan hazard faktörler sıralanmıştır. Son aşamada da; genel sağkalıma ait en uygun hazard fonksiyonu dağılımının Log-lojistik dağılıma uygun olduğu görülmüş dolayısıyla yine aynı veri setine Log-lojistik HBZ modeli uygulanmıştır. Modellerin anlamlı çıkan ilk 5 hazard faktörü tespit edilerek en uygun model saptanmaya çalışılmıştır.

Bulgular: Erken evre meme kanseri tanısı konulan 697 hastanın makine öğrenmesi algoritmalarıyla ile Kaplan-Meier sağkalım yöntemleriyle hesaplanan 5 ve 10 yıllık sağkalım olasılıkları hemen hemen aynı değerlere sahiptir. COH modelinde hazard faktörlerinin etkisi HO (hazard oranı) ile yorumlanırken, HBZ modellerinde ZO (zaman oranı) şeklinde yorumlanmaktadır. ZO, deneklerin sağkalım eğrisinde katettiği mesafenin oranlarının bir karşılaştırması olduğundan HO gibi yorumlanmasa da; hazard faktörleri açısından yapılan karşılaştırmada, meme kanseri hastaları için klinik açıdan önemli bir risk göstergesi olan Her-2 faktörü makine öğrenme algoritmalarıyla oluşturulan sağkalım modellerinin ikisinde (Gbm, Xgboost) önemli ilk 5 hazard faktörün içerisinde sıralanmış olmasına rağmen COH Regresyon ve HBZ modelinde anlamlı çıkmamıştır.

Sonuç: Sağlık alanına ait, özellikle takibe dayalı çalışmalarda eksik gözlem veya dengesiz dağılım sergileyen veri setlerinde seçenek yöntemlerin tercih edilmesi analiz sonuçlarını klinik açıdan daha güvenilir yapacaktır.

Anahtar Kelimeler: Erken Evre Meme Kanseri, Cox Orantısız Hazard Regresyon, Parametrik Hızlandırılmış Başarısızlık Zamanı Modeli, Makine Öğrenmesi Teknikleri

Makine Öğrenme Yöntemlerinin Başarılarını Değerlendirirken Hangi Performans Ölçütleri Kullanılmalıdır.

¹Yaşar SERTDEMİR, ²İlker ÜNAL, ³Hülya BİNOKAY

¹Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana

²Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana

³Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana

Amaç: Bu çalışmadaki amaç, makine öğrenme yöntemlerinin değerlendirmesinde kullanılan performans ölçütlerinden hangilerinin benzer ve hangilerinin farklı bilgi içerdiğini belirleyerek performans değerlendirmesinde kullanılacak ölçüt gruplarını belirlemektir.

Gereç ve Yöntem: Literatürde makine öğrenme yöntemlerinin karşılaştırılmasında kullanılan ve özellik olarak birbirinden 41 farklı gerçek veri setinin her biri %70'i eğitim ve %30'u test seti olarak ayrıldı. Bu ayırma işlemi rasgele olarak 300 kere tekrarlandı. Her bir tekrarda lojistik regresyon analizi uygulanarak oluşturulan modellerin test veri setlerinde PÖ'leri elde edildi. Her bir veri seti için, ölçütler üzerinde faktör analizi yapılarak hangi PÖ'lerinin birliktelik gösterdiği kayıt edildi.

Bulgular: Performans ölçütleri veri setlerindeki bağımlı değişkenin prevalansına göre <0.25, 0.25-0.39 ve 0.40-0.5 setlerine ayrılarak değerlendirildiğinde Performans ölçütleri grup sayısı ve bu gruplarda yer alan performans ölçütleri farklılık göstermektedir. Bu PÖ arasında prevalansa göre farklı PÖ gruplarında yer alan ölçütleri; RMSE, MCC, Cohen Kappa ve YI olduğu gözlenmiştir.

Sonuç: Makine öğrenme yöntemlerinin performanslarını karşılaştırırken ACC, Adj_F ve (AUCROC veya RMSE) den biri kullanıldığında üç PÖ den ikisinin aynı grupta yer alması nedeniyle ilk iki PÖ başarısına daha fazla ağırlık verilmesi neden ile yanlış yöntem başarılı bulunabilmektedir. Önerimiz bağımlı değişkenin prevalasına göre belirlenen gruplarda oluşan PÖ gruplarından birer PÖ seçilerek yöntemlerin karşılaştırılmasıdır.

Anahtar Kelimeler: Performans ölçütleri, makine öğrenme

Propensity Skora Dayalı Eşleştirme Analizi İçin Yeni Bir R Shiny Web Uygulaması (A New R Shiny Web Application for Propensity Score Matching Analysis)

¹Emre DEMİR, ²Gülçin AYDOĞDU, ³Muhammed Ali YILMAZ, ⁴Yasemin YAVUZ

¹Hitit Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Çorum

²Hitit Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Çorum

³Sağlık Bakanlığı, Çağrı Hizmetleri Dairesi Başkanlığı, Ankara

⁴Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Amaç: Birçok tıp alanında randomizasyonun gerçekleştirilemediği gözlemsel çalışmalarda bir maruziyetin veya tedavinin gerçek etkisinin ortaya konabilmesi amacıyla vaka ve kontrol gruplarının ortak değişkenler (covariate) açısından benzer dağılımı istenilmektedir. Gözlemsel araştırmalarda, vaka ve kontrol grupları arasında birçok karıştırıcı (confounder) değişken açısından farklılıklar bulunabilir. Gözlemsel araştırmalarda karıştırıcı değişkenleri kontrol altında tutarak yansız tahminler elde etmek için literatürde yaygın olarak Lojistik Regresyon analizi ile hesaplanan Propensity skor (PS) tahminlerine dayalı eşleştirme analizi kullanılmaktadır. Eşleştirme analizinin amacı vaka grubundaki bireylerin sahip olduğu ortak değişken özelliklerine benzer olan bireyleri kontrol grubu içerisinde seçmektir. Bu çalışmanın amacı PS'ye dayalı eşleştirme analizi yapan ve eşleştirme sonrası birçok farklı denge değerlendirme yöntemi içeren web tabanlı yeni bir interaktif R Shiny uygulaması geliştirmektir.

Gereç ve Yöntem: Web tabanlı uygulama R Shiny (R Core Team, 2019) kullanılarak geliştirilmiştir. Eşleştirme analizi ve PS tahmini için "matchIt" kütüphanesi kullanıldı. Dengeyi değerlendirmek ve grafikler oluşturmak için "matchIt" kütüphanesine ek olarak "cobalt" kütüphanesi kullanıldı. Ortak değişkenlere/faktörlere göre eşleştirme dengesinin değerlendirilmesi için, Standardize Ortalama Farklar (SOF), Varyans Oranları (VO), Kolmogorov Smirnov (KS) test istatistikleri ve çeşitli grafiksel yaklaşımlar kullanıldı. İnteraktif uygulamanın kullanımı bir benzetim veri seti üzerinde açıklanmıştır. Araştırmacılar "<https://psmatching.shinyapps.io/propensityscore>" web adresinden uygulamaya erişebilirler.

Bulgular: Üretilen benzetim veri seti ile eşleştirme analizi gerçekleştirildi. Propensity skorlar lojistik regresyon analizi ile hesaplandı. Eşleştirme analizi için en yakın komşuluk algoritması, optimal eşleştirme ve tam eşleştirme yöntemleri, 1:1 ve 1:2 oranlarında sırasıyla 0.1, 0.15, 0.2, 0.25 ve 0.3 caliper değerleri kullanılarak hesaplandı. Farklı istatistiksel ve grafiksel yaklaşımlar ile gerçekleştirilen denge değerlendirilmesi sonucunda en iyi dengeye sahip eşleştirilmiş veri seti elde edildi.

Sonuç: Randomizasyonun yapılamadığı gözlemsel araştırmalarda PS'ye göre vaka ve kontrol gruplarının eşleştirilmesi karıştırıcı faktörlerin etkisini kontrol ederek maruziyetin saf etkisini belirlemek amacıyla araştırmalarda dikkate alınması gereken ve son dekatta önemi artan bir yöntemdir. Geliştirmiş olduğumuz interaktif uygulamanın küresel olarak araştırmacılar tarafından gözlemsel çalışmalarda vaka grubuna en uygun kontrol grubunu seçmek amacıyla rahatlıkla kullanılabilecek bir uygulama olacağını düşünmekteyiz.

Anahtar Kelimeler: Propensity skor, eşleştirme, R Shiny, gözlemsel çalışmalar, randomizasyon



Şekil-1. Uygulamada denge değerlendirmesini gösteren bulgular

Kantil Regresyon Yöntemi İle Oluşturulan Pediatrik Büyüme Eğrilerinin Model Uyumunun Worm Grafiği İle Değerlendirilmesi

¹Eda ÇAKMAK, ²Ergun KARAAĞAOĞLU, ³Serhat KILIÇ, ⁴Pınar ÖZDEMİR

¹Başkent Üniversitesi Sağlık Bilimleri Fakültesi, Ankara

²Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

³Başkent Üniversitesi Tıp Fakültesi, Çocuk Sağlığı ve Hastalıkları Anabilim Dalı, Ankara

⁴Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Amaç: Bir çocuğun büyüme ve gelişiminin incelenmesinde, hastalık durumunun erken tanınmasında en iyi araçlardan biri pediatrik büyüme eğrileridir. Büyüme eğrilerinin oluşturulmasında kullanılan yöntemlerden biri de kantil regresyon yöntemidir. Pediatrik büyüme eğrileri oluşturulurken model yeterliğinin incelenmesinde worm grafiğinden yararlanılmaktadır. Bu çalışmanın amacı, kantil regresyon yöntemi aracılığıyla oluşturulan büyüme eğrilerinin model yeterliğinin worm grafiği ile incelenmesidir.

Gereç ve Yöntem: Kantil regresyon yöntemi veride herhangi bir dağılım varsayımı gerektirmediğinden dolayı worm grafiğini elde ederken veride bu şart aranmamaktadır. Kantil regresyon yöntemi ile oluşturulan büyüme eğrilerinin veriye ne derece uyduğunu göstermek için kullanılan worm grafiği yararlı bir görsel araçtır. Worm grafiği her bir yaş grubu için uydurulan (fitting) büyüme eğrilerinin model uyumunun değerlendirilmesinde kullanılmaktadır. Bu çalışmada Başkent Üniversitesi Ankara Hastanesi'nde 0-3 yaş aralığındaki 3320 çocuğun boy uzunluğu ve vücut ağırlığı ölçümleri kullanılarak kantil regresyon yöntemine göre büyüme eğrileri oluşturulmuş ve büyüme eğrilerinin model yeterliğinin incelenmesinde worm grafiği kullanılmıştır. Çalışmada R 4.1.2 versiyonu kullanılarak kantil regresyon yöntemi ile büyüme eğrilerinin elde edilmesinde "quantregGrowth" paketi ve kantil regresyon yöntemine bağlı olarak worm grafiklerinin oluşturulmasında "AGD" ve "quantreg" paketleri kullanılmıştır.

Bulgular: Gerçek veri seti kullanılarak kantil regresyon yöntemi ile cinsiyete göre boy uzunluğu ve vücut ağırlığı ölçümlerine ilişkin uydurulan büyüme eğrilerinin model yeterliği worm grafiği ile değerlendirilmiştir. Kız ve erkek çocuklarının 0-3 yaş aralığındaki boy uzunluğu ve vücut ağırlığı verileri için uydurulan büyüme eğrilerinin model uyumunun yeterli olduğu söylenebilmektedir. Yaş grupları için ayrı ayrı model uyumu incelendiğinde ise bazı yaş gruplarında hafif sapmalar mevcut ancak genel olarak wormlar %95 güven aralığında yer almaktadır.

Sonuç: Özellikle dağılım varsayımı gerektirmeyen kantil regresyon yöntemi ile uydurulan büyüme eğrilerinin model yeterliğinin incelenmesinde worm grafikleri kolaylık sağlayan görsel bir araçtır. Worm grafiği belirli bir yaş aralığında tüm gruba ilişkin model uyumunu değerlendirmekle birlikte, bu yaş aralığının her bir alt grupları için de model uyumunu değerlendirmek amacıyla kullanılmaktadır.

Anahtar Kelimeler: Kantil regresyon, worm grafiği, büyüme eğrileri

gw-Exp-Tool: A Snakemake Workflow for Genome-wide Characterization and Expression Profiling of Transcription Factors or Gene Families

Vahap ELDEM

İstanbul Üniversitesi Fen Fakültesi Biyoloji Bölümü

Amaç: Genome-wide characterization and gene expression profiling of any protein-coding genes of interest, particularly for transcription factor (TF) genes, attract great interest among biologist and medical parasitologist because these analyses serve as a quick solution for gene characterization at the genome-wide level, expression profiling and gene function estimation. Therefore, determining the TFs might particularly hold promise for understanding molecular and genetic determinants of pathogenesis.

Gereç ve Yöntem: We present a Snakemake workflow called “gw-Exp-Tool” built primarily for biologists to run a wide variety of useful genome-wide characterization, expression analyses and statistical test to determine whether or not TFs are expressed correctly in terms of biological meaning. Analyses consisted of several steps: (i) homology search of genes of interest against well-annotated proteomes of closely-related or model species, (ii) searching for presence of conserved domain of TFs, (iii) multiple sequence alignment and a neighbour joining (NJ) phylogenetic tree construction, (iv) motif annotation, chromosome distribution and collinearity analysis, (v) transcript abundance estimation and differential gene expression analysis, (vi) Markov-chain based rank aggregation method called Weighted MC test for most biologically relevant TFs.

Bulgular: We run the tools on four genomes including serious human pathogens (*Anopheles gambiae*, *Ixodes scapularis*, *Leishmania major*) and ecologically important insects (*Bombus terrestris*) to determine TFs, expression profiles and statistical evaluations. We successfully extracted nearly all TFs (>94%) in four genomes and further identified all TF motifs. The tools (DESeq2, edgeR, and limma) were used for the differential gene expression of TFs at genome-wide level and edgeR and limma identified the largest number of significantly differentially expressed TFs ($p < 0.01$). As for biological relevance, the final list of TFs by weighted rank aggregation showed that the TFs stayed on top of the list assessed as mostly compatible with DESeq2 results.

Sonuç: In this study, we presented a Snakemake workflow called “gw-Exp-Tool” built primarily for biologists to run a wide variety of useful genome-wide characterization, expression and statistical analysis of TFs in a manner easy enough to be employed by users without significant computational training. By providing only three inputs including genome and annotation (proteins), conserved domain IDs and transcriptomic data (or SRR accession numbers), one can easily implement and run.

Anahtar Kelimeler: Transcription factors, genome-wide analysis, snakemake workflow, rank

Performance Analysis of Simulated Data Sets by Machine Learning Algorithms

¹Senem KOC, ²Leman TOMAK, ³Erdem KARABULUT

¹Nişantaşı Üniversitesi Tıbbi İstatistik Anabilim Dalı, İstanbul

²Ondokuz Mayıs Üniversitesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Samsun

³Hacettepe Üniversitesi Biyoistatistik Anabilim Dalı, Ankara

Amaç: Machine Learning provides new tools for interpreting the high-dimensional and complex data sets that clinicians are confronted with. The aim of this study is comparing different sample size of simulated data sets by using machine learning algorithms for the performance analysis.

Gereç ve Yöntem: The simulated data were created using an original data set. This data set contain genetic information with a sample size of 174 which consist of 10 factors of periodontitis patients. Correlation matrices were created according to the periodontitis data set with the help of mean and variance values. Within these results, two data sets of 500 and 1000 sizes were created. Decision trees, K Nearest Neighborhood, Naive Bayes, Support Vector Machines, and Random Forest algorithm were compared as classification models. In addition, the 10-fold cross-validation method was used to save classifiers from side effects such as overlearning during the training phase. Accuracy, sensitivity, and specificity criteria were used to evaluate the classification performance of the models. The analyses were performed using R software.

