

Bayes Ağlarda Kümeleme Metotunu Kullanarak Meme Kanseri Tanısının Modellenmesi

Modeling Breast Cancer Diagnosis Using Clustering Method in Bayesian Networks

Yrd.Doç.Dr. Hülya OLMUŞ,^a
Prof.Dr. Semra ERBAŞ^a

^aİstatistik Bölümü,
Gazi Üniversitesi Fen Fakültesi,
Ankara

Geliş Tarihi/Received: 13.09.2011
Kabul Tarihi/Accepted: 23.11.2011

Yazışma Adresi/Correspondence:
Yrd.Doç.Dr. Hülya OLMUŞ
Gazi Üniversitesi Fen Fakültesi,
İstatistik Bölümü, Ankara,
TÜRKİYE/TURKEY
hulya@gazi.edu.tr

ÖZET Amaç: Olasılık teorisi ve grafik teorisinin birlikte kullanımı ile elde edilen modellere olasılıklı ağ modelleri denir. Bu modelleri oluşturma, istatistik modellerini oluşturmamanın bir çeşidi olup, bu modelleri göstermek için grafikler kullanılır. Grafikler, ortak olasılık fonksiyonlarının gösterimini ve gözlemlerden etkili sonuç çıkarımlarını kolaylaştırmak için önemli rol oynarlar. Böyle grafiklerde, değişkenler düğümleri, kenarlar ise değişkenler arası bağımlılıkları ve nedensel etkiyi gösterir. Olasılıklı ağ modellerin en çok kullanılan çeşitlerinden biri olan Bayes ağlar, yön verilmiş döngüsel olmayan grafiklerle belirlenir. İstatistiksel metotların, tıbbi tanımlara uygulama alanlarında Bayes ağlar önemli bir yer tutar. Bayes ağlar, gözlenmiş olan belirtilerden nedenleri ortaya çıkarmak için, tıp ile ilgili birçok çalışmada birçok alana uygulanabilir. Amaç, kesikli değişkenler için, Bayes ağları tanıtmak ve bu konuda verilen bir problemin nasıl çözüleceğini ve yorumlanacağını açıklamaktır. **Gereç ve Yöntemler:** Bu çalışmada, Bayes ağlar tanıtılmış, veriden en iyi modeli tahmin etmek için K2 algoritması verilmiştir. Bu model elde edildikten sonra, Bayes ağlar için kümeleme metotları kullanılarak, veriyi en iyi açıklayan küme ağacı elde edilmiştir. Tabloların düzenlenmesi ve tanımlayıcı istatistikler için SPSS paket program ve Bayes ağlarda gerekli olan olasılık hesaplamaları için HUGIN paket program kullanılmıştır. **Bulgular:** Uygulamada, Ankara ili ve çevresinde meme kanserine yakalanmış 74 kadın üzerinde bir anket çalışması yapılmıştır. Meme kanserinin erken teşhisinin önemi, kansere neden olan etkiler ve bazı önemli sonuçlara yer verilmiştir. Çalışmaya katılan kadınların kadınların %35'i kendi kendine meme muayenesi yaptığı ve %65'inin mamografi yoluyla meme kanseri teşhisi konulduğu gözlenmiştir. **Sonuç:** Toplumun meme kanseri konusunda bilgilendirilmesi ve kadınların yılda en az bir kez doktor muayenesine gitmeleri sağlanmalıdır.

Anahtar Kelimeler: Bayes ağ; öğrenme; kümeleme algoritması; meme kanseri

ABSTRACT Objective: Models obtained by probability theory and graph theory are called probabilistic network models. Formation of these models is a type of statistical model formation and graphs are used for showing these models. Graphs play an important role in displaying joint probabilistic functions and easing the derivation of effective results from observations. In these graphs, variables represent nodes and the edges represent dependence between variables and causal effect. Bayes network models, one of the probabilistic network models used most frequently are determined by directed acyclic graphs. Bayesian networks have an important place in the application of statistical methods to medical diagnosis. Bayesian networks may be applied to numerous problems in medicine in order to reveal causes from observed symptoms. The aim is to introduce Bayesian networks for discrete variables and to interpret and explain how to solve a given problem in this issue. **Material and Methods:** In this study, Bayesian networks are introduced and K2 algorithm is given to estimate the best model from data. After the best Bayesian network is obtained, clustering methods are given for Bayesian networks. Finally the cluster trees are obtained for the best model from data. Here, SPSS software are used to constitute contingency tables and descriptive statistics and HUGIN software for probability calculus in Bayesian networks. **Results:** In practice, a survey study has been made on 74 women who have breast cancer in and around Ankara province. The importance of early diagnosis of breast cancer, cancer-causing effects, and some important results are given. According to this study, 35% of these women did breast self-examination and put into 65% of breast cancer through mammography has been observed. **Conclusion:** Society should be informed about breast cancer and women should go to doctors visits at least once a year.

