İmmunohistokimyasal Boyalar İle Tiroid Tümörü Teşhisinde Naive Bayes Algoritması Kullanılması

Ahmet Haltaş[1], Ahmet Alkan[2]

[1] GAUN, Gaziantep Meslek Yüksekokulu, Gaziantep

[2] KSU, Elektrik Elektronik Mühendisliği Bölümü, Kahramanmaraş

haltas@gantep.edu.tr, aalkan@ksu.edu.tr

Özet: Tiroid nodüllerinin ayrımında preoperatif olarak kolay, pratik ve güvenilir yöntem olarak kabul edilen ince iğne aspirasyon biyopsisi (İİAB) kullanılmaktadır. Nodüllerin ayrımında patologlar morfolojik benzer özellik gösteren lezyona farklı tanılar koyabilmektedirler. Bu durum tiroid nodüllerinin ayırıcı tanısında ilave immunohistokimyasal çalışmaların gerekliliğini göstermektedir. Çalışmamızda histopatolojik incelemeye ek olarak yapılan İmmünohistokimyasal boyama verilerinin Naive Bayes algoritması ile değerlendirilerek, nodülün malign olup olmadığı bilgisi için doktora yardımcı bir karar destek sistemi gerçekleştirilmiştir. Bu tanı sisteminde, 63 olgudan alınan veriler, Naive Bayes algoritması ile sınıflandırılmış ve J48 algoritması ile karşılaştırılarak %92 başarım oranı ile karar destek sistemi olarak kullanılabileceğini göstermiştir.

Anahtar Sözcükler: Tiroid Hastalığı, İmmünohistokimyasal, Naive Bayes, J-48

Use of Naive Bayes algorithm in the Immunohistochemical dyeing based diagnosis of Thyroid Tumor

Abstract: Fine needle aspiration biopsy (FNAB) is considered as an easy, practical and reliable preoperative method in classification of thyroid nodules. While classifying these nodules, pathologists may make different diagnosis decisions for same kind of lesions with different morphologic features. Therefore, it is required that additional information such as Immunohistochemical data should also be considered in diagnosis of thyroid nodules. This paper presents a Naïve Bayes algorithm based methodology to support decision making of a physician to classify a nodule into either benign or malignant. The proposed methodology takes Immunohistochemical data, which is obtained after the histopathologic investigation process, as input to the algorithm. This automatic diagnosis system is tested over data of 63 patients and showed %92 accuracy. The system is also compared with J48 algorithm.

Keywords: Thyroid Disease, Immunohistochemical, Naive Bayes, J-48

1. Giriş

Tiroid bezinden kaynaklanan hastalıklar, WHO (Dünya Sağlık Örgütü) raporlarına göre dünya nüfusunun %7 sini etkilemekte ve bu orandan yola çıkarak Türkiye'de yaklaşık olarak 4.5-5 milyon kişinin troid ile ilgili hastalıklardan etkilendiği tahmin edilmektedir [2, 12].

Tiroid nodüllerinin ayrımında preoperatif olarak kolay, pratik ve güvenilir yöntem olarak kabul edilen ince iğne aspirasyon biyopsisi (İİAB) kullanılmaktadır [15].

Nodüllerin ayrımında patologlar morfolojik benzer özellik gösteren lezyona farklı tanılar koyabilmektedirler [3, 5, 8]. Konulan bu tanı tedavi sürecini etkileyerek yanlış tanı koyma durumunda, gerçekte malign, tanıda benign ise erken tedavinin önemli olduğu bu durumda tedavi gecikir. Tersi durumlarda ise gereksiz cerrahi işlem yapılması daha önemli bir sorundur.

Ayırt edilemeyen tiroid nodül tanısında, doğru tanı oranını artırmak için İmmünohistokimyasal boyalar kullanılarak tanı koymaya yardımcı olarak kullanılmıştır [3, 4, 10, 11, 13, 14]. Araştırmalar sonucu kullanılan immünohistokimyasal boyalarının tiroid tümörlerinin tanısında destekleyici olduğu belirtmektedirler [7, 11, 13, 14] .