Bulgular: As a result of the analysis, the accuracy was obtained as 70% with decision trees and nearest k neighborhood for the periodontitis data set, 71% accuracy for support vector machines, 68% accuracy for random forest and 72% accuracy with naive bayes algorithm. For the first simulated data set, the accuracy was obtained as 68% with decision trees, 70% accuracy with k nearest neighborhood and naive bayes, 72% accuracy with random forest and 73% accuracy with support vector machines. For the second simulated data set, the accuracy for naive bayes with decision trees was obtained as 75%, the accuracy for the k nearest neighborhood was 73%, the accuracy for random forest was 76%, and the accuracy for support vector machines was 78% as shown in Table 1.

Sonuç: This study contained different sample size of simulated data set. The results show that the second simulated data show the best performance by support vector machines algorithm. Which shows that by enlarging the sample sizes the performance can be improved by using machine learning algorithms.

Anahtar Kelimeler: Simulated data, machine learning, support vector machines, random forest, naive bayes.

Makine Öğrenmesinde Sınıf Dengesizliği Sorunu için Bazı Veri Ön İşleme Yöntemlerinin Karşılaştırılması: Bir Uygulama

¹Hülya ÖZEN, ²Doğukan ÖZEN

¹Sağlık Bilimleri Üniversitesi Gülhane Tıp Fakültesi, Tıp Bilişimi Anabilim Dalı, Ankara

²Ankara Üniversitesi Veteriner Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Amaç: Dengesiz veriler tipik olarak sınıfların eşit olarak temsil edilmediği sınıflandırma problemleriyle ilgili bir sorunu ifade eder. Bu çalışmada; operasyon sonrası hasta dataları içeren dengesiz bir veri seti üzerinde, sınıflandırmada kullanılan veri ön işleme yöntemleri ve sonrasında uygulanan sınıflandırma algoritmalarının performanslarının karşılaştırılması amaçlanmaktadır.

Gereç ve Yöntem: Bu çalışmada postoperatif iyileşme alanındaki hastaların bir sonraki aşamada hastane içerisinde bir birime sevk edilmesi ya da taburcu olmaları durumunu tahmin edecek uygun bir sınıflandırma algoritması belirlenmeye çalışılmıştır. Bir birime sevk edilen hastalar çoğunluk, taburcu olan hastalar ise azınlık sınıfını oluşturmaktadır. Dengesiz veri setleri kullanılarak doğrudan oluşturulan sınıflandırma algoritmaları, çoğunluk sınıfı yönünde yanlı tahminlerde bulunur. Bu nedenle sınıf dengesizliğini düzenlemeye yönelik veri ön işleme yöntemleri geliştirilmiştir. Bu çalışmada aşağı örnekleme (rasgele, tek taraflı seçim, Tomek links), yukarı örnekleme (rasgele, ADASYN, SMOTE) ve hibrit yöntemler kullanılmıştır. Veri ön işleme yöntemleri, eğitim veri setlerine uygulandıktan sonra, oluşan yeni veri setlerinin sınıflandırılması için C5.0, Random Forests ve Adaboost algoritmaları kullanılmıştır. Kullanılan algoritmaların performansları kesinlik, duyarlılık, F1 skoru metrikleri ile kıyaslanmıştır.

Bulgular: Tek taraflı seçim veri ön işleme yönteminin kullanılmasının ardından Random Forests algoritmasının uygulanması ile en yüksek tahmin sonuçları elde edilmiştir. Azınlık sınıfına ait kesinlik 0.667, duyarlılık 0.571; çoğunluk sınıfına ait kesinlik 0.842, duyarlılık ise 0.889 bulunmuştur. Modele ait F1 skoru 0.615, dengeli doğruluk (balanced accuracy) değeri ise 0.730 olarak elde edilmiştir.

Sonuç: Bu çalışma; dengesiz veri setleri için belirlenecek uygun sınıflandırma algoritmasının tahmin geçerliliğinin veri ön işleme yöntemleri ile arttırılabileceğini ortaya koymaktadır. Veri setine uygun ön işleme yönteminin seçimi doğru sonuca ulaşmak için oldukça önemlidir.

Anahtar Kelimeler: Dengesiz veri seti, sınıflandırma, veri ön işleme, postoperatif veri

Twowaytests: İki Yönlü Bağımsız Grup Tasarımı İçin Bir R Paketi

¹Muhammed Ali YILMAZ, ²Osman DAĞ, ³Samaradasa WEERAHANDI, ⁴Malwane M.A. ANANDA

¹Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara

²Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara

³Department of Mathematical Statistics, University of Nevada, Las Vegas, NV, USA

⁴Department of Mathematical Statistics, University of Nevada, Las Vegas, NV, USA

Amaç: Bu çalışmada ise iki yönlü bağımsız grup tasarımında kullanılabilecek tüm yöntemler ve analizler tek bir R paketi altında toplanarak, araştırmacılara kolaylık sağlamak amaçlanmıştır.

Gereç ve Yöntem: İki faktörün hedef değişken üzerindeki etkisi araştırılırken en çok kullanılan yöntemlerden birisi iki yönlü varyans analizi (ANOVA)'dır. ANOVA yönteminin uygulanabilmesi için normallik ve varyans homojenliği gibi iki temel varsayımın sağlanması gerekir. Bu varsayımlar sağlanmadığı durumda, iki yönlü varyans analizine alternatif yöntemler geliştirilmiştir. Geliştirilen bu yöntemler gpQ tabanlı değişen varyanslı iki yönlü ANOVA, PB tabanlı değişen varyanslı iki yönlü ANOVA, ortanca için iki yönlü ANOVA, kırılmış ortalama için iki yönlü ANOVA ve M-kestiricileri için iki yönlü ANOVA yöntemleridir. Bahsi geçen yedi farklı test yönteminin fonksiyonları "twowaytests" R paketi içerisinde yayınlanmıştır.

Bulgular: "twowaytests" paketi içerisinde bulunan her analiz yöntemi için oluşturulan fonksiyonun çıktısında faktörlere ait test istatistiklerine ve bu test istatistiklerine ait p değerlerine yer verilmektedir. Araştırmacılar alfa değerini kendileri belirleyebilme ve analiz sonuçlarının yazılı çıktısını elde etme imkanına sahiptirler. Kütüphane içerisinde yer alan çeşitli fonksiyonlarla araştırmacılar tanımlayıcı istatistiklerle ve çeşitli grafik türleriyle veri üzerinde ilk incelemelerini yapabilirler.

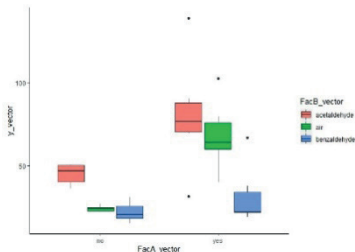
Sonuç: Bu çalışmada iki yönlü ANOVA yapmak isteyen araştırmacıların kullanması amacıyla geliştirdiğimiz "twowaytests" R paketi hazırlanmıştır. İleride çalışmalarda, bu fonksiyonlara ek olarak post-hoc testlerine de yer verilmesi öngörülmektedir.

Anahtar Kelimeler: İki yönlü ANOVA, ortanca, kırılmış ortalama

```
R> library(twowaytests)
R> data("alveolar")
R> gpTwoWay(count ~ ovalbumin*treatment, data = alveolar, method = "gPB")

Generalized p-value by PB method (alpha = 0.05)
=====
      Factor      P-value      Result
ovalbumin    0.0044      Reject
treatment    0.0002      Reject
ovalbumin:treatment 0.0404      Reject
=====
```

Analiz-1. Kütüphane içerisinde yer alan örnek bir fonksiyon ile yapılan bir analizin çıktısına yer verilmiştir.



Grafik-1. Paket sayesinde örnek bir veri seti kullanılarak elde edilen bir grafiğe yer verilmiştir.

Referans Aralığı Belirlemede Kullanılan Tek Değişkenli ve Çok Değişkenli Yöntemlerin Karşılaştırılması

¹Esra Kutsal MERGEN, ²Sevilay KARAHAN

¹Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara

²Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara

Amaç: Çok değişkenli referans alan yöntemi, klinik laboratuvar literatüründe tek değişkenli referans aralık yöntemine bir alternatiftir. Çok değişkenli referans alanı elde etme ve analiz etmedeki zorluklar nedeniyle bu çalışma, klinikte uygulanması kolay ve yanlış pozitif oranını azaltan çok değişkenli referans aralık yöntemleri önermektedir.

Gereç ve Yöntem: Bu çalışmada; sağlık alanında anemi teşhisi için kullanılan Serum Ferritin ve Transferrin Satürasyon yüzdesi laboratuvar değerlerinin tek değişkenli referans aralık yöntemi ile önerilen çok değişkenli referans aralık yöntemlerinin kapsayıcılığını karşılaştırmak için R Studio programı kullanılarak farklı korelasyon ve örneklem genişliği senaryoları için benzetim çalışması yapılmıştır. Anemisi olmayan, 10-20 yaş grubunda bulunan kadın ve erkekler iki ayrı alt grup olarak seçilmiştir. Bu iki laboratuvar değerinin kadın ve erkeklere ilişkin ortalama ve standart sapma değerleri kullanılarak iki değişkenli normal dağılıma sahip veriler türetilmiştir. Benzetim çalışmasında 9 farklı senaryo ve 5 farklı referans aralık yöntemi ile bireyi ilgilenilen hastalık bakımından hasta ve sağlıklı olarak sınıflandırarak karşılaştırmalar yapılmıştır.

Bulgular: Tüm senaryolarda hastalık sınıflandırma oranı en yüksek Uni1 ve en düşük Multi2 yöntemlerinde bulundu. Karşılaştırıldığında, Multi1 yöntemi, Mahalanobis yöntemine yaklaşık olarak benzer bir hastalık oranına sahipti. Tüm yöntemlerde örneklem sayısı arttıkça bulunan hasta sayısının değişkenliği azalmaktadır.

Sonuç: Referans aralıklar, sağlıklı kitlenin merkezdeki %95'ini karakterize etmek için kullanılan istatistiksel olarak türetilmiş araçlar olduğu için, Mahalanobis yöntemiyle her senaryoda ulaşılan yaklaşık %5'lik hasta oranı bu yöntemin çok değişkenli referans alanının belirsizliği ve uygulama zorluğu karşısında önerilebileceğini kanıtlamaktadır. Mahalanobis %5 değerine en yakın hasta sınıflama oranını vermesine rağmen hekimin kolaylıkla değerlendirme yapabileceği alt ve üst referans sınırını vermediği için bu yöneme en yakın sonuçları veren çok değişkenli 1 yönteminin kullanılması klinik değerlendirmede kolaylık sağlayacaktır.

Anahtar Kelimeler: Mahalanobis Uzaklığı, referans aralığı, çok değişkenli referans bölgesi, tek değişkenli referans aralığı, simülasyon çalışması

Yayın Yanlılığını Önlemede P-Eğrisi Yöntemi

¹Arzu BAYGÜL EDEN, ²Alev Bakır KAYI

¹Koç Üniversitesi Tıp Fakültesi

²İstanbul Üniversitesi, Çocuk Sağlığı Enstitüsü, Sosyal Pediatri Anabilim Dalı, İstanbul

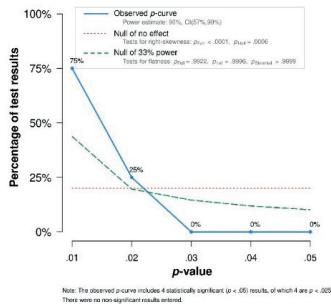
Amaç: Yayın yanlılığı ve P-korsanlığı özellikle meta analizlerde rastlanılan etik dışı ve gerçek dışı sonuçların elde edilmesini sağlayan ve direkt olarak insan sağlığını etkileyebilecek sonuçlara yolaçabilecek kavramlardır. Bu sorunları bertaraf edebilmek adına günümüzde farklı yöntemler kullanılmaktadır. En sık kullanılan yöntemler Funnel plot, Eggers rank korelasyon testi gibi yöntemlerdir. Çalışmamızda bu yöntemlere ek olarak p-eğrisi yöntemini inceleyeceğiz.

Gereç ve Yöntem: P-eğrisi yöntemi ilk olarak Simonsohn ve arkadaşları tarafından 2014 yılında, araştırmacıların özellikle kendi tarafında olan sonuçları yayınlama eğilimlerini fark edip, “Acaba bu etkiler doğru mudur, yoksa bir selektif yayın sonucu mu elde edilmiştir?” sorusuna bir cevap bulma amacıyla ortaya atılmıştır. Bu yöntemde herhangi bir konuda yayınlanmış çalışmalardan elde edilen istatistiksel olarak anlamlı “p” değerlerinden ($p < 0.05$) oluşan bir dağılım incelenir. Bu dağılımın şekli bize çalışmaların kanıtsal bir bilgi sağlayıp sağlamayacağını söyleyecektir. Başka bir deyişle elde edilen şekillere göre tekrar edilebilir sonuçlar elde edilip edilmeyeceği hakkında bir fikir sahibi olunacaktır.

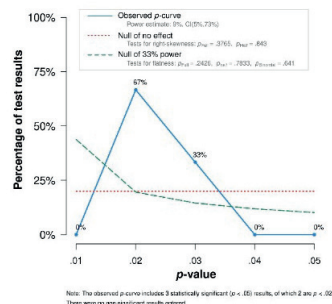
Bulgular: Uygulama1 Uygulama çevrimiçi P-curve.com/app uygulaması kullanılarak (<http://www.p-curve.com/app4/pcurve4.php>) gerçekleştirilmiştir. Farklı çalışmaların p değerlerini ekleyerek p- eğrisi çizdirilmiş ve elde edilen şekle göre sonuçların “kanıtsal” olup olmaması incelenmiştir. Girdiler: $r(147)=0,246$; $F(1,100)=9,1$; $f(2,210)=4,45$; $t(200)=5.2$ Sonuçların kanıtsal olması için yarı p-eğrisinin sağa çarpık olması için yapılan testte $p < 0,05$ veya yarı ve tam p eğrisinin sağa çarpık olması için yapılan testte $p < 0,10$ olmalıdır. Bu şartlar sağlandığı için çalışmaların kanıtsal sonuç sağladığı söylenebilir. Uygulama 2 Girdi: $f(2,210)=4$; $r(77)=0,25$; $\chi^2(2)=8$ Sonuçların kanıtsal olduğunu gösteren şartlar sağlanmadığından bu çalışmalar açısından bir yayın biasından söz edilebilir.

Sonuç: p-eğrisi yayın hatasını ve selektif raporlamayı belirlemede önemli ve etkin rol oynamaktadır. Özellikle meta analiz çalışmalarında çalışmaya dahil edilecek yayınların bulunması ya da ortak bir etki arandığında “gerçek” etkiyi ifade eden çalışmaların bulunması için p-eğrisi yöntemine başvurulması iyi bir tercih olacaktır.