Key Words: Bayesian network; learning; clustering algorithm; breast cancer

Bayes ağlar, istatistikte uzun bir tarihe sahiptir. 1980'lerin başında uzman sistemlerde kesin olmayışlık ile ilgili çalışmalar yapılmıştır. Uzman görüşler sayesinde farklı değişkenler arası bağımlılıklar Bayes ağlar ile verilebilir. Son on yıl süresince, Bayes ağ, uzman sistemlerde kesin olmayan bilgiden yararlanmak için popüler bir gösterim oluşturmuştur. Daha yakın bir geçmişte, araştırmacılar, verilerden Bayes ağı öğrenilmesi için metotlar geliştirmişlerdir. Bu teknikler yeni olup, gelişimleri sürmektedir. Bu tekniklerin birçok veri-analiz problemleri için etkili olduğu gösterilmiştir. Ayrıca, son yıllarda yapılan çalışmalarda bilgi alanı ile olasılığa dayalı uzman sistemlerin gelişimi üzerinde durulmaktadır. Kesin olmayan sonuçların olasılıkları üzerinde olayların etkisini genişletmek mümkündür.

Grafikler, ortak olasılık fonksiyonlarının gösterimini ve gözlemlerden etkili sonuç çıkarımlarını kolaylaştırmak için önemli rol oynarlar. Böyle grafiklerde, değişkenler düğümleri, kenarlar ise değişkenler arası bağımlılıkları ve nedensel etkiyi gösterir. Bayes ağ modelleri, yön verilmiş döngüsel olmayan grafiklerle belirlenir. Bayes ağ, değişkenler ve değişkenler arası yön verilmiş kenarların kümesinden oluşur. Kenarlar, değişkenler arası olasılık bağımlılıkları gösterir. Bu bağımlılıklar, koşullu olasılıkların kümesinden oluşur. Her bir değişkenin ebeveynleri verildiğinde değişkenin koşullu olasılığı belirlenir. Bir düğümün ebeveynleri olmadığı zaman, bir değişken koşulsuz (marjinal) bir olasılığa sahiptir. Kısaca, bu ağ modelleri, değişkenler kümesi arasındaki ilişkileri grafik olarak göstermek ve uzman sistemlerde belirsizlik ile ilgilenmek için güçlü bir yöntemdir.¹ Bayes ağları modellemek için model yapısı ve koşullu olasılıklar için uzman görüşler yerine, veriden tahmin etme işlemi yapılacaktır. Ayrıca, veri tabanının kayıp veri içermediği durum ele alınacaktır.

Bir grafikteki düğümlerin bir araya getirilmesi ile elde edilen düğümlerin grubu, *küme*ler olarak adlandırılır. Bir grafiğin ortak olasılık dağılımının yapısını elde etmek için düğümleri bir araya getiren metoda kümeleme metodu denir. İlk olarak, Lauritzen ve Spiegelhalter² tarafından tanıtılmıştır. Bu kümeleme metodu grafikteki kümeler ile

ilişkilendirilmiş koşullu olasılıkların hesaplanması temeline dayanır.

Kümeleme metotları, verilen bir grafiğin ortak olasılık dağılımının yapısı ile ilgilenir. İlk olarak, orijinal grafikteki kümeler elde edilir. Daha sonra, birçok koşullu olasılık dağılımları hesaplanarak, her bir kümenin ortak olasılık dağılımı elde edilir. Verilen bir düğümün koşullu olasılığı, düğümü kapsayan herhangi bir kümenin ortak olasılık dağılımı marjinalleştirilerek elde edilecektir.