Bu çalışmanın amacı morfolojik benzerlik gösteren tiroid nodüllerin tanısında, tanıyı destekleyici olduğu düşünülen immünohistokimyasal boyaların tanıya etkisini araştırmaktır. Doğru tanı için Naive Bayes sınıflandırıcısı ile belirlenen belirteç bilgilerin başarılı sınıflandırılarak gereksiz tedavilerin uygulanmaması ve erken teşhis ile hekime karar vermelerinde yardımcı olması amaçlanmaktadır. Yapılan çalışmalarda immünohistokimyasal boyalar ile tiroid tümör teşhis sistemine dair sınıflandırma modelli bir çalışmaya rastlanmadığı için, önerilen çalışmanın ileriki çalışmalara önayak olabileceği düşünülmektedir.

2. Materyal ve Metot

2.1. Materyal

Bu çalışmada S. Çekiç'in 2009 yılında yaptığı çalışmada [6] bulunan ilgili veriler kullanılmıştır. Veri tabanında toplam 63 örnek bulunmaktadır. Bunlar iki sınıfa ait örneklerdir. Bu sınıflar Sınıf 1:benign (18 adet), Sınıf 2: malign (45 adet) sınıflarıdır.

Veri tabanındaki her örnekte 8 özellik (attribute) bulunmaktadır. Her özellik sayısal bir değer almaktadır. Bu özellikler şunlardır:

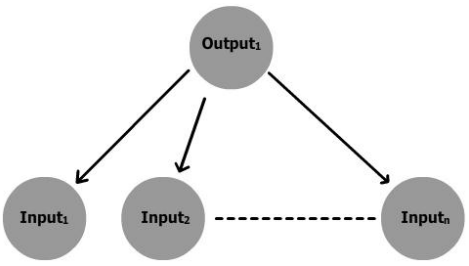
Tablo 1. Veri tabanındaki giriş parametreleri

İsim	Tanım	Tip
Yaş	Kişilerin yaşı	Sayısal
Cinsiyet	Kadın/Erkek	Sınıflandırılabilir 0/1
CK-19 boyanma yaygınlığı	İlgili boyaya karşı verdikleri tepkileri temsil etmektedir. Sayısal olarak 0 ile 3 arası bir değer almaktadır.	Sayısal
CK-19 boyanma şiddeti	İlgili boyaya karşı boyanma şiddeti temsil etmektedir. Sayısal olarak 0 ile 3 arası bir değer almaktadır.	Sayısal
HBME-1 boyanma yaygınlığı	İlgili boyaya karşı verdikleri tepkileri temsil etmektedir. Sayısal olarak 0 ile 3 arası bir değer almaktadır	Sayısal
HBME-1 boyanma şiddeti	İlgili boyaya karşı boyanma şiddeti temsil etmektedir. Sayısal olarak 0 ile 3 arası bir değer almaktadır.	Sayısal
Cyclin D1 boyanma yaygınlığı	İlgili boyaya karşı verdikleri tepkileri temsil etmektedir. Sayısal olarak 0 ile 3 arası bir değer almaktadır.	Sayısal
Cyclin D1 boyanma şiddeti	İlgili boyaya karşı boyanma şiddeti temsil etmektedir. Sayısal olarak 0 ile 3 arası bir değer almaktadır.	Sayısal

2.2. METOT

2.2.1. NAİVE BAYES

Naive Bayes (NB) Sınıflandırıcısı, Bayes teorisini esas alan olasılık tabanlı sınıflandırma tekniklerinden biridir. Bayes Network olaylar ile arasındaki olası ilişkileri, koşullu olasılıklar kullanarak gösteren grafiksel modellerdir. Grafiksel modelde değişkenler düğüm olarak gösterilerek olası ilişkiler ise düğümler arasındaki çizgiler olarak gösterilir. Eğer düğüm bağlı değilse bu diğer değişkenlerden şartlı olarak bağımsız anlamına gelir [9].



