

Araştırma Makalesi

Prostat Spesifik Antijeni Yardımı ile Prostat Kanserinin Değişik Yapay Sinir Ağı Modelleri ile Tahmini

The Prediction of Prostate Cancer Using Different Artificial Neural Network Models with The Help of Prostate Specific Antigen

Mehmet Onur KAYA¹, Cemil ÇOLAK², Enver ÖZDEMİR³

¹Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik AD, Bursa

²Fırat Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilimleri AD, Malatya

³Fırat Üniversitesi Tıp Fakültesi Üroloji AD, Elazığ

Özet

Bu çalışmada, prostat spesifik antijeni (PSA) değerlerinin yardımıyla prostat kanseri olan ve prostat kanseri olmayan vakaların yapay sinir ağları (YSA) modelleri yardımıyla tahmin edilmesi amaçlanmıştır. Çalışma, geriye yönelik veri toplama yöntemi ile 203 erkek bireye ait olup, Fırat Üniversitesi Tıp Fakültesi Üroloji Anabilim Dalı Polikliniği'nden sağlanmıştır. Prostat kanserinin oluşumuna ilişkin PSA tipleri ve prostat kanserinin tanısında kullanılan değişkenler incelenmiştir. YSA, değişik öğrenme algoritmaları kullanılarak eğitilmiştir. Girdi katmanında 9 değişken kullanılmıştır. Çıktı katmanındaki değişken ise PSA değerlerine göre prostat kanseri olup olmamasıdır. Değişik YSA modelleri ve öğrenme algoritmaları denenerek, en iyi sonuç elde edilmeye çalışılmıştır. YSA modelleri, ileri beslemeli geriye yayımlı ağı kullanılarak tahminde bulunulmuştur. Çalışmadaki YSA modellerinde, en iyi açıklayıcılık katsayısı (R^2) ve en küçük hata kareleri ortalaması (MSE) sırasıyla; 0.75 ve 0.07 olarak bulunmuştur. Yapılan çalışmada değişik YSA modellerinin, PSA'nın yardımı ile prostat kanserinin tahmin edilmesinde daha etkili ve ümit verici sonuçlar verebildiği belirlenmiştir. Böylece klinik tanı sürecinde kullanılabilir olduğu görülmüştür.

Anahtar Kelimeler: Parametre tahmini, prostat spesifik antijeni, yapay sinir ağları.

Abstract

In this study, with the help of prostate specific antigen (PSA) values, patients with prostate cancer and without prostate cancer were predicted by different artificial neural networks (ANN) models. The study that was carried out retrospectively with 203 men was achieved in Department of Gazi University Faculty of Medicine, Department of Urology, Ankara, Turkey. The formation of prostate cancer was investigated, and some variables were used in the diagnosis of prostate cancer. ANN models were trained with different training algorithms. 9 process elements have been used in input layer. The output layer of value, the presence or absence of prostate cancer, was according to the PSA values. Trying out different ANN models and learning algorithms, the best results were achieved. ANN models were estimated using feed-forward back propagation network. Of the ANN models in the study, the best ANN model had determination coefficient (R^2) of 0.75 and mean squared error (MSE) of 0.07, respectively. Different ANN models can give promising results to predict prostate cancer with the help of PSA. Hence, ANN models may be used in the clinical diagnosis.

Key words: Parameter estimates, prostate specific antigen, artificial neural Networks.