Bu çalışmada amaç, kesikli değişkenler için Bayes ağları tanıtmak ve bu konuda verilen bir problemin nasıl çözüleceğini ve yorumlanacağını açıklamaktır. Bayes ağlarda, olasılıkların hesaplanmasını kolaylaştırmak için, koşullu olasılıklar kullanılmıştır. Ayrıca, değişkenler arasındaki neden-etki ilişkilerine değinilmiş ve böylece bu değişkenler arasındaki ilişki modellenmiştir.

Meme kanseri özellikle gelişmiş ülkelerde kadınlarda görülme sıklığı en yüksek olan kanserdir. Bu durum Türkiye için de geçerlidir. Türkiye'de kadınlar arasında kanser ölümleri bakımından birinci sırada yer almaktadır.³ Türkiye'de her yıl 30 bin kadının meme kanserine yakalandığı söylenebilir. Türkiye'de tüm ölüm nedenlerine bakıldığında; kanserlerin hem erkeklerde hem de kadınlarda sırasıyla %15,04 ve %10,74 oranlarıyla ikinci sırada yer aldığı görülmektedir. Türkiye'de kadınlarda en sık görülen kanser türü olduğundan erken tanı ve tedaviye yönelik çalışmalar oldukça önem taşımaktadır.

İstatistiksel metotların, tıbbi tanılara uygulama alanlarında Bayes ağlar önemli bir yer tutar. Bayes ağlar, gözlenmiş olan belirtilerden nedenleri ortaya çıkarmak için, tıp ile ilgili birçok çalışmada birçok alana uygulanabilir.⁴⁻⁶

Bu amaçla uygulamada, Ankara ilinde meme kanserine yakalanmış 74 kadın üzerinde bir anket çalışması yapılmıştır. Bu çalışmada, ele alınan değişkenler için en iyi model UnBBayes paket programı kullanılarak elde edilmiştir. Elde edilen Bayes ağ modeli için kümeleme algoritmaları uygulanmıştır. Ayrıca, modele uygun analizlerin yapılmasında HUGIN paket programından yararlanılmıştır. Böylelikle, ele alınan değişkenlerin kümelmesi

gösterilecektir. Bu kümeleme sonucunda, meme kanseri teşhisinin erken yapılmasının önemi, memesinde tumor veya kitlenin mamografi ile tespitinin önemi, protez veya yapay meme kullanımının önemi gibi bazı önemli sonuçlara yer verilmiştir.

BAYES AĞ MODELLERİ

Grafik, yön verilmiş ve döngü olmama durumlarının her ikisini de sağlarsa grafiğe yön verilmiş döngüsel olmayan grafik denir (Şekil 1).^{2,7,8}

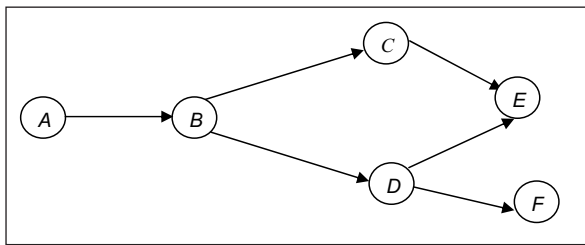
Yön verilmiş bir grafikte ebeveyn/çocuk ilişkilerinden söz edilir. A 'dan B 'ye bir kenar varsa A , B 'nin ebeveyni ve B , A 'nın çocuğudur denir ($A \rightarrow B$). Dede/torun ilişkisi, ebeveyn/çocuk ilişkisinin uzantısıdır. A , B 'nin ebeveyni ve B , C 'nin ebeveyni ise ($A \rightarrow B \rightarrow C$), A , C 'nin dedesi ve C , A 'nın torunudur denir. Aile, düğüm ve düğümün ebeveynlerinden oluşan bir kümedir.^{2,7,8}

Örneğin; Şekil 1'de, B , A 'nın çocuğu; A , B 'nin ebeveynidir. C ve D , B 'nin çocukları; B , C ve D 'nin ebeveynleridir. B , C , D , E ve F , A 'nın torunları; A , B , C , D , E ve F 'nin dedesidir. E düğümünün ailesi, $\{C, D, E\}$ düğümleridir.^{2,7,8}