Şekil-1 Naive Bayes yapısı

Bağımsız ve rasgele iki olayın olması durumunda, birinci olayın olduğu durumda ikinci olayın olma olasılığı P(A∩B) ifadesi ile gösterilir. Bu ifade aşağıdaki gibi iki farklı eşitlik şekilde de yazılabilir.(Denklem 1)

$$P(A\cap B)=P(A|B)P(B)=P(B|A)P(A) \quad (1)$$

Bayes teoremi şartlı olasılığı A ve B iki değişken için, Denklem 2 ile tanımlanmaktadır.

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)} \quad (2)$$

Denklem 2'de P(A) ifadesi problemin girdi olasılığını, P(B) ifadesi olası çıkıs olayının olasılığını ve P(A|B) ifadesi ise A girişi gerçekleştiği durumda B çıkış olayının meydana gelme olasılığını ve P(B|A) ifadesi ise B çıkışı gerçekleştiği durumda A giriş olayının meydana gelme olasılığını temsil etmektedir.

Burada X giriş değerlerini, n adet özellikten X={ X_1, X_2,...., X_n } oluşacak şeklinde ifade edilmekte ve düğümdeki değişkenlerin ortak olasılığı zincir kuralı ile denklem 3'teki gibi hesaplanır.

$$P(X|Y_j)=\prod_{i=1}^{n}P(X_i|Y_j) \quad (3)$$

Denklem 3'de Yj fonksiyonu belli bir düğümün komşuluğunu, n ise Naive Bayes yapısındaki düğüm sayısını ifade etmektedir. Denklem 3'de paydaların eşitliğinden dolayı pay değerlerini karşılaştırarak sınıflandırma yapılır. Çıkan sonuçlar içinden en büyük olanı seçilerek sınıflandırma işlemi yapılmış olur. Bu durumda Naive Bayes formülü aşağıdaki denklem 4'teki gibi hesaplanır.

$$Y'=\underset{y_j\epsilon Y}{\arg max}\ (P(Y=y_j)\prod_{i=1}^{n}P(X=x_i|Y=y_j)) \quad (4)$$

3. Deneysel Çalışmalar

Materyal bölümünde anlatılan veriseti 10 kat çapraz doğrulama (cross-validation) ile Naive Bayes modeline göre sınıflandırma yapılmıştır. Tablo-1 de ilgili sonucu göstermektedir:

Tablo 2 – Sonuç değerlendirmesi

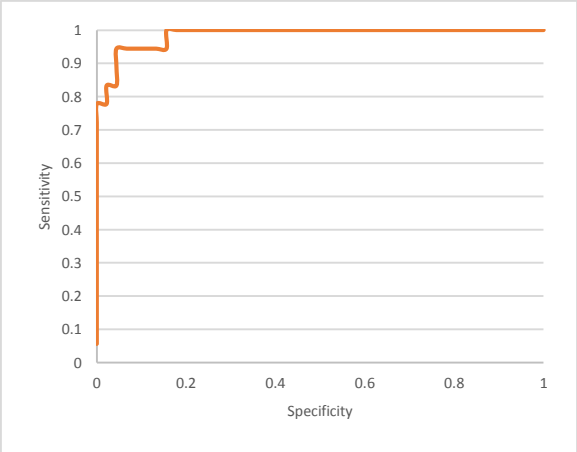
	Precision (Kesinlik)	**Recall (Anma)**	**F-Measure (F-Ölçütü)**	**Sınıf**
	0.882	0.833	0.857	benign
	0.935	0.956	0.945	malign
Ağırlıklı Ort.	0.92	0.921	0.92	

Benign olan örnek grubunda testin pozitif olma olasılığı, testin duyarlılığıdır.
Precision(Kesinlik) = (TN) / (TN+FN) (5)

Bir testin Recall (Anma) oranı, malign olan örnek grubunda testin nagatif olma olasılığıdır.
Recall (Anma) = (TP) / (TP+FN) (6)

F-ölçütü aşağıdaki şekilde hesaplanmıştır.
F−Ölçütü=(2×Kesinlik×Anma)/(Kesinlik+ Anma) (7)

Deney sonucu, performans üç kritere göre test edilmiştir. Precision(Kesinlik), Recall (Anma) ve F-Measure(F-ölçütü) . Sonuca göre, ağırlıklı ortalamada tüm sınıflar için %92 başarım elde edilmiştir. Ayrıca önerilen modelin sınıflandırma performansına ait ROC (Receiver Operating Characteric=Alıcı İşletim Karakteristiği) eğrileri de üretilmiştir.