Giriş

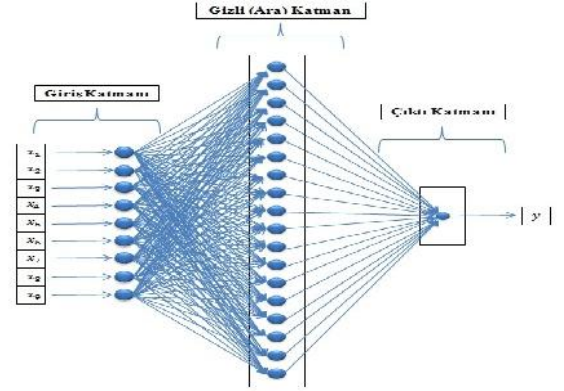
20. yüzyılın ikinci yarısında bilgisayarların hesaplama gücü ve hızının inanılmaz bir şekilde artması ve bilgisayar kullanımı yaygın bir hale gelmesi istatistik biliminin pratik uygulaması ve hatta teorik gelişmesi üzerine çok büyük etki yapmıştır. Pratik istatistik hesaplamasının çok zor olması nedeniyle veri analizi devamlı olarak hesaplamasının kolaylaştırılması üzerine odaklanıp daha çok deneysel modellere dayanmıştır. Çok yaygın kullanılan ve çok güçlü bilgisayarların kullanılmaya başlanması ve sayısal algoritmaların geliştirilip bilgisayar yazılımları geliştirilmesi ile yeni deneysel olmayan modeller (örneğin deneysel olmayan regresyon, genelleştirilmiş deneysel modeller, çok-seviyeli model gibi) pratikte kullanılmaya başlanmıştır (1-8). Bilgisayar devrimi tekrar örnekleme yöntemi, özyikleme yöntemleri, Gibbs örnekleme, permütasyon testleri gibi çok bilgisayara dayanan teknikler kullanılmaya başlanmıştır. Diğer taraftan istatistik gibi temeli ileri matematiğe bağlı olmayan ve büyük bilgisayar gücüne dayanan (veri madenciliği

gibi) araştırma ve pratik veri inceleme yöntemleri gelişmiştir. Bunun devamında ilk olarak 1950'lerde kendisini "yapay zekâ" adıyla tanıtan sezgisel yöntemler ortaya çıkmış ve çıktı gibi büyüyerek günümüzde hayatımızın vazgeçilmez yapı taşlarından biri olmuştur. Yapay zekâ ya da diğer bir ifadeyle "akıllı sistemler" oldukça karmaşık olan istatistik bilimine de çok büyük katkılar sağlayarak gelişmesine yardımcı olmuştur (9,10).

Gereç ve Yöntem

Çalışmada geriye dönük veri toplama yöntemi ile 203 erkek bireye ait yaş, DRE, TPSA, fPSA, f/t, Tvolüm, TZvolüm, PSAD, TZPSAD ve patolojiye ilişkin veriler, Fırat Üniversitesi Tıp Fakültesi Üroloji Anabilim Dalı Polikliniği'nden sağlanmıştır. PSA değerlerinin tahmininde yapay sinir ağları modelleri kullanılmıştır. YSA modelleri, ileri beslemeli geriye yayımlı ağı (feed-forward back propagation) kullanılarak tahminde bulunulmuştur. YSA modellerinin oluşturulmasında MATLAB

programlama tabanı kullanılmıdır. YSA, 1 giri katmanı, 1 gizli (ara) katman ve 1 çıktı katmanı olacak şekilde oluşturulmuştur ve de iki öğrenme algoritmaları kullanılarak eğitilmiştir. Girdi katmanında 9 elemanı kullanılmıştır. Bu girdi elemanları (x_1, x_2, \dots, x_9), ya da DRE, TPSA, fPSA, f/t, Tvolüm, TZvolüm, PSAD, TZPSAD de i kenleri idi. Gizli (ara) katmanda ise de iki sayıda yapay nöron kullanıldı. Çıktı katmanındaki değer (y) ise PSA değerlerine göre prostat kanser olup olmaması idi. Oluşturulan YSA modellerinden 9 girdili, ara katmanında 18 adet yapay nöron bulunan ve 1 çıktı katmanı bulunan YSA mimarisi, ekil 1'de gösterilmektedir.



ekil 1. PSA tahmini için oluşturulan YSA mimarisi

Bulgular

Çalışmanın girdi ve çıktı de i kenlerine ait tanımlayıcı bilgiler Tablo 1'de gösterilmiştir.