Anahtar Kelimeler: p-eğrisi, yayın biası, meta analiz



Şekil-1. Uygulama1



Şekil-2. Uygulama 2

Tablo-1. Uygulama 1

Binomial Test <i>(Share of results $p < .025$)</i>		Continuous Test <i>(Aggregate with Stouffer Method)</i>	
		Full p -curve <i>($p < .05$)</i>	Half p -curve <i>($p < .025$)</i>
1) Studies contain evidential value. <i>(Right skew)</i>	$p = .0625$	$Z = -4.04, p < .0001$	$Z = -3.25, p = .0006$
		Statistical Power	
Power of tests included in p -curve <i>(correcting for selective reporting)</i>		Estimate: 90% 90% Confidence interval: (57% , 99%)	

Tablo-2. Uygulama 2

Binomial Test <i>(Share of results $p < .025$)</i>		Continuous Test <i>(Aggregate with Stouffer Method)</i>	
		Full p -curve <i>($p < .05$)</i>	Half p -curve <i>($p < .025$)</i>
1) Studies contain evidential value. <i>(Right skew)</i>	$p = .5$	$Z = -0.31, p = .3765$	$Z = 1.01, p = .843$
		Statistical Power	
Power of tests included in p -curve <i>(correcting for selective reporting)</i>		Estimate: 9% 90% Confidence interval: (5% , 73%)	

Kanıtsal Güç ve Literatür Bütünlüğünün p-Eğrisi ile İncelenmesi

¹Alev BAKIR KAYI, ²Arzu BAYGÜL EDEN

¹İstanbul Üniversitesi, Çocuk Sağlığı Enstitüsü, Sosyal Pediatri Anabilim Dalı, İstanbul

²Koç Üniversitesi, Tıp Fakültesi, Biyoistatistik Bölümü, İstanbul

Amaç: Kanıt piramidinin en üstünde yer alan, tıbbi kararlar alırken en çok güven duyulan ve kanıta dayalı tıbbın bel kemiği olan meta-analiz çalışmalarında yanlış seçimler ve p-korsanlığı (p-hacking) yapılabilme ihtimalinden dolayı, meta-analiz sonuçlarının şüpheli bir geçerliliğe sahip olabileceğine dair endişeler artmaktadır. Meta-analiz sonucunda anlamlı ortak bir etki kararına varılsa da sonuçların kanıtsal olup olmadığını, tekrarlanabilirliğini p-eğrisi yöntemi ile incelenmesi amaçlanmaktadır.

Gereç ve Yöntem: Farkındalık temelli müdahalelerin dikkat eksikliği hiperaktivite bozukluğu semptomları üzerindeki etkisi meta analizi ve diyabetin transkateter aort kapak replasmanından sonra artan 1 yıllık mortalite ile ilişkisi meta-analizi çalışmalarındaki test istatistikleri kullanılarak elde edilen p dağılımı ile meta-analizlerin kanıt güçleri p-eğrisi yöntemi ile incelendi. P-eğrisi yönteminin R kodu ile çalışan hazır bir web uygulamasına başvurulmuştur.

Bulgular: Farkındalık temelli müdahalelerin dikkat eksikliği tedavisinde etkili olduğu ile ilgili derlenmiş 682 katılımcıyı içeren 11 makale ile yapılan meta-analizin, p-eğrisi yöntemi ile kanıtsal sonuç sağladığı görülmüştür (yarı p-eğrisi için $Z=-6,97$, $p<0,0001$, tam p-eğrisi için $Z=-6,46$, $p<0,0001$ ve istatistiksel güç %96 (%90 G.A.: %85-%99)). Diyabetin, transkateter aort kapak replasmanından sonra 1 yıllık mortalite ile ilişkisini incelemek için 17.503 börek hastasını kapsayan 22 çalışmayı içeren meta-analiz sonucunda anlamlı ortak bir karara varılmasına rağmen, p-eğrisi yarı p-eğrisi $Z=0,29$, $p=0,614$, tam p-eğrisi $Z=1,03$, $p=0,847$ ve istatistiksel güç %5 (%90 G.A.: %5-%22) şeklinde olup kanıtsal bir sonuç sağlamadığı bulunmuştur.

Sonuç: Meta-analiz sonuçlarının p-eğrisi yöntemi ile de değerlendirilmesi hem kanıtsal bir sonuç elde edilip edilmediğini hem de bulunan kanıtsal güce olan güveni arttıracak için literatüre katkı sağlayacaktır.

Anahtar Kelimeler: p-eğrisi, meta-analiz, kanıtsal güç

The Effectiveness of Blood Routine Parameters and Some Biomarkers as a Potential Diagnostic Tool in the Diagnosis and Prognosis of COVID-19 Disease

¹Mehmet Tahir HUYUT, ²Siddik KESKİN

¹Erzincan Binali Yıldırım Üniversitesi, Biyoistatistik ve Tıbbi Bilişim A.D, Erzincan

²Yüzüncü Yıl Üniversitesi, Biyoistatistik A.D, Van

Aim: Due to the complex pathogenesis of COVID-19, early diagnosis and timely treatment are important. There is still a lack of information about the effects of routine blood values (RBV) on the disease process, and it is necessary to generalize and increase the reliability of existing studies. The aim of this study is to evaluate the effectiveness of new biomarkers derived from these blood values instead of using direct blood values for the diagnosis and prognosis of COVID-19 in a large patient group

Materials and Methods: Two groups were reported in this study, those treated in the intensive care unit and subjects in the all services unit. In addition, the same number of control groups was determined as the patient group. The following criteria were applied when designing derived biomarkers to determine the diagnosis and prognosis of the disease: 1) Having a significant probability ratio as a result of Multivariate Logistic Regression analysis, 2) Having significant AUC values as a result of ROC analysis, 3) Reporting of biomarkers that meet the Inflation Factor of Variance (VIF) <10 criterion among variables. In addition, a decision tree suitable for clinician use was created to check whether the evaluation of predictors together is more successful than evaluations with a single feature in determining the diagnosis and prognosis of the disease. The accuracy rate of the decision tree in predicting the diagnosis prognosis of the disease was calculated.

Results: Low lymphocyte (LYM) and white-blood-cell (WBC), high CRP and Ferritin were effective in the diagnosis of the disease. The (d-CWL) = CRP*WBC*LYM and (d-CFL) = CRP*Ferritin*LYM biomarkers derived from them were the most important risk factors in diagnosing the disease and were more successful than direct RBV values. High d-CWL and d-CFL values largely confirmed the Covid-19 diagnosis. The most effective RBV in the prognosis of the disease was CRP. (d-CIT) = CRP*INR*Troponin; (d-CT) = CRP*Troponin; (d-PPT) = PT*Troponin*Procalcitonin biomarkers were found to be more successful than direct RBV values and biomarkers used in previous studies in the prognosis of the disease. In this study, biomarkers derived from RBV were found to be more successful in both diagnosis and prognosis of COVID-19 than previously used direct RBV and biomarkers.

Conclusion: It was observed that the biomarkers we derived from routine laboratory tests determined the diagnosis and prognosis of COVID-19 more successfully than direct blood values and biomarkers previously entered in the literature. The independent biomarkers derived in this study are thought to provide an alternative to clinical approaches in the diagnosis and prediction of the prognosis of COVID-19.

Keywords: Covid-19, biochemical, hematological and inflammatory biomarkers, blood routine parameters, covid-19's independent biomarkers.

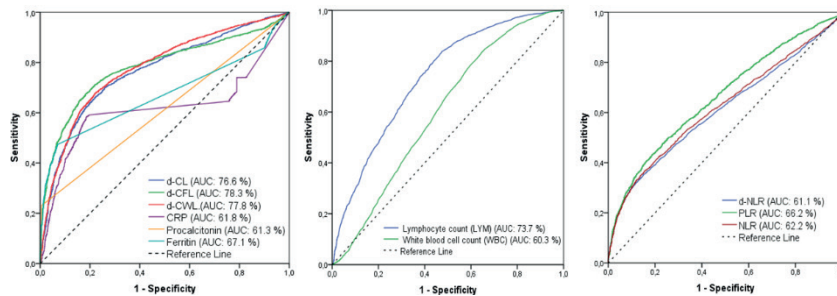


Fig-1. The left figure and the middle figure are the ROC curves for the diagnosis of Covid-19 of biomarkers derived with blood parameters. The figure on the right is the ROC curves of the biomarkers used in the literature for the diagnosis of this study.

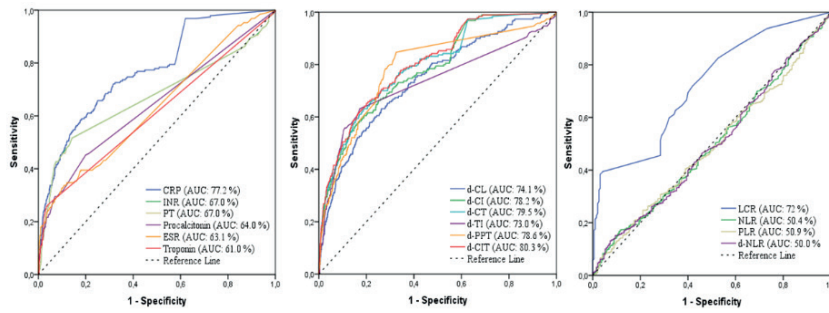


Fig-2. In the distinction between non-ICU and ICU (ie, prognosis of the disease): the left figure is the blood parameters; the middle figure is of newly derived biomarkers; The figure on the right is the ROC curves of biomarkers in the literature.

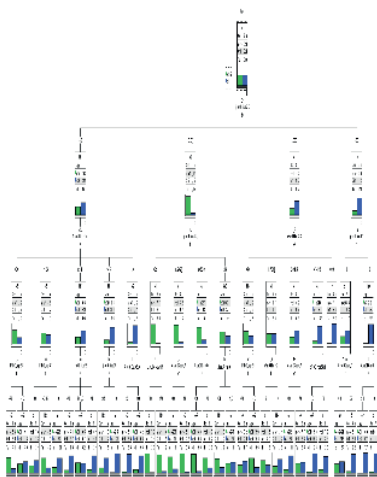


Fig-3. Decision tree evaluating the effects of predictive variables together in the diagnosis of Covid-19.

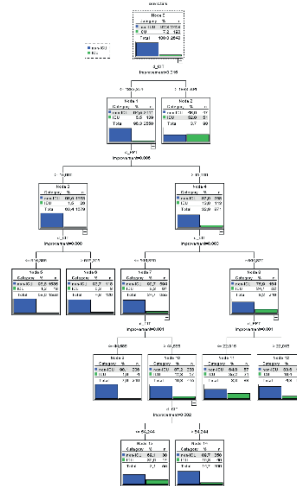


Fig-4. Decision tree evaluating the combined effect of predictive variables in determining patients to non-ICU and ICU in Covid-19.

Çok Değişkenli Varyans Analizinde Kullanılan Test İstatistiklerinin Tip I Hata Oranları Bakımından Karşılaştırılması

¹Fatma Ezgi CAN, ²Büşra EMİR, ³Ferhan ELMALI, ⁴Can ATEŞ, ⁵Mustafa Agah TEKİNDAL

¹İzmir Kâtip Çelebi Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, İzmir

²İzmir Kâtip Çelebi Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, İzmir

³İzmir Kâtip Çelebi Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, İzmir

⁴Aksaray Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Aksaray

⁵İzmir Kâtip Çelebi Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, İzmir

Amaç: Çok değişkenli varyans analizi, MANOVA (Multivariate Analysis of Variance), iki veya daha fazla bağımlı değişkenden oluşan verilerin analizinde kullanılmaktadır. Bu çalışmada dengeli ve dengesiz örneklerde MANOVA yönteminin normallik ve varyansların homojenliği varsayımlarının sağlanmadığı durumlarda Wilks Lamda, Pillai iz, Hotelling iz ve Roy'un en büyük özdeğer test istatistiklerinin Tip I hata oranları araştırılmıştır.

Gereç ve Yöntem: Test istatistiklerinin Tip I Hata oranlarını karşılaştırmak amacıyla 10000 tekrarlı farklı senaryolardan oluşan simülasyon çalışması yapılmıştır. Lognormal (1;0,5), Weibull (2,1;1) ve Weibull (0,68;1) dağılımlarından farklı grup, farklı değişken sayıları, dengeli ve dengesiz örneklem büyüklüklerinde, sabit ve artan grup varyansları dikkate alınarak rasgele sayılar türetilmiştir. $\alpha = 0,05$ alınarak her test istatistiği için Tip I hata oranları hesaplanmıştır. Simülasyon çalışmasında R v.4.2.1 program dili ve MASS v. 7.3-57 paketi kullanılmıştır.

Bulgular: Yapılan simülasyon çalışması sonucunda Lognormal dağılımda sabit varyans ve dengesiz örneklerde Pillai iz test istatistiğinin Tip I hata oranları $\alpha = 0,05$ değerine en yakın sonuçları vermiştir. Değişen varyans, 2 ve 3 değişkenli dengeli örneklerde ise Hotelling iz istatistiği en yakın sonuçları göstermiştir. Değişen varyans, 3 değişkenli dengesiz örneklerdeyse Roy'un en büyük özdeğer test istatistiği en yakın sonuçları vermiştir. Weibull dağılımında sabit varyans ve 2 değişkenli dengesiz örneklerde $\alpha = 0,05$ değerine en yakın sonuçlar Hotelling iz test istatistiğinden elde edilirken; 3 değişkenli olduğu durumda Wilks Lamda istatistiği en yakın sonucu göstermiştir. Weibull (0,68;1) dağılımından elde edilen sonuçlarda 2 değişkenli dengeli ve dengesiz örneklerde Pillai iz, 3 değişkenli dengeli örneklerde Hotelling iz, 3 değişkenli dengesiz örneklerdeyse Pillai iz test istatistiği en yakın sonuçları göstermiştir.

Sonuç: Simülasyon çalışmasında elde edilen sonuçlara göre Lognormal dağılımın Weibull dağılımına göre daha tutarlı sonuçlar verdiği görülmüştür.

Anahtar Kelimeler: Wilks lamda istatistiği, pillai iz istatistiği, hotelling iz istatistiği, roy'un en büyük özdeğer istatistiği, MANOVA

Obuchowski Parametrik olmayan ROC Eğrisi Analizi ve Küme Bootstrap Yaklaşımları ile İlişkili Verilerin Değerlendirilmesi: Multipl Skleroz Hastalarında Optik Nöropati Teşhisi Üzerine Bir Uygulama

¹Büşra EMİR, ²Fatma Ezgi CAN, ³Mustafa Agah TEKİNDAL, ⁴Ferhan ELMALI, ⁵Ertuğrul ÇOLAK

¹İzmir Katip Çelebi Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı

²İzmir Katip Çelebi Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı

³İzmir Katip Çelebi Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı

⁴İzmir Katip Çelebi Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı

⁵Eskişehir Osmangazi Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı

Amaç: Bu çalışmada, ilişkili verilerin ROC (Receiver Operating Characteristic) eğrisi ile değerlendirilmesinde önerilen Obuchowski parametrik olmayan ROC eğrisi analizi ve Küme Bootstrap yaklaşımları incelenmiştir. Temel motivasyonumuz bağımlı kümelenmiş verilerin değerlendirilmesinde üç yöntemin karşılaştırılmasıdır.

Gereç ve Yöntem: Multipl Skleroz hastalarında Optik Nöropati (ON) teşhisi üzerine (ON +/-), iki farklı görüntüleme tekniğine (OCT, GDx) ait sağ ve sol göz retina sinir lifi tabakası kalınlığı ölçümleri için Obuchowski parametrik olmayan ROC eğrisi analizi ve Küme Bootstrap yöntemlerine ait eğri altında kalan alan (AUC) değerleri ve %95 güven aralıkları hesaplanmıştır. Orijinal veri setine ait sağ ve sol göz RNFL kalınlığı ölçüm ortalama ve standart sapma değerleri, örneklem büyüklükleri ve değişen gözler arası korelasyon yapısı ($r=0.25, 0.50$ ve 0.75) kullanılarak çok değişkenli normal dağılımdan üç farklı veri seti oluşturulmuştur. Değişen gözler arası korelasyonu dahil etmeden gerçekleştirilen parametrik ROC eğrisi analizi sonuçları ile hesaplanan yaklaşımlar karşılaştırılmıştır. Obuchowski parametrik olmayan ROC eğrisi analizi RStudio, parametrik ROC eğrisi analizi ve küme bootstrap yaklaşımı SAS University Edition programı kullanılarak gerçekleştirilmiştir.