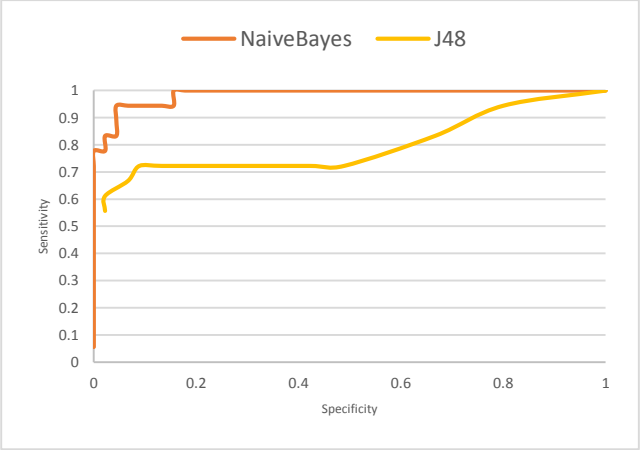


Şekil 2. Naive Bayes'e ait ROC eğrisi

Çalışmada kullanılan model ayrıca başka sınıflandırma algoritmaları ile de kıyaslanmıştır. Bu amaçla Naive Bayes'den farklı olarak J48 algoritması ile karşılaştırılmıştır. J48 ise karar ağacı algoritması C4.5'in java uygulamasıdır. Önerilen Naive Bayes modeli, J48 modeli ile kıyaslanmış ve kıyaslama sonuçları Tablo 3 de gösterilmiştir.

Tablo 3. Önerilen Naive Bayes modelinin diğer sınıflandırma algoritması ile kıyaslanması

Algoritma	F-Measure
Naive Bayes	0.920
J48	0.856



Şekil 3. Her sınıflandırmaya ait ortalama sınıflandırma performans ROC eğrileri

Tablo 3'de görüldüğü üzere, önerilen Naive Bayes tabanlı model diğer sınıflandırıcıya (J48) göre tiroid teşhis probleminde daha iyi performans göstermiştir. Bu sonuçlar önerilen modelin, güvenirliğini ispatlamaktadır.

4. Sonuç ve Değerlendirme

Hızla gelişen uzman sistemler, tıp alanında da katkıda bulunarak daha hızlı ve doğru kararlar alınmasında fayda sağlamaktadır. Bu sistemler sayesinde daha doğru kararlar alınarak hekime yardımcı olacak karar destek sistemleri oluşturulmaktadır.

Bu çalışmada tiroid nodül tanısında yardımcı olacak Naive Bayes tabanlı sınıflandırma yapılmıştır. Veri kümesindeki test başarım oranları Naive Bayes modeli %92 elde edilirken Karar Ağaçları modeli ile % 85

başarım hesaplanarak kıyaslama yapılmıştır. Yapılan çalışmalar Naive Bayes yönteminin %92 başarım oranı ile karar destek sistemi olarak immunohistokimyasal boyalar ile tiroid tümörü teşhisinde kullanılabileceğini göstermiştir.

Bu çalışma devam etmekte olan çalışmalarımızın ilk bölümü olup, farklı algoritmalar ile daha başarılı yöntemler geliştirilmek hedeflenmektedir. Çalışmada kullanılan veri sayısı ve nitelik sayısı artırılarak daha başarılı doğruluk derecesi elde edilecek sistemin geliştirilmesi hedeflenmektedir.