YSA'ya 203 bireye ait veri kümesinden örnekler rastgele seçilip verilerek eğitim, doğrulama ve test

yapılmıştır. De i iki YSA modelleri ve öğrenme algoritmaları denenerek en iyi sonuç elde edilmeye çalışılmıştır, bu modellerden en başarılı olan modellerden bazıları ve bu modellere ait bilgiler ile çalışmaların sonuçları Tablo 2'de verilmiştir.

Tablo 1. PSA Tahmini için Oluşturulan De i iki YSA Modellerine İlişkin De i kenler ve Tanımlayıcı De i kenleri

Giriş De i kenleri		Ortanca	(Min-Maks)
X_1	(Yaş)	65	40-92
X_2	(DRE)	0	0-2 cm
X_3	(TPSA)	9.33	0.4-779 ng/ml
X_4	(fPSA)	1.70	0.1-41 ng/ml
X_5	(f/t)	0.17	0.01-0.85 ng/ml
X_6	(Tvolüm)	56	12.2-216 mg
X_7	(TZvolüm)	30.20	4.9-175 mg
X_8	(PSAD)	0.15	0.01-16.1 mg
X_9	(TZPSAD)	0.32	0.03-32.4 mg
Çıktı De i kenleri		Çıktı Sonucu	
Y	(Patoloji)	BPH-CA	

BPH: Benign Prostat Hiperplazisi, CA: Kanser, DRE: (Digital Rectal Examination) Parmakla Rektal Muayene, fPSA: (Free) Serbest Prostat Spesifik Antijen, f/t: Serbest/Toplam Prostat Spesifik Antijen Oranı, PSAD: Prostat Spesifik Antijen Yo unlu u (PSA/Prostat Hacmi), TPSA: Toplam Prostat Spesifik Antijen, Tvolüm: Toplam Prostat Spesifik Antijen Hacmi, TZvolüm: Transizyonel Zon Hacmi, TZPSAD: Transizyonel Zon Prostat Spesifik Antijen Yo unlu unu (PSA/Transizyonel Zon Hacmi) göstermektedir.

Tablo 2. De i iki YSA Modellerine İlişkin Bilgiler ve Sonuçları

YSA	Katmanlardaki Nöron Sayısı			Örnek Büyüklükleri			Ö.A.	T.F.1	T.F.2	Sonuç	
	G	A	Ç	E	D	T				R ²	MSE
1	9	10	1	163	20	20	LM	T.Sig.	T.Sig.	0.66	0.15
2	9	12	1	163	20	20	LM	T.Sig.	T.Sig.	0.68	0.06
3	9	15	1	163	20	20	LM	T.Sig.	T.Sig.	0.72	0.09
4	9	18	1	121	41	41	LM	T.Sig.	T.Sig.	0.61	0.14
5	9	18	1	121	41	41	LM	T.Sig.	Purelin	0.62	0.16
6	9	18	1	183	10	10	LM	T.Sig.	T.Sig.	0.75	0.07
7	9	18	1	163	20	20	SCG	T.Sig.	T.Sig.	0.49	0.15
8	9	18	1	183	10	10	SCG	T.Sig.	T.Sig.	0.72	0.09
9	9	18	1	163	20	20	QN	T.Sig.	T.Sig.	0.45	0.14
10	9	18	1	121	41	41	BR	T.Sig.	T.Sig.	0.48	4.43

A: Ara Katman, BR: Bayesian Regulation Algoritması, Ç: Çıktı Katmanı, D: Doğrulama, E: Eğitim, G: Girdi Katmanı, LM: Levenberg Marquardt Algoritması, MSE: (Mean Squared Error) Hata Kareler Ortalaması, Ö.A.: Öğrenme Algoritması, Purelin: Doğrusal Aktivasyon Fonksiyonu, QN: Quasi-Newton Algoritması, SCG: Scaled Conjugate Gradient Algoritması, T: Test, T.F: Transfer Fonksiyonu, T.Sig.: Tangent Sigmoid Aktivasyon Fonksiyonunu göstermektedir.