Bulgular: Yetmiş iki hasta ve 144 retina sinir lifi tabakası kalınlığı ölçümlerinin yer aldığı optik nöropati (ON) çalışmasında, ON teşhisi için optik koherens tomografi görüntüleme tekniği (OCT) ile elde edilen retina sinir lifi tabakası (RNFL) kalınlığının AUC değeri 0.705 (%95 güven aralığı (%95 GA): 0.616, 0.793) olarak elde edilmiştir. Gözler arası korelasyonu hesaba katmadan parametrik yaklaşım ile elde edilen AUC değeri ve güven aralığı, Obuchowski parametrik olmayan ROC eğrisi analizi (%95 GA: 0.607, 0.803) ve küme bootstrap (%95 GA: 0.600, 0.800) yaklaşımlardan daha dardı. Tarayıcı lazer polimetri görüntüleme tekniği (GDx) ile elde edilen RNFL kalınlığının AUC değeri 0.666 (%95 GA : 0.576, 0.757) olarak elde edilmiştir. Gözler arası korelasyonu hesaba katmadan parametrik yaklaşım ile elde edilen AUC değeri, Obuchowski parametrik olmayan ROC eğrisi analizi (%95 GA: 0.564, 0.769) ve küme bootstrap (%95 GA: 0.555, 0.764) yaklaşımlardan daha dardı.

Sonuç: İlişkili verilerin ROC analizinde, veriler arası korelasyonun göz ardı edilerek parametrik ROC eğrisi analizinin kullanılması, korelasyonun büyüklüğüne bağlı olarak %95 güven aralığının daha dar olmasına sebep olmuştur. İlişkili verilerin analizinde Obuchowski parametrik olmayan ROC eğrisi analizi ya da küme bootstrap yaklaşımlarının kullanılması ile daha geniş güven aralığı elde edilebilir.

Anahtar Kelimeler: Küme bootstrap, obuchowski, parametrik olmayan roc eğrisi

Genom-boyu İlişki Çalışmalarında Poligenik Risk Skorunun Makine Öğrenimi ve Derin Öğrenme Yöntemleri ile Tahmin Edilmesi

¹Ragıp Onur ÖZTORNACI, ²Erdal COŞGUN, ³Bahar TAŞDELEN, ³Cemil ÇOLAK

¹Koç Üniversitesi Translasyonel Araştırma Merkezi, Koç Üniversitesi, İstanbul, Türkiye

²Microsoft Research, 14820 NE 36th Street, Building 99, Redmond, WA, 98052, USA

³Mersin Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Mersin, Türkiye

³İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Malatya, Türkiye

Amaç: Son yıllarda genetik yatkınlığın değerlendirilmesinde Poligenik risk skorunun (PRS) kullanılması yaygın bir yaklaşımdır. PRS, aynı anda birden çok SNP kullanılarak, bir hastalık için genetik risk skoru sağlayan bir ölçüttür ve tipik bir hastalık skoru olarak hesaplanabilir. Bu çalışmada, klasik PRS hesaplama yöntemlerine alternatif olarak derin öğrenme (DL) ve makine öğrenmesi (ML) kullanılarak literatüre alternatif yöntemler katılması ve klasik yöntemin validasyonu amaçlanmıştır.

Gereç ve Yöntem: PLINK Programı ile Genom-boyu İlişki Çalışmaları (GWAS) verileri farklı allel frekanslarında, farklı örnek genişliklerinde yüz kez tekrar edilerek oluşturulmuştur. DL ve ML yöntemleri ile elde edilen yeni ağırlıklar ile PRS hesaplanmıştır. Elde edilen PRS sonuçlarının dağılımı üzerinde durularak, çevresel faktörlerle birlikte nasıl yorumlanacağı üzerine durulmuştur. Kullanılan programlar ve işletim sistemleri; R, Python ve Ubuntu Linux Operating System'dir.

Bulgular: Farklı örnek genişliklerinde ve farklı SNP sayılarında, ML ile hesaplanan PRS, klasik yöntem (PRS) ile hesaplanan PRS'e göre ortalamalar bakımından hasta-kontrol ayırımını yapmada daha tutarlı sonuçlar elde edebileceği görülmüştür. Benzer bir yaklaşımla SNP sayısını sabit tutulup örnek genişliğindeki artışın sırasında meydana gelen değişiklikler gözlemlenirse; klasik yöntem (PRS), destek vektör makineleri (SVM) ve DL için etkili olabileceği görülürken, rastgele orman (RF) yöntemi için değişim gözlemlenmemiştir. SNP sayısı sabitken örnek genişliğindeki değişimden, klasik PRS ve DL etkilenmiştir.

Sonuç: Klasik yöntemle PRS hesaplanırken eğitim ve test seti olmak üzere iki farklı veri seti kullanılması çok sık kullanılmaktadır. Bu her zaman kolay olmamakla birlikte klasik yöntem için bir dezavantaj getirmektedir. ML ve DL yöntemleri tek bir veri seti kullanarak eğitim aşamasında yapılan çapraz doğrulama ile hem aşırı uyum sorunundan kaçınmaktadır hem de klasik yöntemin iki farklı veri seti kullanımı dezavantajına karşın iyi bir alternatif yöntem olarak kullanılabilir. Klasik PRS hesaplama yöntemine göre ML ve DL hem daha dar değişim aralığı ve daha dar güven aralıkları elde etmişlerdir. PRS hesaplamak için klasik yöntemlere alternatif olarak ML ve DL yöntemlerinin kullanılması, alternatif yöntemler olarak önerilmiştir.

Anahtar Kelimeler: Genom-boyu ilişki çalışmaları (GWAS), Poligenik risk skoru (PRS), Makine öğrenimi, derin öğrenme.

Doğrusal Olmayan Yapısal Eşitlik Modellemesi Yaklaşımlarının İncelenmesi ve Bir Uygulama

¹Nur Efsan TİĞLİ, ²Şengül CANGÜR

¹Düzce Üniversitesi Biyoistatistik Anabilim Dalı, Düzce

²Düzce Üniversitesi Biyoistatistik Anabilim Dalı, Düzce

Amaç: Bu çalışmada değişkenler arasındaki ilişkilerin doğrusal olmadığı ya da normal dağılım varsayımı sağlanamadığı durumlarda kullanılan gizil moderatörlü yapısal eşitlikler (GMYE), yapısal eşitlik karışım modelleri (YEKM) ve doğrusal olmayan yapısal eşitlik karışım modellemesi (DOYEKM) yaklaşımlarının tanıtılması ve gerçek veri seti üzerinde uygulamasının gösterilmesi amaçlanmıştır.

Gereç ve Yöntem: Doğrusal olmayan YEM kapsamında üç yaklaşım üzerinde durulmuştur. Birincisi, etkileşim etkisi veya ikinci dereceden etkiler gibi gizil değişkenleri kullanarak doğrusal olmayan ilişki türlerini analiz etmek için kullanılan GMYE yaklaşımı, ikincisi ise doğrusal olmayan gizil değişkenlerin doğrusal ilişkilerin karışımı ile yaklaşık olarak tahmin edilmesi için kullanılan YEKM yaklaşımıdır. Bu iki yaklaşımın birleşimini gösteren üçüncü bir yaklaşım olarak da yarı parametrik yöntem olarak geliştirilen DOYEKM yaklaşımı ele alınmıştır. Bu üç yaklaşım kullanılarak Düzce Üniversitesi Tıp Fakültesi öğrencilerinin (n=272) yalnızlık ve yaşam doyumunun internet bağımlılığı üzerindeki etkilerini incelemek için yapısal modeller oluşturulmuştur. Model karşılaştırmaları için Akaike bilgi kriteri (AIC), Schwarz Bayes bilgi kriteri (SBIC) ve loglikelihood değerleri hesaplanmıştır. Verilerin analizi için R-CRAN 3.6.2 ve SPSS 24 programları kullanılmıştır.

Bulgular: AIC ve SBIC kriterlerine göre en küçük AIC ve SBIC değerlerine sahip model en uygun model olarak kabul edildiğinden, DOYEKM yaklaşımının diğer iki yaklaşıma göre daha kullanışlı olduğu görülmüştür. Yalnızlık gizil değişkeninin internet bağımlılığı üzerinde anlamlı pozitif bir etkisi olduğu, yaşam doyumu gizil değişkeninin ise anlamlı negatif bir etkisi olduğu saptanmıştır. Ayrıca hem yalnızlık ve yaşam doyumu gizil değişkenlerinin etkileşim etkisinin hem de yaşam doyumu gizil değişkeninin karesel etkisinin anlamlı düzeyde negatif etkilere sahip oldukları bulunmuştur.

Sonuç: Değişkenler arasındaki ilişkilerin doğrusal olmadığı ya da normal dağılım sağlanamadığı durumlarda klasik YEM yaklaşımlarının yetersiz kalmasına karşı geliştirilen doğrusal olmayan YEM yaklaşımlarının özellikle sağlık ve davranış bilimleri başta olmak üzere tüm alanlarda yaygın bir şekilde kullanılmasını önermekteyiz.

Anahtar Kelimeler: Gizil moderatörlü yapısal eşitlikler, Yapısal eşitlik karışım modelleri, Doğrusal olmayan yapısal eşitlik karışım modellemesi

Machine Learning Sensors for Diagnosis of COVID-19 Disease Using Routine Blood Values for Internet of Things Application

¹Mehmet Tahir HUYUT, ²Andrei VELICHKO, ³Belyaev MAKSIM

¹Erzincan Binali Yıldırım Üniversitesi, Biyoistatistik ve Tıbbi Bilişim A.D, Erzincan

²Petrozavodsk State University, Institute of Physics and Technology, 33 Lenin Str., 185910 Petrozavodsk, Russia

³Petrozavodsk State University, Institute of Physics and Technology, 33 Lenin Str., 185910 Petrozavodsk, Russia

Aim: Healthcare digitalization needs effective methods of human sensorics, when various parameters of the human body are instantly monitored in everyday life and connected to the Internet of Things (IoT). In particular, Machine Learning (ML) sensors for the prompt diagnosis of COVID-19 is an important case for IoT application in healthcare and Ambient Assistance Living. This study aims to provide a fast, reliable and economical tool for the diagnosis of COVID-19 based on routine blood values.

Materials and Methods: In this study, 51 features belonging to the same number of patients with positive and negative COVID-19 test results were used (a total of 5296 patients). In this study, the performance of 13 popular supervised ML models and LogNet deep neural network models in the diagnosis of the disease was investigated. Before the models were trained, the data were normalized and then additional features were generated. Quantitative transformer (QT) for the normalization process; Robust scaler (RS); MinMax (MM) methods were used. Polynomial (PN) to create additional properties; Random trees embedding (RTE); Extra trees preprocessor (ETP); Linear SVM preprocessor (LSVMP); Independent component analysis (ICA); Nystroem sampler (NS) methods were used. The “auto-sklearn” software package adjusted to the target indicator-classification accuracy was used for optimization of each classifier model. The accuracy of the models was evaluated by the K-fold cross validation method.

Results: The most successful classifier model in terms of time and accuracy in the detection of the disease was the Histogram-based Gradient Boosting (HGB) (accuracy: 100%, time: 6.39 sec). The HGB classifier identified the 11 most important features to detect the disease with 100% accuracy. In addition, the accuracy rate of the single, double and triple combinations of these features operated with HGB in the diagnosis of the disease was determined. Paired combinations of these features run with HGB achieved higher accuracy in diagnosing the disease than the direct features. Also, the implementation of the LogNet network on the arduino demonstrated the practical applicability of the model in low-power devices (ie the concept of the Internet of Things).

Conclusion: We propose to use these 11 traits and their binary combinations as important biomarkers for ML sensors in diagnosis of the disease, supporting edge computing on Arduino and cloud IoT service. The results make it possible to create the concept of Machine Learning Sensors for the Diagnosis of COVID-19 Disease. In addition, the results of this study can be effectively used in IoT (IoT/M) medical peripheral devices with low RAM resources, ML sensors, portable point-of-care blood testing devices, clinical decision support systems, remote internet medicine and telemedicine.

Keywords: COVID-19; biochemical and hematological biomarkers; routine blood values; feature selection method; LogNet neural network; Machine Learning sensors, Internet of Medical Things; IoT.

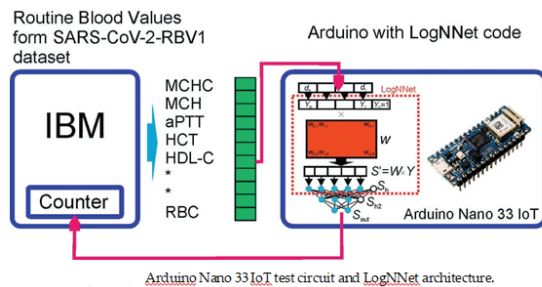
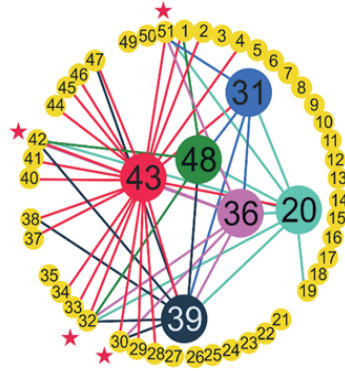


Fig-1. Arduino Nano 33 IoT test devresi, LogNet mimarisi ve kartı Fig-1’de gösterilmiştir.

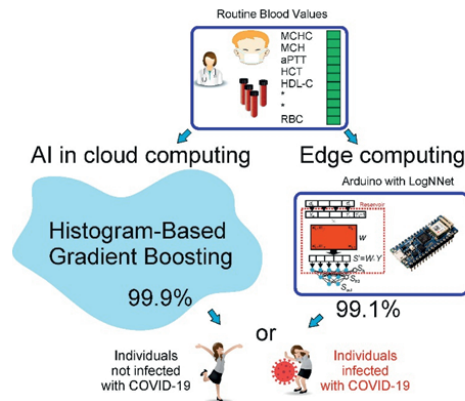
Table-1. Results of assessing the classification accuracy of machine learning models for the SARS-CoV-2-RBV1 dataset.

Results of assessing the classification accuracy of machine learning models for the SARS-CoV-2-RBV1 dataset.				
Classification algorithm	Average model accuracy Ası, %	Average learning time, s	Normalization method	Methods for generating additional features
Histogram based Gradient Boosting	100	6.39	-	-
Random Forest	99.943	13.15	QT	-
K-nearest neighbors	99.924	3.17	QT	ETP
Extra Trees classifier	99.905	18.73	RS	-
Multilayer Perceptron	99.886	3.99	RS	LSVMP
Multinomial Naive Bayes	99.792	2.48	QT	RTE
Linear Discriminant Analysis	99.773	9.15	QT	PN
Support Vector Machine with non-linear kernel	99.754	222.41	QT	NS
Decision Tree	99.660	1.46	RS	LSVMP
Passive-Aggressive	99.641	2.91	QT	RTE
Bernoulli Naive Bayes	99.622	2.59	QT	RTE
Support Vector Machine with linear kernel	99.584	5.21	MM	PN
Gaussian Naive Bayes	98.565	1.68	QT	ICA
LogNet [37]	99.509	100	-	-



Pairs of features with high classification efficiency SARS-CoV-2-RBV1 datasets for the Histogram based Gradient Boosting classifier.

Fig-2. Pairs of features with high classification efficiency SARS-CoV-2-RBV1 datasets for the Histogram based Gradient Boosting classifier.



Two architectures of the IoT system, which includes an IoT device with LogNetNet implementation (edge computing) and a cloud service containing a trained HGB model (AI computing).