Bir Bayes ağ modeli veya Bayes ağ, yön verilmiş döngüsel olmayan grafik, $P = \{p(x_1/Eb(m)_1), \dots, p(x_n/Eb(m)_n)\}$ n tane koşullu olasılık dağılımların kümesidir. Eb_i , grafikte X_i düğümünün ebeveynlerinin kümesidir. P kümesi, ilişkilendirilmiş ortak olasılık dağılımı tanımlar.^{2,7,8}

$$p(x) = \prod_{i=1}^n p(x_i / Eb(m)_i) \quad (1)$$

Bayes ağların ortak olasılık dağılımının çarpanlarına ayrılması yön verilmiş döngüsel olmayan grafikten kolaylıkla elde edilebilir. Şekil 2'deki gi-



ŞEKİL 1: Yön verilmiş döngüsel olmayan grafik.

bi verilen bir yön verilmiş döngüsel olmayan grafik düşünülün. Tüm değişkenler $\{A, B, C, D, E, F, G\}$ 'nin iki sonuçlu olduğu yani her bir değişkenin iki mümkün değer aldığı varsayılınsın.

Bu grafik verildiğinde, bir Bayes ağ Eşitlik 2'de ortaya çıkan koşullu olasılık dağılımlarının kümesi belirlenerek tanımlanmıştır.

$$p(a,b,c,d,e,f,g) = p(a).p(b).p(c/a).p(d/a,b).p(e).p(f/d).p(g/d,e) \quad (2)$$

BAYES AĞLARI ÖĞRENME

Modelin bağımlılık yapısı ve modele ait koşullu olasılık dağılımları uzman kişiler tarafından sağlanabilir. Birçok uygulamada, bu bilgi elde edilemeyebilir. Buna ek olarak, farklı uzman kişiler farklı görüşler verecektir. Bu da ele alınan problemde bilgi karmaşasına neden olacaktır. Bu nedenle, bağımlılık yapısı ve koşullu olasılık dağılımları veriden tahmin edilecektir. Bağımlılık yapısı ve koşullu olasılık dağılımlarının veriden tahmin edilme süreci *öğrenme* olarak adlandırılır.⁹ Bu çalışmada amaç var olan veritabanını kullanarak, en iyi modeli elde etmektir.

Öğrenmenin iki farklı türü vardır: Yapıyı öğrenme, parametre öğrenme.

Yapıyı öğrenme: Grafikte yer alan bağlantıları belirleyen bir bağımlılık yapısını öğrenmeyi ifade eder. Bir Bayes ağda, yön verilmiş döngüsel olmayan grafik, yapı olarak adlandırılır.⁹

Parametre öğrenme: Koşullu olasılık dağılımlarının parametrik yapısını öğrenmeyi ifade eder. Bir Bayes ağda, yön verilmiş döngüsel olmayan grafiğe ait koşullu olasılık değerleri parametreler olarak adlandırılır.⁹

BAYES AĞ ARAMA ALGORİTMALARI

Yüksek kaliteli Bayes ağların en iyi olanını seçmek için kullanılır. Değişkenlerin çok küçük sayıları için bile tüm mümkün ağların sayısını denemek hemen hemen imkansızdır. Tüm mümkün farklı ağ yapılarının sayısı çok büyük olduğu durumda, ağ yapılarının her birini değerlendirmek imkansızdır. Bu bölümde, veri ve ön bilgi verildiğinde, en yük-

sek kaliteli Bayes ağı bulmak için iki arama algoritması verilecektir.

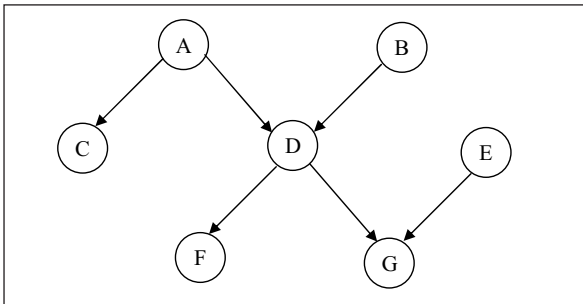
K2-Algorithması: Cooper ve Herkovits,⁷ yüksek kaliteli Bayes ağ bulmak için bir algoritma sunar ve K2-algoritması olarak adlandırılır. Bu algoritmada ilk olarak, bağlantıların olmadığı en basit ağ ele alınır. Ayrıca, düğümlerin sıralı olduğu varsayılır. Her bir değişken X_i için, X_i 'den daha küçük numaralı düğüm, ebeveyn kümesi olan $Eb(m)_i$ 'ye eklenir. Bu süreç, yeni düğümlerin eklenmesiyle tam ağ modeli elde edilene kadar tekrarlanacaktır.