Teşekkür

Turgut Özal Üniversitesi, Patoloji Anabilim Dalı öğretim üyesi Yrd. Doç. Dr. Hacer HALTAŞ' a çalışmaya vermiş oldukları katkı ve yardımlarından dolayı teşekkür ederiz.

Kaynakça:

[1] Beesley Mf, M. K., "Cytokeratin 19 And Galectin-3 İmmunohistochemistry İn The Differential Diagnosis Of Solitary Thyroid Nodules", **Histopathology** , 41: 236-43. , (2002).

[2] Bender Ö, Y. E. Ç. H. H. A. A. B. A. O. E. A. ,"Total Tiroidektomi Deneyimlerimiz", **Endokrin Diyalog**, 1: 15-18. , (2004).

[3] Birigitte F., "Observer Variation Of Lesions Of The Thyroid". **Am J Surg Pathol** , 27: 1177-1178. , (2003).

[4] Casey Mb, L. C. L. R. "Distinction Between Papillary Thyroid Hyperplasia And Papillary Thyroid Carcinoma By İmmunohistochemical Staining For Cytokeratin-19, Galectin-3, And Hbme-1", **Endocr Pathol** , 14(1): 55-60. , (2003).

[5] Colı A, B. G. Z. F. E. A. ,"Galectin-3, A Marker Of Well Differentiated Thyroid Carcinoma, İs Expressed İn Thyroid Nodules", **Histopathology**, 40: 80-87. , (2002).

[6] Çekiç, S. "Tiroid Papiller Karsinomda Ck 19, Hbme-1, Cyclin D1 Ekspresyonu", **Uzmanlık Tezi** İstanbul, (2009).

[7] Haltaş, H. Et Al. ,"Diagnostic Value Of Cytokeratin 19, Hbme-1, And Galectin-3 İmmuno-Staining Of Cell Block", **Tubitak-Sag**-1208-66, (2013).

[8] Hırokawa M, C. J. G. J. E. A. ,"Observer Variation İn Encapsulated Follicular Lesions Of The Thyroid", **Am J Surg Pathol** , 26: 1508-1514. , (2002).

[9] Luis, E. Et. Al. ,"Fault Diagnosis Of Industrial Systems With Bayesian Networks And Neural Networks, Advances İn Artificial Intelligence", **Springer-Verlag**, P. 998- 1008, (2008).

[10] Lıberman E, W. N. "Papillary And Follicular Neoplasms Of The Thyroid Gland: Differential İmmunohistochemical Staining With High-Molecular-Weight Keratin And İnvolucrin. Appl Immunohistochem", **Mol Morphol**, 8(1):42-8. , (2000).

[11] Matos Ps, F. A. O. F. F. A. L. M. K. W. L.," Usefulness Of Hbme-1, Cytokeratin 19 And Galectin-3 İmmunostaining İn The Diagnosis Of Thyroid Malignancy", **Histopathology Oct**,47(4): 391-40. , (2005).

[12] Müller Pe, K. S. R. E. S. F. Indications, Risks And Acceptance Of Total Thyroidectomy For Multinodular Benign Goiter. **Surg Today**, 31: 958-962. , (2001).

[13] Prasad Ml, P. N. H. Y. N. H. C. A. K. R. ,"Galectin-3, Fibronectin-1, Cıted-1, Hbme1, And Cytokeratin-19 İmmunohistochemistry İs Useful For The Differential Diagnosis Of Thyroid Tumors", **Mod Pathol**,18(1): 48-57. , (2005).

[14] Teng Xd, W. L. Y. H. L. J. D. W. Y. L. ,"Expression Of Cytokeratin19, Galectin-3 And Hbme-1 İn Thyroid Lesions And Their Differential Diagnoses [Abstract]", **Zhonghua Bing Li Xue Za Zhi**, 33(3): 212-6. , (2004).

[15] Yetkın E.,"Tiroidektomi Komplikasyonları", **Tiroit Hastalıkları Ve Cerrahisi**, İşgör A.(ed): 1.Baskı, İstanbul, 10: 583-95. , (2000).