Fig-3. Two architectures of the IoT system, which includes an IoT device with LogNetNet implementation (edge computing) and a cloud service containing a trained HGB model (AI computing).

Sınıflamada Kullanılan Veri Madenciliği Yöntemlerinin Performanslarının D Vitamini Veri Seti Üzerinde Karşılaştırılması

¹Özlem ARIK, ²İnci ARIKAN, ³Erdem KARABULUT

¹Kütahya Sağlık Bilimleri Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Kütahya

²Kütahya Sağlık Bilimleri Üniversitesi, Tıp Fakültesi, Halk Sağlığı Anabilim Dalı, Kütahya

³Hacettepe Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Kütahya

Amaç: Büyük miktarda biriken ham veri setlerinden anlamlı, değerli ve faydalı bilgilerin ortaya çıkarılmasına veri madenciliği denir. Tahmin edici yöntemler arasında sınıflama yöntemleri bulunmaktadır. Bu çalışmadaki temel amaç, Kütahya Sağlık Bilimleri Üniversitesi Evliya Çelebi Eğitim ve Araştırma Hastanesine 2017-2021 yılları arasında ayaktan başvuran 121727 hastaya ait bazı bilgiler kullanılarak hastanın D vitamini düzeyini sınıflamak ve sınıflama yöntemlerinin performansını ölçmektir.

Gereç ve Yöntem: Çalışmada, hastaneye ayaktan başvuran hastalardan 25-OH VitD düzeyi çalışılanlar hastane veri tabanı siste-minden geriye dönük olarak incelendi. Hastaların yaş, cinsiyet, mevsim ve 25-OH VitD düzeyi bilgileri kullanıldı. Veri ile ilgili ön işlemeden sonra veri madenciliği sınıflama yöntemlerinden k-en-yakın komşu (kNN), karar ağaçları, destek vektör makineleri (DVM), random forest, yapay sinir ağları (YSA), naive bayes (NB) ve lojistik regresyon kullanıldı. Doğru sınıflama oranı, duyarlılık, seçicilik, F1 ölçüsü, kesinlik ve ROC eğrisi altında kalan alan (EAKA) ölçüleri aracılığıyla yöntemlerin performansları belirlendi.

Bulgular: D vitamini verisi için 25-OH VitD'nin 20 ng/ml'den (50 nmol/L) küçük düzeyleri Vitamin D eksikliği olarak ifade edilmiştir. Vitamin D eksikliği var-yok şeklinde ve hedef sınıf: Vitamin D eksikliği var olduğunda sınıflama modellerine ait elde edilen bulgular Tablo 1`de yer almaktadır.

Sonuç: Verilerin analizinin Orange 3.32 programı aracılığıyla yapıldığı çalışmada, Tablo 1`de verilen doğru sınıflama oranı, duyarlılık, F1 ölçüsü, kesinlik ve EAKA değerlerine göre karar ağaçları, random forest ve yapay sinir ağları yöntemlerinin performansları birbirine yakındır ve diğerlerine göre daha iyidir. ROC eğrisi altında kalan alan (EAKA) değerleri incelendiğinde ise hastaları D vitamini düzeylerine göre sınıflara ayırmak için kullanılan yöntemler içerisinde DVM modelinin performansının diğerlerine göre daha düşük olduğu sonucuna ulaşılmıştır.

Anahtar Kelimeler: Sınıflama, performans, D vitamini

Psikiyatrik Hastalıklarda, Psikopatoloji ve Çevresel Etken İlişkisinin Özyinelemesiz Model ile Değerlendirilmesi

¹Selen Begüm UZUN, ²Derya GÖKMEN, ³Meram Can SAKA

¹T.C. Sağlık Bakanlığı, Sağlık Bilgi Sistemleri Genel Müdürlüğü, Sağlık İstatistikleri Dairesi Başkanlığı, Ankara

²Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

³Ankara Üniversitesi, Tıp Fakültesi, Ruh Sağlığı ve Hastalıkları Anabilim Dalı, Ankara

Amaç: Psikiyatri alanında; çevresel etmenler, hastalık durumu ve şiddeti, hastaların işlevselliği, algıladıkları sosyal destek ve sosyal ilişkileri psikopatoloji üzerinde etkili değişkenler olarak değerlendirilmektedir. Bu çalışmada, çevresel etmenler, hastalık durumu, hastalık şiddeti gibi değişkenlerin psikopatoloji ile ilişkisinin özyinelemesiz eş zamanlı denklem modelleri ile değerlendirilmesi amaçlanmıştır.

Gereç ve Yöntem: Çalışma, Mayıs-Haziran 2017 tarihleri arasında Ankara Üniversitesi Tıp Fakültesi Ruh Sağlığı ve Hastalıkları Anabilim Dalı Polikliniğine başvuran ve çalışmaya katılmayı kabul eden 378 hasta üzerinde gerçekleştirilmiştir. Çalışma kapsamında; hastaların demografik özelliklerinin yanısıra, Çocukluk Çağı Travma Ölçeği (ÇÇT), Rosenberg Benlik Saygısı, Çok Boyutlu Algılanan Sosyal Destek Ölçeği (ÇBASD), Bireysel ve Sosyal Performans Ölçeğine (PSP) verdikleri yanıtlar bulunmaktadır. Hastalık durumu; hekim tarafından yapılan konsültasyon sonrasında ICD-10 kodları kullanılarak; belirti/semptom şiddeti ise Klinik Global İzlenim Ölçeği kullanılarak belirlenmiştir. Çalışmada araç değişkenler yardımıyla iki aşamalı probit en küçük kareler (2SPLS), iki aşamalı koşullu en çok olabilirlik (2SCML) ve üç aşamalı en küçük kareler (3SLS) yöntemleri Stata programı aracılığı ile gerçekleştirilmiştir.

Bulgular: Çalışmaya katılan hastaların yaş ortalaması (standart sapma:SS) 37,3 (11,6) olup; % 61,6'sı kadın, % 58,8'i lise ve altı eğitime sahip, %48,7'si evli, % 72,1'i asgari ücretin üstünde gelire sahiptir. Hastaların ölçek puanlarına ilişkin tanımlayıcı istatistikler değerlendirildiğinde; bireylerin, aile/arkadaş veya özel bir kişiden gördüklerini düşündükleri/algıladıkları puanı ortalaması 53,9 (18,9), benlik saygısı puan ortalaması 22,8 (5,7), çocukluklarında istismar/ihmal nedeniyle yaşadıkları travma şiddetinin ortalaması 44,5 (12,1) olarak bulunmuştur. Araç değişkenler yardımıyla uygulanan 2SPLS ve 2SCML yöntemleri sonucunda; kişilerin hastalık grubuna atanmasında-tanı almasında ÇBASD ve PSP ve Rosenberg değerlerinin anlamlı etkisi olmadığı; 3SLS sonucunda ise benlik saygısı için çocukluk çağı travması ve yaş, algılanan destek için çocukluk çağı travması, bireysel ve sosyal işlevsellikte ise semptom şiddetinin anlamlı değişken olarak belirlenebileceğine işaret etmektedir.

Sonuç: Çocukluk çağında yaşanan istismar şiddeti, benlik saygısı ve destek algısının azalmasına sebep olmakta, bu konuda özellikle eğitimcilerin ve ebeveynlerin bilinçlendirilmesi gerekmektedir. Hastalık durumu ve şiddetinin sosyal işlevselliği etkilemesi nedeniyle bu hastalara gerekli psikolojik desteğin verilmesinin hastanın psikopatoloji açısından iyilik hali üzerinde olumlu etkilere sahip olacağı düşünülmektedir. Teorik altyapısı açıklanan özyinelemesiz modellere özgü yöntemlerden 3SLS, 2SPLS ve 2SCML için sonuçlar elde edilmiş ve oluşturulan diyagram ve ilişki yapısı göz önüne alındığında, sağlık alanında özyinelemesiz modellerin uygulanabileceği öngörülmüştür.

Anahtar Kelimeler: Psikopatoloji, üç aşamalı en küçük kareler, iki aşamalı probit en küçük kareler, iki aşamalı koşullu en çok olabilirlik, yapısal eşitlik modellemesi, özyinelemesiz modeller, araç değişkeni

Diagnosis and Prognosis of COVID-19 Disease Using Routine Blood Values and LogNet Neural Network

¹Mehmet Tahir HUYUT, ²Andrei VELICHKO

¹Erzincan Binali Yıldırım Üniversitesi, Biyoistatistik ve Tıbbi Bilişim A.D, Erzincan

²Petrozavodsk State University, Institute of Physics and Technology, 33 Lenin Str., 185910 Petrozavodsk, Russia

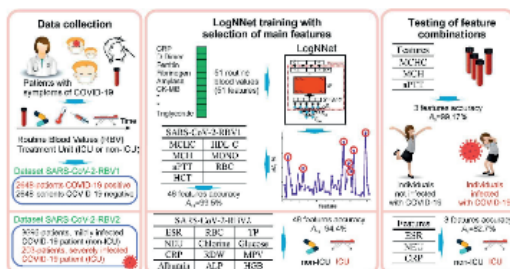
Aim: Since February 2020, the world has embarked on an intense struggle with the COVID-19 disease, and health systems have come under a tragic pressure as the disease has turned into a pandemic. The aim of this study is to determine the most effective routine blood values (RBV) in the diagnosis and prognosis of COVID-19 by using a retrospective feature elimination algorithm for the LogNet reservoir neural network and to evaluate the classification performance of the method.

Materials and Methods: This study consists of two data sets. The first dataset includes data from 2648 patients with positive RT-PCR test results and 2648 patients with negative results. The LogNet neural network is a feedforward network that improves classification accuracy by passing the feature vector through a special reservoir matrix and transforming it into a feature vector of different size. Previous studies have shown that the higher the entropy of a chaotic mapping that fills a reservoir matrix, the better the classification accuracy. Therefore, the procedure for optimizing chaotic map parameters plays an important role in the presented data analysis method using the LogNet neural network. In addition, due to the characteristics of chaotic mapping, RAM usage by a neural network can be significantly reduced.

Results: The LogNet-model achieved 99.5% accuracy in diagnosing the disease with 46 features. Moreover, the LogNet model in diagnosing the disease achieved an accuracy of 99.17 with only three features (mean corpuscular hemoglobin concentration, mean corpuscular hemoglobin and activated partial prothrombin time). The second data set consists of the data of 3899 patients, 203 of whom were severe and 3696 of whom were mild, with a diagnosis of COVID-19 who were treated in hospital. The LogNet model reached 94.4% accuracy in determining the prognosis of the disease with 48 features. Moreover, in predicting the prognosis of the disease, the LogNet model achieved 82.7% accuracy with only three features (erythrocyte sedimentation ratio, neutrophil count, and C reactive protein).

Conclusion: This study provides a fast, reliable, and economic alternative mobile tool for the diagnosis and prognosis of COVID-19 based on the RBV values measured only at the time of admission to the hospital. The results of this study can be effectively used in medical peripheral devices of the IoT (IoT) with low RAM resources, including clinical decision support systems, remote internet medicine, and telemedicine.

Keywords: COVID-19; biochemical and hematological biomarkers; routine blood values; feature selection method; LogNet neural network; Internet of Medical Things; IoT



Work flow chart

The main stages of the study for the diagnosis and prognosis of COVID-19 using the routine blood values: data collection, LogNet training with the selection of main features, testing combinations of the most important features that influence the diagnosis and prognosis of the disease.

Ölçme Araçlarının Boyutluluk Yapısının Belirlenmesinde Yapısal Eşitlik Modellemesine Dayalı Yöntemler

¹Serhat HAYME, ²Derya GÖKMEN, ³Şehim KUTLAY, ⁴Ayşe Adile KÜÇÜKDEVECİ

¹Erzincan Binali Yıldırım Üniversitesi, Biyoistatistik ve Tıbbi Bilişim A.D, Erzincan

²Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik A.D, Ankara

³Ankara Üniversitesi Tıp Fakültesi Fiziksel Tıp ve Rehabilitasyon A.D, Ankara

⁴Ankara Üniversitesi Tıp Fakültesi Fiziksel Tıp ve Rehabilitasyon A.D, Ankara

Amaç: Çok boyutluluk kavramsal olarak iki farklı şekilde tanımlanır: gerçek çok boyutluluk ve psikometrik çok boyutluluk. Gerçek çok boyutluluk, birden fazla boyutu farklı sorularla değerlendirmek için tasarlanmış ölçme araçlarında gözlenen çok boyutluluk türü iken; psikometrik çok boyutluluk, soruların birden fazla boyut ile ilişkili olabilmesi durumudur. Sağlık alanında kullanılan ölçme araçlarında psikometrik çok boyutluluğun her iki kaynağının varlığı, farklı modellerle değerlendirilmiştir.

Gereç ve Yöntem: Psikometrik çok boyutluluğun yapısal olarak iki kaynağı vardır. Bunlardan ilki “Soruların, tek bir yapının mükemmel göstergeleri gibi değerlendirilmesinden kaynaklı yanılabilir doğası”, ikincisi ise “Değerlendirilen yapıların hiyerarşik doğası” şeklindedir. Bu çok boyutluluğun ilk kaynağı doğrulayıcı faktör analizi (DFA) ile açıklayıcı yapısal eşitlik modellemesi (AYEM) çözümlerinin karşılaştırılması ile değerlendirilirken, ikinci kaynağı ise birinci merteye ve hiyerarşik ölçüm modellerinin karşılaştırılması ile belirlenir. Hiyerarşik ölçüm modelleri kapsamında yer alan ikinci merteye modeller ve bifaktör modeller yardımıyla soruların altında yatan global bir yapının olup olmadığı test edilir. Her iki çok boyutluluk kaynağının bir arada olduğu durumlarda hiyerarşik modellerde AYEM modellerinin tercih edildiğini gösteren çalışmalara literatürde yaygın olarak rastlanmaktadır. Veri seti, Romatoid Artridli (RA) 270 hastanın, 109 soruluk ölçme aracına verdikleri cevaplardan oluşmaktadır

Bulgular: Faktör yapısı açıklayıcı yaklaşım altında hem birinci merteye AYEM modeli ile hem de global yapı varlığında Bifaktör AYEM(-B-AYEM) modeli ile incelenmiştir. İlk model kapsamında veri yapısının iki faktörlü bir yapıya uygun olduğu gösterilmiş, faktörler “fiziksel durum” ve “ağrı-bilişsel-psikososyal durum” olarak alan uzmanları tarafından isimlendirilmiştir. Psikometrik çok boyutluluğun ikinci kaynağı bakımından değerlendirme yapıldığında ise, bu faktörler alan uzmanları tarafından global faktör “işlevsellik”, birinci spesifik faktör “kişilerarası ve sosyal etkileşim/ilişki” ve ikinci spesifik faktör “mobilitateyle ilişkili zorluk/ağrı” olacak şekilde tanımlanmıştır.

Sonuç: Bu çalışma kapsamında elde edilen bulgular, psikometrik çok boyutluluğun her iki kaynağının da mevcut olduğunu B-AYEM modeli ile ortaya koymuştur. Daha açık olarak ifade edilirse, bu çalışma ile; AFA, DFA, AYEM ve B-AYEM, psikometrik çok boyutluluğun her iki kaynağının beraber incelenmesine olanak sağlayan genel bir çerçeve sağladığı gösterilmiştir. B-AYEM modelinin özellikle karmaşık çok boyutlu ölçme araçlarıyla çalışıldığında göz önünde bulundurulması gereken bir model olduğu düşünülmektedir.