Sonuçlar de erlendirildi inde, Tablo 2’de de görülece i gibi en ba arılı YSA modelinin 6. model oldu u görülmektedir. E itime ayrılan örnek hacminin büyük olması, gerçek sistemlerle daha iyi uyum sa laması açısından Tanjant Sigmoid aktivasyon fonksiyonunun kullanılması, hatayı azaltmada daha ba arılı olan Levenberg Marquardt ö renme algoritmasının kullanılması ve ara katmanda 18 adet yapay nöron kullanılması, 6. modelin daha ba arılı sonuçlar verdi ini göstermi tir.

Di er yandan olu turulan YSA modelleri incelendi inde ara katmandaki yapay nöron sayılarının sonucu de i tirdi i görülmü tür. Bu problem için ara katmandaki yapay nöron sayısı en uygun 18 adet olarak bulunmu tur. E itime ayrılan örnek hacimlerinin büyüklü ü de aynı ekilde çıktıkları oldukça önemli derecede etkilemekte oldu u görülmü tür. Ayrıca en ba arılı ö renme algoritmasının da Levenberg Marquardt ö renme algoritması oldu u görülmü tür. Aynı zamanda YSA modellerinde Tanjant Sigmoid aktivasyon fonksiyonunun yapıya daha uygun bir aktivasyon fonksiyonu oldu u da görülmü tür.

Tartı ma

statistiksel veri analizinde kullanılan yöntemlerden biri olan yapay sinir a ları di er birçok istatistiksel yöntemden daha ba arılı sonuçlar vermektedir. Veri analizinde kullanılan yöntemin verinin yapısına uygun olması istatistiksel analizin temel artıdır. Bu bakımdan veri analizinde kullanılan yöntemler parametrik ve parametrik olmayan yöntemler olmak üzere temel iki kısma ayrılmaktadır. Bu çalı mada parametrik olmayan istatistiksel yöntemlerden biri sayılan yapay sinir a ları, PSA’nın prostat kanseri tanısı koymada ne denli etkili bir de i ken oldu unun belirtilmesinde kullanılmı tir. Elde edilen ve istatistiksel açıdan ba arılı olan sonuçlar bulgular kısmında verilmi tir.

Çalı mada incelenen problemin çözümünde di er sezgisel yöntemlerin de kullanılması mümkündür. Fakat ba langıç çözümünden ba ımsız olma, paralel çözüm arama, hızlı çalı ma ve daha ba arılı sonuçlar verme özelliklerinden dolayı yapay sinir a ları daha uygun çözüm yöntemi olarak kabul görmektedir.

Yapılan çalı manın uygulama kısmında ise prostat kanserinin olu umuna ili kin PSA tipleri ve prostat kanserinin tanısında kullanılan de i kenler incelenmi tir. Çalı mada ise YSA’ya 203 bireye ait veri kümesinden örnekler rastgele seçilip verilerek e itim, do rulama ve test yapılmı tir. De i ik YSA modelleri ve ö renme algoritmaları denenerek en iyi sonuç elde edilmeye çalı ılmı tir. Her model için de i ik sayıda örnek e itim, do rulama ve test için kullanılmı tir. Çok katmanlı YSA modeli kullanılmı tir. Tahmin performansı ise, açıklayıcılık katsayısı (R^2) ve hata kareleri ortalaması (MSE) de erlerine göre de erlendirilmi tir. Çalı madaki YSA modellerinde, en iyi açıklayıcılık katsayısı ve hata kareleri ortalaması sırasıyla 0.75 ve 0.07 olarak 6. model bulunmu tur. Bu modelde 9 adet girdi katmanı

i lem elemanı, 18 adet ara katman i lem elemanı ve 1 adet çıktı katmanı i lem elemanı kullanılmı lardır. Girdi ve ara katmanda transfer fonksiyonu olarak tanjant sigmoid aktivasyon fonksiyonu kullanılmı tir. Ö renme algoritması olarak ise Levenberg Marquardt geriye yayılım ö renme algoritması kullanılmı tir.