B-Algorithması: Düğümlerin herhangi bir sıralanmasını gerektirmeyen diğer bir algoritma B-algoritmasıdır. Bu algoritma, K2-algoritmasında olduğu gibi boş ebeveyn kümesiyle başlar. Her bir adımda, döngü elde edilmeksizin yeni bağlantı eklenecektir. Süreç, mümkün veya tam ağ elde edilene kadar tekrar edilecektir.⁹

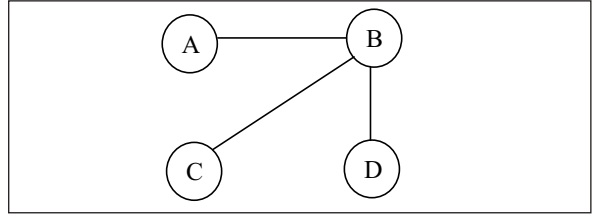
Bu çalışmada, en iyi modelin elde edilmesinde ve veritabanından yapıyı öğrenmede UnBBayes API 2.2 paket programı kullanılmıştır.^{9,10} Bu program, hem K2 ve B algoritmasını da içermektedir. Bu çalışmada, sadece K2 algoritması kullanılmıştır. En iyi Bayes ağ modeli belirlendikten sonra, yön verilmemiş bir grafik elde edilecek ve grafiğin üçgenleştirilmiş bir grafik olup olmadığını belirlemek için En büyük sayıyı arama doldurma algoritması kullanılacaktır.

Yön verilmemiş grafik: Bir grafikteki kenarlar arasında sadece yön verilmemiş kenarlar varsa, bu grafiğe yön verilmemiş grafik denir.¹¹ Şekil 3'te, yön verilmemiş bir grafiğe örnek verilmiştir.

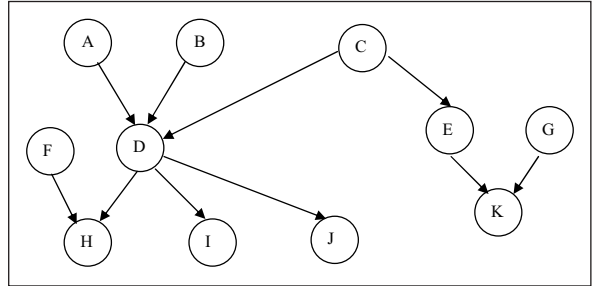
Moral grafik: Bir yön verilmiş grafikte düğümlerin tüm çiftleri arasında ve her düğümün ebeveynlerinin arasında yön verilmemiş kenarlar eklenerek



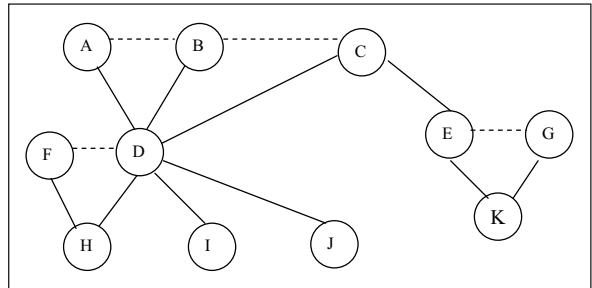
ŞEKİL 2: Yedi düğümden oluşan bir Bayes ağı.



ŞEKİL 3: Yön verilmemiş grafik.



ŞEKİL 4: On bir düğümden oluşan yön verilmiş bir grafik.



ŞEKİL 5: Şekil 4'den elde edilen moral grafik.

elde edilen yön verilmemiş grafiğe moral grafik denir.⁹⁻¹¹ Şekil 4'de, on bir düğümden oluşan bir yön verilmiş grafik verilmiştir.