Anahtar Kelimeler: Açıklayıcı yapısal eşitlik modellemesi, psikometrik çok boyutluluk, bifaktör model

The Interaction Term and Its Graphical Interpretation In Regression Models

¹Handan ANKARALI, ²Özge PASIN, ³Senem GÖNENÇ

¹İstanbul Medeniyet Üniversitesi, Tıp Fakültesi, Temel Tıp Bilimleri, Biyoistatistik Ve Tıp Bilişimi Anabilim Dalı, İstanbul

²Bezmialem Vakıf Üniversitesi, Tıp Fakültesi, Temel Tıp Bilimleri Bölümü, Biyoistatistik Ve Tıp Bilişimi Anabilim Dalı, İstanbul

³Atatürk Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Erzurum

Aim: One of the most important steps in the models to be established in order to define the relationships between the measured characteristics in health field is to correctly define the effects. In this study, the concept of interaction between independent variables in a continuous structure, which is little known and used very rarely in regression-like models, is to be presented with an easily interpretable graphical result.

Materials and Methods: The data used to emphasize the interpretation and importance of the interaction in the regression model were produced by simulation based on the descriptive statistics and distribution patterns of the data in a real study. The data set includes the Systolic blood pressures of 167 people aged between 40 and 76 and a body mass index between 21 and 52. Age and body mass index were defined as independent variables and systolic blood pressure as dependent variables.

Results: In the non-interaction model, an increase in body mass index was increased systolic blood pressure when age was kept constant, and an increase in age increased systolic blood pressure when body mass index was kept constant. Although this result is sufficient, appropriate, and meaningful for the practitioner, it will not make sense without knowing the importance and meaning of the interaction between body mass index and age. When the interaction term was added to the model, the effect of 1-unit change observed in body mass index on systolic blood pressure differs at different ages and the effect of 1-year increase in age on systolic blood pressure differs according to body mass index values. In this case, it has emerged that a physician who will make a clinical decision should also consider age when deciding on systolic blood pressure according to body mass index.

Conclusion: Faulty model design means producing faulty results. The modeling process in healthcare research involving complex relationships requires substantial knowledge, domain knowledge, modeling knowledge, and accurate interpretation of results. Examining the interaction terms is of great importance in the modeling process. If this effect is significant, the actual effects of the interacting effects are meaningless and their interpretation will yield erroneous results. The contour graphic method, which will facilitate the work of the practitioners in the interpretation of the interaction, will make a significant contribution to the evaluation of this term in models.

Keywords: Regression model, interaction, contour chart, independent variable

BKSY: Bilgi Keşfi Süreci Yazılımı

¹Ahmet Kadir ARSLAN, ²Cemil ÇOLAK

¹İnönü Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Malatya

²İnönü Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Malatya

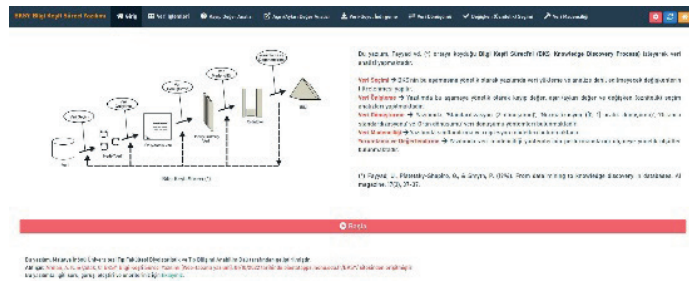
Amaç: Bu çalışmada, veri bilimi süreçlerine yönelik olarak bilgi keşfi süreci (BKS, Knowledge Discovery Process) tabanlı, uçtan uca veri analizi yapan kapsamlı bir web-tabanlı aracın geliştirilmesi amaçlanmıştır.

Gereç ve Yöntem: BKS, üzerinde çalışılacak veri setinde saklı olarak bulunan bilgi, örüntü ve özellikleri ortaya çıkarmak için uygulanan ve beş adımdan oluşan bir süreçtir. BKS, analize konu veri setinin ilgili veri tabanlarından temin edilmesi, veri önışlemesi, veri dönüşümü, değişken seçimi, veri madenciliği ve veri madenciliği çıktıların değerlendirilmesi ve yorumlanması aşamalarından oluşur. BKS temelli geliştirilen yazılımda, analize konu veri setinin yazılıma yüklemesi yapılır. Veri önışleme aşamasında, veri setindeki kayıp değerlere çeşitli yöntemler kullanılarak değer ataması yapılır, çeşitli istatistiksel teknikler kullanılarak uç değerlerin tespiti ve veri setinden çıkarılması işlemi gerçekleştirilir. Veri dönüşümü aşamasında veri setine standardizasyon (z-dönüşüm) ve diğer normalizasyon dönüşümleri uygulanabilmektedir. Değişken seçimi aşamasında, Boruta veya LASSO regresyon tekniklerinden biri kullanılarak veri setindeki değişken sayısı en uygun sayıya indirgeme yapılır. Veri madenciliği aşamasında ise sınıflandırma ve regresyon görevlerinde kullanılan çeşitli makine öğrenmesi modelleri kullanılarak veri setindeki potansiyel desen ve örüntülerden bilgi çıkarma işlemi gerçekleştirilir. Yorumlama ve değerlendirme aşamasında ise veri madenciliği çıktıları yorumlanır ve modelin performansı çeşitli ölçütlerle ölçülür.

Bulgular: Geliştirilen yazılımda, kullanılan makine öğrenmesi yönteminin amacına göre (sınıflandırma veya regresyon) bulgular oluşturulmaktadır. Sınıflandırma problemlerinde genel doğruluk, duyarlılık, seçicilik, ROC eğrisi altına kalan alan vb. bulgular elde edilebilirken, regresyon problemlerinde ise ortalama karesel hata, hata kareler ortalamasının karekökü ve ortalama mutlak hata gibi bulgular sunulmaktadır. Ayrıca modellere ilişkin değişken önemlilik skorları da geliştirilen yazılımda raporlanmaktadır.

Sonuç: Geliştirilen yazılım ile, • Tıp, biyoloji ve diğer alanlara yönelik çeşitli dosya uzantılarına sahip (.xlsx, .csv, .txt, .arff, .sav, vb.) veri dosyalarının analizine imkân tanınması, • Yüklenen veri setlerinin kapsamlı önışlemesi, • Önışlemesi yapılan veri setinden belirli bir bağımlı değişkene (hastalık vb.) göre tahmin sonucu üretmesi, • Sonuç/çıktı değişkeni ile ilişkili olabilecek faktörleri modelleme öncesinde tespit etmesi, • Yapılan tahmin sonuçlarının açıklanabilmesi sayesinde ilgili veri setindeki hastalık risk faktörlerinin/biyobelirteçlerin tahmini ve değişken önemliliklerinin (variable importance) belirlenmesi, • Açık erişimli ve web-tabanlı olması nedeniyle tüm kişisel bilgisayarlarda kullanılabilir olması, • Sınıflandırma ve regresyon problemlerine uygulanabilmesi gibi özgün özelliklerinin olması nedeniyle veri bilimi ve ilişkili alanlarda çalışan araştırmacılar için faydalı bir analiz aracı olması beklenmektedir.

Anahtar Kelimeler: Bilgi keşfi süreci, shiny, web-tabanlı yazılım



Şekil-1. Keşfi Süreci Yazılımı'nın genel görünümü

KOAH Hastalarında Fonksiyonel Kapasite Üzerindeki Etkilerin İstatistiksel Yöntemlerle Analizi

¹Sinan SARAÇLI, ²Aydın BALCI, ³Berkalp TUNCA

¹Balıkesir Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim ABD

²Afyon Sağlık Bilimleri Üniversitesi Tıp Fakültesi Göğüs Hastalıkları ABD

³Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü İstatistik ABD

Amaç: Bu çalışmanın amacı, KOAH Hastalarında Fonksiyonel Kapasite Üzerindeki Etkili olan faktörleri İstatistiksel Yöntemlerle Analiz etmektir.

Gereç ve Yöntem: Çalışmada Afyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi Hastanesi göğüs hastalıkları polikliniğine Haziran 2021 - Haziran 2022 tarihleri arasında başvuran 120 Amfizem fenotipli KOAH hastasının fonksiyonel kapasitelerini ölçmek için 6 dakika yürüme testi, 1 dakika otur kalk testi ve 5 tekrarlı otur kalk testi uygulanarak ilgili veriler derlenmiştir. İlgili verilere ilişkin betimleyici istatistikler verildikten sonra KOAH hastalarının fonksiyonel kapasiteleri üzerinde etkili olan etmenler CRT analizi ve PLS analizi ile incelenmiştir.

Bulgular: Her bir fonksiyonel kapasite ölçüm testi için uygulanan CRT analizi sonucunda 6 dakika testine bağlı fonksiyonel kapasite üzerinde BORG indeksi, 1 dakika otur kalk testine bağlı fonksiyonel kapasite üzerinde nabız değerleri, 5 dakika otur kalk testine bağlı fonksiyonel kapasite üzerinde ise saturasyon değerleri en etkili değişkenler olarak belirlenmiştir. Tüm fonksiyonel kapasite testlerine bağlı olarak kurulan PLS modeli üzerinde ise BORG indeksi, diğer değişkenlere göre en etkili değişken olarak belirlenmiştir.

Sonuç: KOAH hastalarının 6 dakika yürüme testi sonucunda, BORG indeksi için kritik değer 4,5 olarak belirlenmiş, BORG indeksi 4,5'in altında olan hastaların fonksiyonel kapasitesi diğerlerine göre daha yüksek olarak belirlenirken, BORG indeksi 4,5'in üzerinde olan hastalar için saturasyon değeri anlamlı olarak bulunmuş, saturasyonu 89'un üzerinde olan hastaların fonksiyonel kapasitesinin diğerlerine göre yüksek olduğu belirlenmiştir. 1 dakika otur kalk testi sonucunda, nabız değeri 139,5'in üzerinde kalanların fonksiyonel kapasitesi diğerlerine göre daha yüksek olarak belirlenmiş, nabız değeri 139,5'in altında kalan nabız değerleri üzerinde BORG indeksi etkili bulunmuş ve BORG indeks değeri 3,5'in altında kalan hastaların fonksiyonel kapasitesinin daha yüksek olduğu belirlenmiştir. 5 dakika otur kalk testi sonucunda, saturasyon değişkeni en etkili değişken olarak belirlenmiş, saturasyon değeri 89,5'in altında olan hastaların fonksiyonel kapasitelerinin diğerlerine göre daha yüksek olduğu belirlenmiştir. Tüm fonksiyonel kapasite testlerine bağlı olarak kurulan PLS modeli sonucunda ise, BORG indeksindeki 1 birimlik artışın fonksiyonel kapasite üzerinde 0,42'lik katsayı ile azalışa sahip olan en etkili değişken olduğu belirlenmiştir.

Anahtar Kelimeler: Koah, fonksiyonel kapasite, CRT, PLS.

Türkiye Covid-19 Verileri için Parçalı Fonksiyona Dayalı Büyüme Modelleri

¹Leyla BAKACAK KARABENLİ, ²Serpil AKTAŞ ALTUNAY

¹Hacettepe Üniversitesi, İstatistik Bölümü, Ankara

²Hacettepe Üniversitesi, İstatistik Bölümü, Ankara

Amaç: Türkiye'nin Covid-19 verilerindeki ölüm sayısı belli bir trend içinde olmayıp dalgalı, sürekli artan ve azalan bir trend göstermekte ve homojen olmayan bir dağılım sergilemektedir. Bu nedenle, tek bir büyüme modeli üzerinden ölüm sayısını tahmin etmek zorlaşmaktadır. Bu çalışmada, çeşitli büyüme modelleri dikkate alınarak Covid-19 verileri üzerinden ölüm sayılarını tahmin etmek amaçlanmıştır.

Gereç ve Yöntem: Büyüme modellerinde tek bir model kullanarak ölüm sayısını tahmin etmek yerine, dalgalanmaları da dikkate alan parçalı bir fonksiyon tanımlanmıştır. Bu fonksiyonlar log-üstel ve log-Gompertz modeller ile tanımlanmıştır. Parçalı fonksiyonların belirlenebilmesi için değişim noktaları tahmin edilmiştir. Böylece parçalı büyüme fonksiyonları sayesinde veri seti modellenmiştir. Değişim noktalarının tahmininde bilgi kriterlerinden yararlanılmıştır.

Bulgular: İncelenen zaman aralığında farklı eğilimlere sahip Covid-19 verileri, değişim noktalarına göre farklı büyüme fonksiyonları ile temsil edilmiştir. Bu yöntem tahminlerde daha fazla esneklik sağlamış ve daha iyi tahmin eğrilerinin elde edilmesine olanak vermiştir. 17/03/2020-21/03/2021 tarihleri arasındaki birbirinden farklı eğilimlere sahip kümülatif ölüm sayıları, beş parçalı log-Gompertz modeli ile temsil edilmiştir.

Sonuç: Bu çalışmada ölüm sayıları incelenmiş ve verilerdeki dalgalanmalardan dolayı ölüm sayılarını tahmin etmek için tek bir modelin yeterli olmayacağı düşünülmüştür. Bu nedenle verilerin değişim noktaları araştırılmış ve bu noktalara göre en uygun büyüme modelleri seçilmiştir.

Anahtar Kelimeler: Büyüme modelleri, parçalı fonksiyon, Türkiye Covid-19 verileri

Recurrent Neural Network for Complex Survival Problems

¹Pius MARTIN, ²Nihal ATA TUTKUN

¹Hacettepe university

²Hacettepe university

Aim: Survival analysis has become one of the paramount procedures in the modeling of time-to-event data. When we encounter complex survival problems, the traditional approach remains limited in accounting for the complex correlational structure between the covariates and the outcome due to the strong assumptions that limit the inference and prediction ability of the resulting models. There exist several studies on the deep learning approach to survival modeling nevertheless, the application for the case of complex survival problem remain limited. In addition, the existing models do not fully address the complexity of the data structure and are subject to noise and redundancy information. In this study, we design a deep learning technique (CmpXRnnSurv_AE) that obliterates the limitations imposed by traditional approaches and addresses the above issues to jointly predict the risk-specific probabilities and survival function for recurrent events with competing risks.

Materials and Methods: We introduce the component termed Risks Information Weights (RIW) as an attention mechanism to compute the weighted cumulative incidence function (WCIF) and an external auto-encoder (ExternalAE) as a feature selector to extract complex characteristics among the set of covariates responsible for the cause-specific events.

Results: We train our model using synthetic and real data sets. For the performance evaluation, we compute the time-dependent concordant index for complex survival models. We select both traditional and machine learning models as benchmarks. Our model demonstrates better performance across all datasets with the maximum concordance indices of 0.86, 0.89, and 0.92 on the synthetic, MIMIC-III clinical, and the UNOS-OPTN(KIDPAN) datasets accordingly.

Conclusion: We developed a novel deep learning model termed CmpXRnnSurv-AE for the complex survival problem. Using the external autoencoder we addressed the noise invariant problem in the complex survival data. By computing the weighted cumulative incidence function (WCIF) we managed the complex correlation structure in recurrent events with competing risks survival data. Since deep learning models are overparameterized and deterministic in nature, the point estimates provided are not reliable for uncertainty measurements in the predictions. In addition, the RNN is subjective to temporal inconsistency. In future work, we propose to apply the Historical Consistent Neural Networks (HCNN) in the Bayesian framework to quantify uncertainties in the model's predictions for the case of complex survival problems.

Keywords: Cumulative incidence function (CIF) Risk Information Weight (RIW), Autoencoders (AE), Survival analysis, Recurrent events with competing risks, Recurrent neural networks (RNN), Long short term memory (LSTM), Self attention, Multilayers perceptrons.