Benzer ekilde, Çolak ve arkadaş ları, yaptıkları çalı mada koroner arter hastalığının (KAH) tahmin edilebilmesi için de i ik sekiz ö renme algoritması ile farklı YSA modelleri olu turulmu ve tanıtılmasını amaçlamı lardır (1). Çok katmanlı “perceptron” YSA uygulanmı ve de i ik sekiz ö renme algoritması ile e itilen farklı YSA modelleri, toplam 237 kayıta, 171’i e itimde ve 66’sı ise teste kullanılarak olu turulmu tur. Tahminin performansı, duyarlılık, seçicilik ve do ruluk oranlarına dayalı olarak de erlendirilmi tir. Çolak ve arkadaş larının çalı masında (1), olu turulan YSA modelleri ile KAH’ın tahmininde yüksek oranda duyarlılık, seçicilik ve do ruluk oranları elde ettikleri için modellerin performansının iyi oldu unu göstermi lerdir. Çalı malarında olu turdukları YSA modellerinde, en iyi do ruluk, duyarlılık ve seçicilik de erleri; e itimde sırasıyla %83.63 - %100, %86.46 - %100 ve %74.67 - %100 arasında iken, testte ise duyarlılık %71’den daha büyük, seçicilik %76’dan daha büyük ve do ruluk %81’den daha büyük olarak 8. modeli bulmu lardır. Bu modelde 17 adet girdi katmanı i lem elemanı, 17 adet ara katman i lem elemanı ve 1 adet çıktı katmanı i lem elemanı kullanılmı lardır. Girdi ve ara katmanda transfer fonksiyonu olarak tanjant sigmoid aktivasyon fonksiyonu kullanılmı tir. Ö renme algoritması olarak ise geriye yayılım ö renme algoritması kullanılmı tir.

Son birkaç yılda prostat kanseri tanısında bazı çalı malar yapılmı tir. Literatürdeki bazı çalı malar incelenmi tir. Literatürde yer alan Gülkesen ve arkadaş larının yaptı ı çalı mada, prostat kanserinin tahmininde, PSA’nın tanısız do rulu unu artırmak için kullanılan yöntemleri kar ıla tırılması amaçlanmı tir (2). Bunun için 997 hastadan prostat biyopsisi alınmı , PSA, PSAD ve f/t kullanılarak ikili lojistik regresyon (LR), yapay sinir a ları (YSA) ve karar a açları (DTs) tahmin modelleri geli tirilmi tir. LR, YSA ve DTs için 241 vakadan olu an bir grup do rulamaya ayrılmı tir. Gülkesen ve arkadaş ları yaptıkları çalı mada (2), 997 vakadan aldıkları biyopsi sonuçlarına göre 193 vakanın prostat kanseri oldu unu bulmu tur. Çalı mada, do rulama grubunda i lem karakteristik e risi (ROC) altında kalan alan (AUC) LR, YSA ve DTs için sırasıyla 0.717, 0.516 ve 0.629 olarak bulmu lardır. Sonuç olarak, çok katmanlı modellerin birbirine benzer sonuçlar verdi ini; ancak YSA’nın AUC (Area Under Curve) e ri altında kalan alan de erinin dü ük olmasının sınırlı de i ken kullanılmasından kaynaklandı mı belirtmektedirler. Ba ka bir çalı mada, Lawrentschuk ve arkadaş ları, prostat kanserini (PC) özellikle biyopsi sonuçlarını tahmin etmek için çoklu lojistik regresyon (LR) modeli ve YSA kullanılmı lardır (3). Biyopsi uygulanan 3025 erkek birey üzerinde çalı malarını gerçekle tirmi ve