Şekil 4'de verilen grafiğe ilişkin moral grafik Şekil 5'de verildiği gibidir. Bu grafikte, {A,B}, {B,C}, {D,F} ve {E,G} çiftlerinin her biri ortak bir çocuğa sahiptir.

Üçgenleştirilmiş grafik: Bu grafikler (Triangulated graphs), yön verilmemiş grafiklerin özel bir çeşididir. Döngüde iki düğüm arasındaki bağlantıya giriş adı verilir. Uzunluğu üç olan döngüler bir girişi içermezler ve bu döngüler üçgenler olarak adlandırılır.^{10,11}

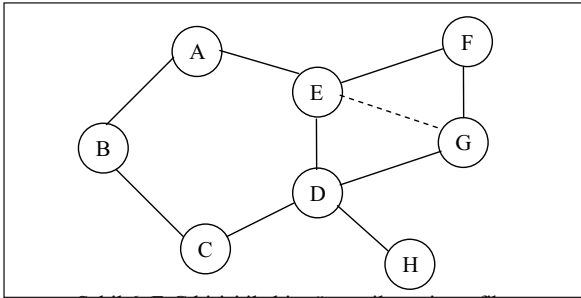
Bu şekilde, E-G girişi, E-F-G-D-E döngüsünü, E-F-G-E ve E-G-D-E olmak üzere iki döngüye par-

çalamaktadır. Şekil 6'da verilen grafik üçgenleştirilmiş değildir. Çünkü A-B-C-D-E-A döngüsü bir kirişe sahip değildir.

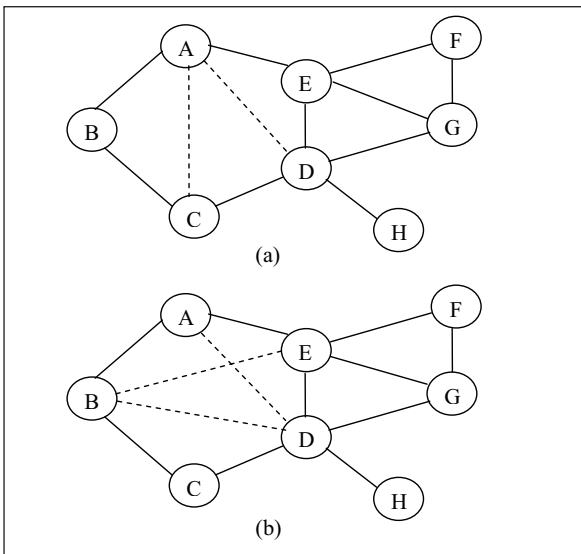
Uzunluğu dört veya daha fazla olan her döngü en az bir kirişe sahip ise, yön verilmemiş grafiğe üçgenleştirilmiş grafik denir. Bir grafik üçgenleştirilmiş değilse, döngüleri ortadan kaldırmak için kirişler eklenerek, grafik üçgenleştirilmiş hale getirilir.¹¹

Şekil 7, Şekil 6'da verilen grafiğin üçgen hale getirilmesinde kullanılan iki farklı durumu göstermektedir.

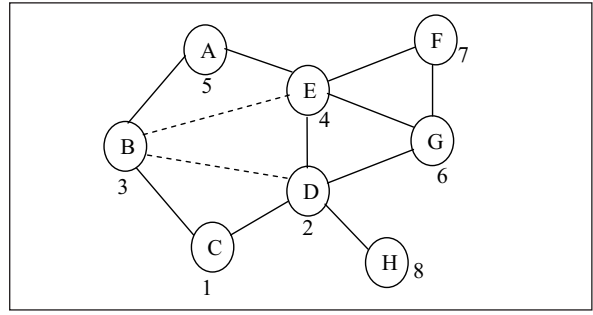
En büyük sayıyı doldurma algoritması: Bu algoritma, yön verilmemiş bir grafiğin düğümlerinin numaralanmasını verir. Ayrıca, ele alınan düğüm, o düğüme ilişkin alt düğümlerin bir araya getirilerek



ŞEKİL 6: E-G kirişi ile bir yön verilmemiş grafik.



ŞEKİL 7: Şekil 6'daki grafiğin üçgen hale getirilmiş iki farklı gösterimi.