KS – 1

Önceden Eğitilmiş Evrişimsel Sinir Ağlarından Elde Edilen Öznitelikler ile Sperm Morfolojisinin Julia Tabanlı Web Uygulaması ile Sınıflandırılması

¹Hüseyin KUTLU, ²Emek GÜLDOĞAN, ³Cemil ÇOLAK

¹Adıyaman Üniversitesi

²Adıyaman Üniversitesi

³Adıyaman Üniversitesi

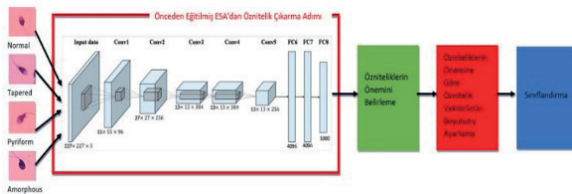
Amaç: Derin öğrenme yöntemleri, eğitim esnasında, binlerce, hatta milyonlarca parametre güncellemesi yaparlar. Dolayısıyla büyük miktarda eğitim verisine ihtiyaç duyarlar. Aksi bir durumda test veri kümesinde düşük doğruluklu, örneklem dışı uygulamada yanlış sonuçlar veren bir model ortaya çıktığı görülür. Tıbbi imgeler gerek yeteri kadar etiketlendirilmiş veri olmayışı, gerek negatif örneğin pozitif örneğe göre düşük oranda oluşu sebebi ile bu durumdan mustarıdır. İmge işlemede bu tür problemler için başka bir veri seti ile önceden eğitilmiş bir evrişimsel sinir ağı (ESA) kullanılmaktadır. Bu çalışmada erkek infertilitesinin teşhis ve tedavisinde önemli bir yere sahip olan sperm baş morfolojisi imgelerinin önceden eğitilmiş ESA mimarilerinden elde edilen özniteliklerle sınıflandırması amaçlanmıştır.

Gereç ve Yöntem: Çalışmada halka açık bir veri kümesi olan HuSHem sperm baş morfolojisi veri setinden elde edilen imgeler kullanılmıştır. İmgelere herhangi bir ön işleme ve çoğaltma uygulanmadan, imgeler önceden eğitilmiş ESA mimarilerinden (squeezeNet, googLeNet, inceptionv3, densenet201, mobilnetv2, resnet18, resnet50, resnet101, xception, inceptionresnetv2, shufflenet, nesnetmobile, nasnetlarge, darknet19, efficientnet, alexnet, vgg16, vgg19) geçirilmiş ESA'ların son katmanlarından imgelerin öznitelikleri alınmıştır. Her imgeden elde edilen 1000 adet özneliğin önem değerleri Lojistik Regresyon katsayıları temelli öznelik önemi belirleme yöntemi, Ağaç tabanlı XGBoost algoritması gini indeksi temelli öznelik önemi belirleme yöntemi ve Temel Bileşenler Analizi (Principal Component Analysis – PCA) öznelik yük değeri (factor loading) temelli öznelik önemi belirleme yöntemi ile hesaplanmıştır. Öznitelikler öznelik önemi belirleme yöntemi ile indirgenerek sınıflandırılmıştır.

Bulgular: Önceden eğitilmiş ESA modellerinden sperm morfolojik imgelerinin en etkili öznitelikleri GoogLeNet ESA mimarisinden elde edilmiştir. Öznelik önem değerlerine göre öz nitelikler önem sırasına göre sıralanmış ve en önemli öznelikten başlayarak bir artırım- la öznelik alt kümeleri sınıflandırılmıştır. Sınıflandırmada İnönü Üniversitesi Biyoistatistik Anabilim dalının Geliştirdiği Julia tabanlı web uygulaması kullanılmıştır. Lojistik regresyon katsayıları ile elde edilen öznelikler 161 öznelikte en yüksek (%81.01) sınıflandırma doğruluğuna ulaşmıştır. XGBoost algoritmasından elde edilen öznelik değerleri 248 öznelikte en yüksek (%79.13) sınıflandırma doğruluğuna ulaşmıştır. PCA faktör yük değerlerinden (elde edilen öznelik önemleri 403 öznelikte en yüksek sınıflandırma doğruluğuna (%79.58) ulaşmıştır.

Sonuç: İmge sayısının yetersiz olduğu problemlerde derin öğrenmenin avantajlarından faydalanmak için Önceden eğitilmiş ESA mimarilerinin kullanılabilir bir yöntem olduğu görülmüştür. Derin özneliklerin önem değerini belirlemede Lojistik regresyon katsayıları diğer yöntemlere göre öne çıkmıştır.

Anahtar Kelimeler: Derin Öznelik Çıkarımı, Önceden eğitilmiş ESA'lar, Öznelik Önemi, Julia, Sınıflandırıcı



Sperm Önerilen Yöntemin Adımlarını Gösteren Diyagram

Lökosit Antijen Molekülleri Toplum Doku Tipi Sıklığı

¹Kemal OLÇA, ²Ahmet KESKİNOĞLU, ³Duygu KARAKAYA, ⁴Ebru ÖZTÜRK, ⁵Pembe KESKİNOĞLU, ⁶Mehmet ORMAN, ⁷Timur KÖSE, ⁸Ahmet DİRİCAN, ⁹Erdem KARABULUT

¹İstanbul Üniversitesi Lisansüstü Eğitim Enstitüsü, Cerrahpaşa Tıp Fakültesi Biyoistatistik ve Tıp Bilişim AD, İstanbul

²Ege Üniversitesi Tıp Fakültesi, Çocuk Sağlığı Nefroloji AD, İzmir

³Dokuz Eylül Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD, İzmir

⁴Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik AD, Ankara

⁵Dokuz Eylül Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD, İzmir

⁶Ege Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD, İzmir

⁷Ege Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD, İzmir

⁸İstanbul Üniversitesi Lisansüstü Eğitim Enstitüsü, Cerrahpaşa Tıp Fakültesi Biyoistatistik ve Tıp Bilişim AD, İstanbul

⁹Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik AD, Ankara

Amaç: Lökosit antijen molekülleri (human leukocyte Antigen-HLA) en çeşitli ve polimorfik molekülerdendir. HLA molekülleri, kromozom 6'nın kısa kolunda bulunur, temel işlevi, yabancı cisimlerin tanınması ve immünojenisitedir. Yapısal ve fonksiyonel özelliklerine göre 3 sınıfa ayrılır, sınıf 1: HLA-A, -B, ve sınıf 2; HLA-DR, -DQ antijenleri önemlidir. Toplumlara göre HLA sıklıkları değişmektedir. HLA doku tipi sıklıkları, antropolojik olarak değerlendirmeler dışında hastalıklara yatkınlık ve organ naklindeki rolü de önemli araştırma konusudur. Ülkemizde hastalıklara yatkınlık ile ilgili sınırlı sayıda hasta popülasyonlarında HLA doku tipleri çalışılmıştır, ayrıca donörler üzerinden sağlıklı kişileri içeren HLA sıklıklarının da araştırıldığı görülmektedir. Ancak bu çalışmalarda toplam "n" sayısı çok sınırlıdır. Bu çalışmada, ülkemizin her yerinden hasta ve yakınlarının başvurduğu bir üniversite hastanesi doku tipi laboratuvarında donörlerin HLA doku tipi sıklıkları değerlendirilmiştir.

Gereç ve Yöntem: Retrospektif olarak, ülkemizin batısında ve organ nakli çok fazla gerçekleştiren donör sayısı çok yüksek olan üniversite hastanesi doku tipi laboratuvarının 40 yıllık verisi taranmıştır. Bu verilerim çoğunluğuna elektronik ortamda düz metin şeklinde elde edilmiştir. Bu metin raporlar kod yazılarak HLA-A, B ve DR için ikişer lokus bilgisi ayrılmış ve veri tabanına ayrı 6 sütun olarak kaydedilmiştir. Her üç HLA doku tipi için allel frekansları belirlendi ve Hardy Weinderg eşitliğime (HWE) uyumları da test edildi.

Bulgular: Doku tipi laboratuvar verilerinde toplam 17130 donörün HLA bilgisine ulaşıldı. HLA-A doku tipleri en sık A*02(%22), A*24 (%16,4), en sık HLA-B için B-35 (%17,4) ve HLA-DR için DRB1*11 (%22,6) ve DRB1-04 (%14) olarak saptandı.

Sonuç: Yüksek sayıdaki donör HLA verilerinden A, B ve DR için allel sıklıkları toplum doku sıklığı olarak değerlendirilebilir. Göçler, epigenetik etkilenimler vb. sonuçlar nedeni ile çok polimorfik olduğu bilinen HLA tipleri belirli aralarla rutin hizmet verilerinden sıklıklarının belirlenmesi bu doku tiplerinin önemli olduğu pek çok hastalık için gerekli olduğu düşünülmüştür. Teşekkür: Bu çalışma 318S166 nolu TÜBİTAK-SBAG projesi verilerinden gerçekleştirilmiştir.

Anahtar Kelimeler: Lökosit Antijen Molekülleri, HLA, Doku Tipi, Hardy Weinderg Eşitliği, HWE

KS – 3

Mamografi görüntülerinde K-ortalamlar kümeleme ve Canny kenar algılama bölütleme yöntemlerinin sınıflandırma performanslarının karşılaştırılması¹Jale KARAKAYA, ²Hanife AVCI¹Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara²Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Amaç: Bu çalışmanın amacı, mamografi görüntülerini çeşitli ön işleme yöntemleriyle iyileştirdikten sonra, görüntülerden çıkarılan özneliklerin hastalık durumunu belirlemede iki farklı bölütleme yönteminin başarısını araştırmaktır.

Gereç ve Yöntem: Çalışmada açık erişimli mini-MIAS veri tabanı kullanıldı. Bu veri seti, 161 hastanın sağ ve sol meme görüntülerini içeren 322 sayısallaştırılmış mamografi görüntüsünden oluşmaktadır. Çalışmada ön işleme yöntemleri olarak sırasıyla medyan filtre, kontrast sınırlı uyarlanabilir histogram eşitleme (CLAHE) ve keskin olmayan maskeleme görüntü iyileştirme teknikleri kullanılmıştır. Ön işleme yöntemleriyle etiketlerden ve yapay çizgilerden arındırılmış mamografi görüntüleri için bir sonraki önemli adım, uygun segmentasyon yöntemiyle görüntülerden pektoral kas vb. temizleyerek ilgi bölgesi (region of interest, ROI) belirlemektir. Bu çalışmada, ön işleme yöntemlerinin performansını incelemek için k-ortalamlar kümeleme algoritması seçilmiştir. Daha sonra, ikinci bir bölütleme yönteminin görüntülerden çıkarılan öznelikler üzerindeki etkisini görmek için mamografi görüntülerine kenar tabanlı görüntü bölütleme yöntemlerinden biri olan Canny kenar algılama algoritması uygulanmıştır. Mamografi görüntülerine görüntü ön işleme ve bölütleme işlemleri uygulandıktan sonra özellik çıkarımı yapılmıştır. Gri seviye eş oluşum matrisi (GLCM) ve Gri seviye çalışma uzunluğu matrisi (GRLM) teknikleri kullanılarak ROI örneklerinden özellikler çıkarılmıştır. Korelasyon matrisi ile öznelik seçimi yapıldıktan sonra bu özneliklerin sınıflandırma başarısına etkileri 6 farklı klasik veri madenciliği yöntemi ile incelenmiştir.

Bulgular: Normal-anormal doku sınıflandırması için Canny kenar algılama algoritması, tüm sınıflandırma yöntemlerinde k-ortalamlar kümeleme algoritmasından daha düşük performans göstermiştir. Bu sonuçlara göre doğruluk, duyarlılık, eğri altında kalan alan (AUC) gibi ölçülere baktığımızda normal-anormal sınıflandırmasında Destek Vektör Makinesi, Rastgele Orman, Yapay Sinir Ağı ve Naive Bayes sınıflandırma yöntemlerinin performansları k-en yakın komşuluk ve Karar Ağacı göre daha başarılıdır.

Sonuç: Bu çalışmada, kenar tabanlı görüntü bölütleme tekniği olan Canny kenar algılama yöntemi ile bölge tabanlı görüntü bölütleme tekniği olan k-ortalamlar kümeleme yönteminin sınıflandırma performansları karşılaştırılmıştır. K-ortalamlar kümeleme algoritmasının Canny kenar algılama algoritmasına göre daha yüksek bir sınıflandırma özelliğine sahip olduğu tespit edilmiştir.

Anahtar Kelimeler: Mamografi görüntüleri, bölütleme yöntemleri, sınıflandırma performansı.

Log Multinomial Regresyon Modeline Hosmer-Lemeshow Uyum İyiliği Testinin Uyarlanması

¹Yasemin ÖZTÜRK, ²Erdem KARABULUT

¹T.C. Sosyal Güvenlik Kurumu, Genel Sağlık Sigortası Genel Müdürlüğü, Ankara

²Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Amaç: İki durumlu veya multinomial lojistik regresyon modelinin uyumunun değerlendirilmesinde Hosmer ve Lemeshow tarafından önerilen uyum iyiliği testi kullanılmaktadır. İleriye yönelik çalışmalarda göreceli risk kestirimi elde edilebildiğinden, iki kategorili bağımlı değişken olduğunda log-binomial ve 3 ya da daha fazla kategorili bağımlı değişken olduğunda log-multinomial regresyon modelinin kullanılması önerilmektedir. Ancak, log-multinomial regresyon modelinin veriye uyumunun değerlendirilmesine yönelik herhangi bir metod bulunmamaktadır. Bu çalışmada Hosmer Lemeshow uyum iyiliği testi log multinomial regresyon modeline uyarlanmış ve benzetim çalışması ile performansı değerlendirilmiştir.

Gereç ve Yöntem: Uyarlanan Hosmer Lemeshow uyum iyiliği testinin uygulanabilirliği çeşitli senaryolar altında benzetim çalışması ile incelenmiştir. Söz konusu senaryolarda bağımlı ve bağımsız değişkenlerin farklı yapıları, çeşitli örneklem genişliklerinde benzetim çalışması yardımıyla karşılaştırılmıştır. Benzetim çalışmasında, ikiden fazla kategorili bağımlı kategorik değişkenler, sürekli ya da kategorik bağımsız değişkenler ve farklı örneklem genişlikleri (200, 400, 600, 800 ve 1000) dikkate alınarak toplamda 70 farklı senaryo kullanılmıştır. Benzetim senaryolarındaki modellerde kullanılan regresyon katsayıları, kullanılan gerçek veri setlerinde yer alan değişkenler yardımı ile kestirilen log-multinomial modellerden elde edilmiştir. Her bir senaryo 1000 kez tekrar edilmiş ve her bir tekrarda uyum iyiliği testi sonuçları tip I ve tip II hata sıklıkları elde edilerek değerlendirilmiştir. Belirlenen senaryolar için benzetim çalışması, modelleme ve uyum iyiliği testinin sonuçlarının incelenmesi R yazılımında yer alan “RStata”, “general hoslem” ve “MASS” paketleri ve STATA yazılımı kullanılarak gerçekleştirilmiştir.

Bulgular: Uyarlanan Hosmer-Lemeshow testi, sürekli sayısal bağımsız değişkene ilişkin karesel terim ve bağımsız değişkenler arasındaki etkileşim teriminin dahil edilmemesi gibi çeşitli uyumsuzlukları belirleyebildiği görülmektedir. Tüm senaryolarda örneklem büyüklüğü arttıkça tip I ve tip II hata oranının düştüğü ve çoğunlukla nominal düzeyin altında olduğu sonucu elde edilmiştir.

Sonuç: Hosmer-Lemeshow uyum iyiliği testinin log multinomial regresyon modelinin uyum iyiliğinin tespit edilmesinde kullanılabilir yapıda olduğu görülmektedir.

Anahtar Kelimeler: Log multinomial regresyon, Hosmer Lemeshow test istatistiği, uyum iyiliği.