girdi de i kenleri olarak; ya , PSA, DRE, TRUS (Trasrektal Ultrasonografi) ve prostat hacmini kullanarak, LR ve YSA ayırmsama modelleri geli tirmi lerdir. Bu çalı mada LR ve YSA'ya ait do ru tahmin yüzdeleri sırasıyla %55 ve %57 olarak birbirine yakın bulunmu ve bu ayırmsamanın iyi yapılmadı ı söylenmi tir. Sonuç olarak; ek de i kenlerin kullanılmasının problemin çözümüne katkı sa layabilece ini ifade etmi lerdir. Lee ve arkadaş ları ise, prostat biyopsi sonucunu tahmin etmek ve her bir modelin do rulu nu kar ıla tırmak için, bir çoklu lojistik regresyon, bir YSA ve bir destek vektör makinası (SVM) modeli geli tirmi lerdir (4). Biyopsi uygulanan 1077 vakadan TRUS bulgularını, PSA, PSAD ve TZPSAD girdi de i kenleri olarak almı lardır. Vakaları rastgele e itim ve test gruplarına ayırarak çalı mayı yapmı lardır. Her bir modelin genel performansını göstermek için ROC analizi ile AUC'yi hesaplamı lardır. LR, YSA ve SVM için AUC de erlerini sırasıyla 0.768, 0.778 ve 0.847 olarak bulmu lardır. Sonuç olarak; bu çalı ma kapsamındaki veriler için SVM performansının YSA ve LR'dan daha iyi oldu unu göstermi lerdir.

Sonuç olarak, olguların ya ve di er de i kenlere göre PSA de erlerinin aracılı ı ile prostat kanseri olup olmama durumunun tahmininde, YSA modellerinin bu çalı ma kapsamındaki veriler için son derece uygun, memnun edici ölçüde ba arılı sonuçları hızlı bir ekilde üreten bir sezgisel yöntem oldu u görülmü tür. De i ik YSA modelleri, PSA'nın tahmin edilmesinde ümit verici sonuçlar vermektedir ve ileriye yönelik klinik tanı sürecinde kullanılabilir. Bundan sonraki çalı malarda, gerek tıp alanında gerekse di er alanlardaki konulara ili kin büyük örneklemeler üzerinde çalı lması daha fazla katkı sa layabilir.

Kaynaklar

1. Çolak MC, Çolak C, Kocatürk H, Sa ıroglu , Barutçu . Predicting coronary artery disease using different artificial neural network models. Anadolu Kardiyoloji Dergisi 2008; (8): 249-54.
2. Gülkesen KH, Köksal IT, Bilge U, Saka O. Comparison of methods for prediction of prostate cancer in Turkish men with PSA levels of 0-10 ng/mL. J Buon 2010; 15(3): 537-42.

3. Lawrentschuk N, Lockwood G, Davies P, Evans A, Sweet J, Toi A, Fleshner NE. Predicting prostate biopsy outcome: artificial neural networks and polychotomous regression are equivalent models. Int Urol Nephrol 2011;43(1): 23-30.
4. Lee HJ, Hwang SI, Han SM, Park SH, Kim SH, Cho JY, Seong CG, Choe G. Image-based clinical decision support for transrectal ultrasound in the diagnosis of prostate cancer: comparison of multiple logistic regression, artificial neural network, and support vector machine. Eur Radiol 2010; 20(6): 1476-1484.
5. Öner Y, Gürcan M. The basic algorithm for parameter estimation in nonlinear models. OMÜ Fen Dergisi 1997; 8(1): 119-125.
6. Florio L, Mussone L. Neural Network Models for classification and forecasting of freeway traffic flow stability. Control Eng. Practice 1996; 4(2): 153-164.
7. Psaltis D, Sideris A. A multilayer neural network controller. IEEE Control System Magazine 1998; 8: 17-21.
8. Binder, DA. On the variances of asymptotically normal estimators from complex surveys. International Statistical Review 1983; 51; 279-292.
9. Levenberg K. A method for the solution of certain problems in least squares. Quart Appl Math 1944; 2; 164-168.
10. Marquardt D. An algorithm for least-squares estimation of nonlinear parameters. Siam J Appl Math 1963; 11; 431-441.

İletişim

Mehmet Onur KAYA
Uluda Üniversitesi Üniversitesi Tıp Fakültesi
Biyostatistik AD, Bursa
monurkaya@gmail.com