ŞEKİL 8: Şekil 7 için En Büyük Sayıyı Arama algoritması kullanılarak düğümlerin kusursuz sıralaması ve üçgenleştirilmiş grafik.

(çocuk, torun, komşu... gibi düğümler) yapılan sıralamadır. Orijinal grafik üçgenleştirilmiş bir grafik ise, yapılan bu numaralama, kusursuz bir numaralamadır.^{10,11} Bir yön verilmemiş grafik üçgenleştirilmiş ise, bu grafik kusursuz numaralamaya olanak sağlar. Bir yön verilmemiş grafikte kusursuz numaralama yapmak mümkün ise, bu grafik üçgenleştirilmiştir.

Bu algoritma, yön verilmemiş bir grafiğin düğümlerinin numaralanmasını verir. Orijinal grafik üçgenleştirilmiş bir grafik ise, yapılan bu numaralama, kusursuz bir numaralamadır.^{1,10}

Şekil 8'de üçgenleştirilmiş olmayan bir grafik verilmiştir. Şekil 8(b)'de ise (a)'da verilen grafik A-C düğümleri arasına kiriş eklenerek üçgenleştirilmiş bir grafik elde edilmiştir. Ayrıca, Şekil 8(b)'de En Büyük Sayıyı Arama algoritması kullanılarak düğümlerin kusursuz numaralanması verilmiştir.

BAYES AĞLARDA KÜMELEME METOTLARI

Herhangi bir grafik verildiğinde, düğümlerin ortak özellikleri bir araya getirilerek, küme grafikleri yardımı ile birçok düğüm gruplanabilir. Bu *kümeleme* olarak adlandırılır. Kümeleme, yeni grafikler elde edilmesine imkan sağlar ve orijinal grafiğin yapısal özellikleri değişmeden aynen kalır.^{2,10,12}

Bir grafikteki düğümlerin bir araya getirilmesi ile elde edilen düğümlerin kümesi, kümeler olarak adlandırılır. Kümeleme metotları, bir grafiğin ortak olasılık dağılımın yapısını elde etmek için kümeleri bir araya getirir. Bu metot, kümeleme metodu olarak adlandırılır. Kümeleme metotları, verilen

bir grafiğin ortak olasılık dağılımının yapısı ile ilgilenir. İlk olarak, orijinal grafikteki klikler (kümeleler) elde edilir. Daha sonra, birçok koşullu olasılık dağılımları hesaplanarak, her bir kliğin ortak olasılık dağılımı elde edilir. Verilen bir düğümün koşullu olasılığı, düğümü kapsayan herhangi kliğin ortak olasılık dağılımı marjinalleştirilerek elde edilecektir.^{2,10,12}

BAYES AĞ MODELLERİNDE KÜMELEME ALGORİTMASI

Bayes ağlarda ortaya çıkan sonuçlar kanıt yayılımı (evidence of propagation) veya basitçe yayılım olarak adlandırılır.^{10,13} Kanıt yayılımı, yeni elde edilebilir kanıta göre, değişkenlerin olasılık dağılımlarının güncelleştirilmesinden oluşur. Örneğin; kanıt (yani semptomlar) verildiğinde, ilgilenilen değişkenler (yani, hastalıklar) kümesinin her eleman için koşullu dağılımını hesaplamak gerekir.

Bayes ağ modellerinde kanıt yayılımını kurmak için kümeleme algoritması tanıtılacaktır. $\{X_1, X_2, \dots, X_n\}$ değişkenler kümesi üzerinde tanımlanan Bayes ağ modeli verildiğinde, ortak olasılık dağılımı eşitlik (1)'de verildiği gibi çarpanlarına ayrılacaktır.

Burada amaç, bir yön verilmiş döngüsel olmayan grafiği, üçgenleştirilmiş yön verilmemiş grafiğe dönüştürmek ve elde edilen grafik üzerinde kümeleme algoritması kullanmaktır. Yön verilmiş grafikler, ilk olarak bağlantıların yönleri kaldırılarak moralize edilir ve böylece sonuçlanan grafik üçgenleştirilmiş olur. Bunun için, her bir aile en az bir klikte (kümede) olacak şekilde üçgenleştirilmiş yön verilmemiş grafik elde edilir. Bu özellik, yön verilmemiş grafik için potansiyel gösterimi tanımlamaya imkan sağlar. Sonuçta, Bayes ağlarda kümeleme algoritması uygulanarak Bayes ağlarda kanıt yayılımı için problem çözülür.¹⁰

UYGULAMA

Bu makalede, Türkiye'de Ankara ilinde meme kanseri teşhisi konmuş ve tedavisi yeni tamamlanmış 74 kadın ile anket çalışması yapılmıştır. Yapılan anket çalışmasından en iyi Bayes ağ modelini belirlemek amacıyla toplam 13 değişken seçilmiştir.