KS – 5

Sobel ve Baron-Kenny Yöntemlerinin Performanslarının Karşılaştırılması

¹Hülya BİNOKAY, ²Yaşar SERTDEMİR

¹Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana

²Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana

Amaç: Bu çalışmada, Sobel ve nedensel adım yaklaşımı yöntemlerinin farklı koşullar altında performanslarının karşılaştırılması amaçlanmıştır. Ayrıca literatürden alınan gerçek bir veri setinde sonuçlar değerlendirilecektir

Gereç ve Yöntem: Farklı örnek büyüklüğü (100,150,200,250,300,350,400) ve $PM=(\text{dolaylı etki})/(\text{toplam etki})=(a*b)/c$ (aracı değişken oranı)(0.05,0.10,0.15,0.20,0.25) olduğu durumlar için normal dağılımdan veri üretilmiştir. $Y=\beta_0+cX+e_1$ $M=\beta_0'+aX+e_2$ $Y=5+\beta_0''+c'X+bM+e_3$ Farklı örnek büyüklüğü ve aracı değişken oranı için 4 farklı senaryo ele alınmıştır. Senaryo 1: sadece a katsayısı anlamlı, Senaryo 2: sadece b katsayısı anlamlı, Senaryo 3: a ve b katsayılarının ikisi de anlamlı değil, Senaryo 4: a ve b katsayılarının ikisi de anlamlı

Bulgular: Aracı değişken oranları için, senaryo 3 te, M değişkeninin, nedensel adım yaklaşımı ve Sobel yöntemlerine göre aracı değişken olmadığı gözlenmiştir. Senaryo 1, 2 ve 4 te, Sobel yöntemine göre M aracı değişken olduğu kararı verilirken, nedensel adım yaklaşımı yöntemine göre aracı değişken olmadığı kararı verilmektedir. Örnek büyüklüğü arttıkça, M aracı değişkeninin sobel yöntemine göre aracı değişken olduğu, nedensel adım yaklaşımı yöntemine göre aracı değişken olmadığı gözlenmiştir.

Sonuç: Farklı senaryolar incelendiğinde, Sobel yöntemine göre aracı değişken olan M değişkeni, nedensel adım yaklaşımına göre aracı değişken değildir. Böyle bir durumda bir değişkenin aracı değişken olup olmadığına karar vermek net değildir. Bundan dolayı yeni bir yöntemin önerilmesi gerekmektedir.

Anahtar Kelimeler: Aracılık, Sobel, Baron-Kenny

KS – 6

Lojistik Regresyon Modeli ile Kardiyovasküler Kalp Hastalığı Sınıflandırılması ve Tahmini

Mehmet KIVRAK

Recep Tayyip Erdoğan Üniversitesi Biyoistatistik ve Tıp Bilişimi

Amaç: Kardiyovasküler hastalıklar dünya çapında önde gelen ölüm nedenidir. Bu nedenle kalp krizlerini etkili bir şekilde öngörmek için nedenlerin belirlenmesi ve bir sistem geliştirilmesi gerekmektedir. Bu çalışma, lojistik regresyon yöntemi ile açık erişimli kardiyovasküler kalp rahatsızlığı verilerini sınıflandırmayı ve etkili faktörleri tahminlemeyi amaçlamaktadır.

Gereç ve Yöntem: Çalışmada açık erişimli kardiyo-kalp veri seti kullanıldı. Veri seti, yaş, cinsiyet, göğüs ağrı tipi, dinlenme anındaki kan basıncı, serum kolesterol düzeyi, açlık kan şekeri, elektrokardiyografi sonuçları, kalp atış hızı, egzersize bağlı göğüs ağrısı, ST segmentinin yükselmesi, floroskopi ile boyalı damar sayısı, talasemi gibi risk faktörlerinden oluşmaktadır. Hastalığı sınıflandırmak için lojistik regresyon modeli kullanıldı. Model performansı için doğruluk, duyarlılık, seçicilik, pozitif tahmin değeri ve negatif tahmin değeri performans metrikleri değerlendirildi. Model anlamlılığı için Hosmer-Lemeshow testi, katsayıların değerlendirilmesinde Odds oranı ve uyum iyiliği için Cox-Snell R2 değerleri incelendi.

Bulgular: Deneysel bulgulara göre model anlamlılığı sağlanırken ($p=0.03$) cp2, cp3'ün hastalığı artırıcı, cinsiyet1, ca1, ca2, ca3'ün ise hastalığı azaltıcı etkisi bulunmaktadır. Y aş, cp1, trestbps, kolesterol, fbs1, restecg1, restecg2, thalach, exang1, oldpeak, slope1, slope2, ca4, thal1, thal2, thal3 değişkenlerinin ise hastalığa anlamlı etkisi görülmedi. Cox-Snell R2 değeri (bağımsız değişkenlerin bağımlı değişkendeki değişimi açıklama yüzdesi) ise yeterli düzeydedir ($R^2=0.544$). Modelin sınıflandırma performansına bakıldığında, 5 katlı çapraz geçerlilik ile parametre optimizasyonu yapıldı (Grid search taramalı yüzeysel). Modelleme sonuçlarından elde edilen doğruluk, duyarlılık, özgüllük, pozitif prediktif değer, negatif prediktif değeri sırasıyla % 88.4, % 88.0, % 83.3, % 92.7, % 90.6 ve % 86.9 olarak bulundu.

Sonuç: Bu araştırmanın bir sonucu olarak, lojistik regresyon yöntemi ile açık erişimli kardiyovasküler kalp rahatsızlığı verilerini sınıflandırma ve etkili faktörleri tahminleme yapıldı. Gelecekte, bu tarz yöntemlerin benzer çalışmalar ile güvenilirliği doğrulanarak bu yöntemlere dayalı terapötik prosedürler oluşturulabilir ve klinik pratikteki yararları belgelenebilir.

Anahtar Kelimeler: Kardiyovasküler Kalp Hastalığı, Sınıflandırma, Tahmin, Lojistik Regresyon.

Çok Boyutlu Ölçekleme Yöntemi ile Trakya'da Tarımsal Üretim Bölgelerinde Ağır Metal Yoğunluğunun Gösterimi

¹Büşra AYDIN, ²Özgecan KORKMAZ AĞAOĞLU, ³İsmayıl SAFA GURCAN

¹Ankara Üniversitesi Veteriner Fakültesi, Biyoistatistik Anabilim Dalı

²Ankara Üniversitesi Veteriner Fakültesi, Biyoistatistik Anabilim Dalı

³Ankara Üniversitesi Veteriner Fakültesi, Biyoistatistik Anabilim Dalı

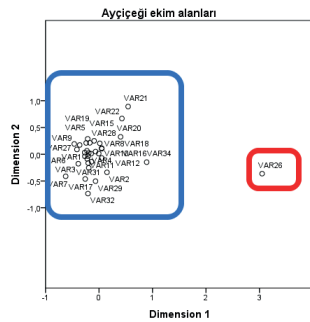
Amaç: Bu çalışma ile havzada tarım yapılan farklı bölgelerden alınan Ayçiçek, Buğday ve Çeltik örneklerinden ağır metal değerlerine göre, bölgeler arası ağır metal yükü bakımından meydana gelen kümelenme, X-Y düzlemi üzerinde incelenmiştir.

Gereç ve Yöntem: Tanımlayıcı istatistikler kullanılarak ağır metal yükleri hakkında bilgiler ışığında uzaklık matrisleri Euclidian uzaklığı kullanılarak çok boyutlu ölçekleme ile uyum iyiliği ve stress değerleri hesaplanarak düzlem üzerinde bölgelerin işaretlemesi yapılmıştır.

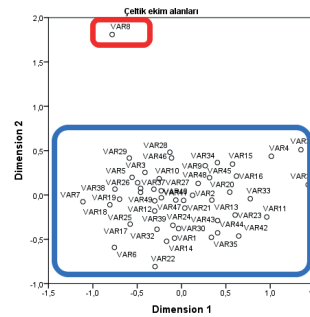
Bulgular: Varyasyon katsayısı yüksek bulunan ağır metaller Buğday için; sırası ile Alüminyum (%188), Krom (%162), Arsenik (%149), Kurşun (%138), Kobalt (%117) ve Demir (%115) ve Nikel (%92), Ayçiçeği için; Alüminyum (%105), Kurşun (%69), Arsenik (%66) ve Kalay (%56), Çeltik için Nikel (%282), ve Kadmium (%65) olmuştur. Çok Boyutlu Ölçekleme yöntemi ile Buğday için yapılan hesaplamalar sonucunda elde edilen koordinat noktalarına göre Yeniköy ve Merkez ve Bayramlı tarım bölgeleri diğer tarım bölgelerinden uzak kalmıştır. Ayçiçeği tarımı yapılan alanlarda Kırcaali bölgesi düzlem üzerinde ana kümeden uzakta yer almıştır. Çeltik ekimi yapılan bölgelerde; Uzunköprü Merkez bölgesi küme oluşturan diğer bölgelerden uzak konumlanmıştır.

Sonuç: Genel olarak sanayi atıklarının yer aldığı havzada belirli dönemlerde nehir taşması yaşanan dönemlerde sulak alan ve yakınlarında tarımsal bölgelerde üretimi yapılan Buğday, Ayçiçeği ve Çeltikte saptanan ağır metal düzeylerinin, iç kesimlerde tarım üretimi yapılan bölgelere göre daha yüksek düzeyde olduğu söylenebilir.

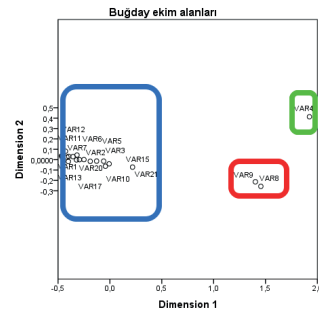
Anahtar Kelimeler: Ağır metaller, çok boyutlu ölçekleme, halk sağlığı, tarımsal bitki



Ayçiçeği için çok boyutlu ölçekleme haritası



Çeltik için çok boyutlu ölçekleme haritası



Buğday için çok boyutlu ölçekleme haritası

Diyabetik Gözlerde Multifokal Elektoretinografide Halka Amplitüdünün ve Halka Amplitüdünü Etkileyen Faktörlerin Değerlendirilmesi

¹Ayşe Büşra GÜNAY ŞENER, ²Hidayet SENER

¹Erciyes Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi ABD, Kayseri, Türkiye

²Erciyes Üniversitesi Tıp Fakültesi, Göz Hastalıkları ABD, Kayseri, Türkiye

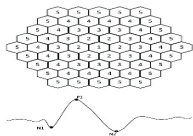
Amaç: Multifokal ERG (MFERG), retinanın elektrofizyolojik yanıtların topografik olarak değerlendirilmesine izin verir. Diyabetik hastalarda halka genliklerini değerlendirmeyi ve diyabetik retinopati için risk faktörlerinin halka genlikleri üzerindeki etkisini değerlendirmeyi amaçladık. Ayrıca halka genliklerinin tanısal değerini araştırmayı amaçladık.

Gereç ve Yöntem: Çalışmada retinopati olmayan (DM) 32 diyabetik hastanın 32 gözü, hafif proliferatif olmayan diyabetik retinopatisi (NPDR) olan 34 hastanın 34 gözü ve 62 kontrolün (CG) 62 gözü dahil edildi. MFERG kayıtları Uluslararası Klinik Elektrofizyoloji Derneği (ISCEV) önerilerine göre yapıldı. Halka genlikleri ve ring alanları Şekil 1 de tanımlanmıştır. Halka genlikleri tek yönlü varyans analizi ve posthoc Scheffe testi karşılaştırıldı. Diyabet ve hipertansiyonun halka genlikleri üzerindeki etkisi iki yönlü ANOVA ile değerlendirildi. Yaş, diyabet süresi, HbA1C ve halka genlikleri arasındaki ilişkiyi belirlemek için stepwise multivariate regresyon analizi yapıldı. Halka genliklerinin tanısal değerini değerlendirmek için üç yönlü bir ROC analizi yapıldı. Analiz sonucunda, ROC yüzeyi altındaki hacim (VUS) hesaplandı. VUS değeri 0.17'den büyükse, test başarılı olarak kabul edilmiştir.

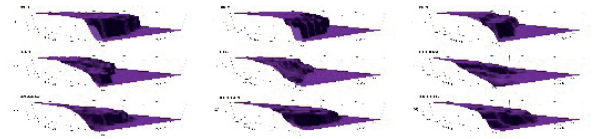
Bulgular: Halka genliklerinin (R1-R5) ve halka genlik oranlarının karşılaştırılmasında, DM ve NPDR gruplarının verileri CG'ye kıyasla anlamlı olarak azalmıştır. Diyabetin genlikler üzerinde etkili olduğu bulunmuştur, ancak hipertansiyon ve diyabet-hipertansiyonun etkileşimi etkili değildir. I2 değerlendirildiğinde, diyabetin etkisinin periferik halkalara doğru azaldığı bulunmuştur. HbA1c, tüm halka genlikleri (R1-R5) ile korelasyon göstermiştir ve en etkili faktördür. Öngörücü değeri periferik halkalara doğru azalmıştır. Diyabet süresi ilk üç halka genliği (R1-R3) ile korelasyon gösterdi. Yaş sadece R1 ile ilişkiliydi. Üç yönlü ROC analizinde değerlendirilen tüm parametrelerin (halka genlikleri ve oranları) VUS değeri 0.17'den büyüktü ve anlamlıydı. $VUS(R1) = 0.65$, $P < 0.001$ ve $[0.50-0.72]$ % 95 güven aralığına sahipti ve üç sınıflı ayırt etmek için en iyi kestirim noktasının $C1 = 217.3$, $C2 = 151.2$ olduğu bulundu (sensitivite 0.72, spesifite 0.59). VUS sonuçları Şekil 2 de sunulmuştur.

Sonuç: Halka oranı karşılaştırması, diyabetin şiddeti arttıkça retina nörodejenerasyonunun arttığını ve merkezi topografik lokalizasyonun daha fazla etkilendiğini ve daha objektif kanıtlar sağladığını göstermiştir. Tüm halka genlikleri diyabet nedeniyle etkilendir, ancak halka oranlarındaki değişiklik, diyabetle ilişkili retinopatide nispi retinal topografik etkinin daha kolay yorumlanmasını sağlayabilir. Grupların ayırımının yapılmasında R1 / R4 ve R1 / R5 oranları R1 / R2 ve R1 / R3 oranlarından daha iyi tanısal performans gösterdi. Bu, merkezden periferik halkaya değişen diyabetin etki büyüklüğü ile açıklanabilir. R1 genliğinin periferik genliklerden daha fazla etkilendiği görülmüştür, kronik hipergliseminin bir sonucu olabilir. Kronik hipergliseminin bir belirtisi olan HbA1c seviyeleri varyansın% 40'ını açıklamıştır. R1 genliği, en iyi tanısal değere sahipti ($VUS = 0.65$). $VUS 0.55$ değeri AUC için 0.8 değerine karşılık gelir. Bu nedenle, VUS'ları kendi sınırları içinde değerlendirdiğimizde, R1 ve R2 en iyi tanısal değere sahiptir. Sonuç olarak, diyabetin etkileri retina üzerinde topografik olarak eşit olmayan bir şekilde dağılmıştır. HbA1C kontrol edilebilir bir faktördür ve retina fonksiyonu düşüşü ile ilişkisi nedeniyle sıkı kontrolü önemlidir. R1 ve R2 genlikleri, retina fonksiyonundaki erken değişiklikleri tanımlamanın etkili yollarıdır.

Anahtar Kelimeler: Diyabetik retinopati, multifokal elektoretinografi, HbA1C, diyabet süresi, diyabetes mellitus



Şekil-1. Halkaların gruplandırılması. Retina eksantrikliği gösterilmemiştir. Yanıt yoğunlukları, her halka için P1 - N1 olarak hesaplandı.



Şekil-2. Halka genliklerinin ve halka oranlarının ROC yüzeyi altındaki hacimleri