Seçilen bu değişkenler ve düzeyleri aşağıda verilmiştir. Her bir değişken sadece iki düzeye sahiptir ve düzeyler küçük harfler ile gösterilecektir. Ayrıca, seçilen değişkenlerin her birine karşılık gelen harfler aşağıda gösterilmiştir. Uygulama süresince bu harfler kullanılacaktır.

A : Hastanın meme kanseri teşhisi öncesi kontrollerini yaptırap yaptırmaması

B : Lenf bezlerinde atlama olup olmaması

C : Memesinde herhangi bir kitle veya tümörün kendi tarafından fark edilip edilmemesi

D : Memesinde kanserin mamografide tespit edilip edilmediği

E : Tedavi yöntemlerinden memenin tümünün alınıp alınmaması

F : Memenin bir kısmının alınıp alınmaması

G : Kemoterapi görüp görmemesi

H : Radyoterapi görüp görmemesi

I : Hormonsal terapi görüp görmemesi

J : Protez meme kullanıp kullanmaması

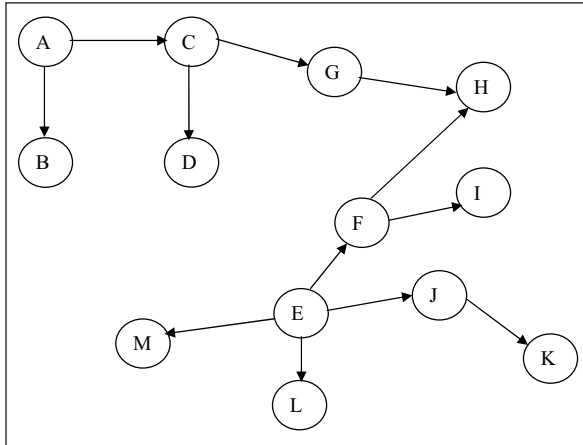
K : Protez meme kullanmaktan memnun olup olmaması

L : Yapay meme yaptırmak isteyip istememesi

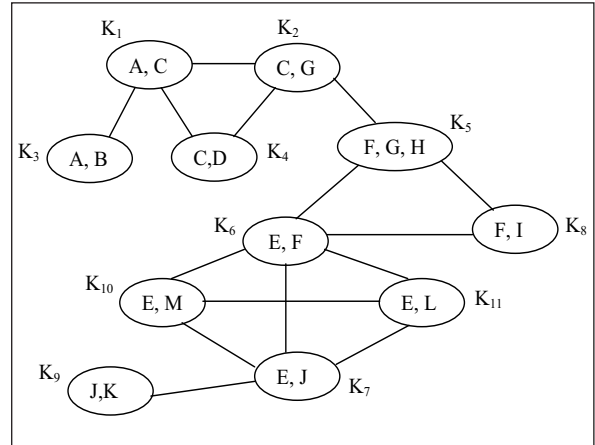
M : Meme kanserinden önceki iki yıl içinde şiddetli bir üzüntü geçirip geçirmediği

Bayes ağları öğrenme için K2 algoritması kullanılmıştır. Bu algoritma Cooper ve Herkovits,⁷ yüksek kaliteli Bayes ağ bulmak için bir algoritma sunar ve K2-algoritması olarak adlandırılır.¹⁰ Bu çalışmada, en iyi modelin elde edilmesinde ve veritabanından yapıyı öğrenmede UnBBayes API 2.2 paket programı kullanılmıştır. En iyi Bayes ağ modeli belirlendikten sonra, yön verilmemiş bir grafik elde edilecek ve grafiğin üçgenleştirilmiş bir grafik olup olmadığını belirlemek için En büyük sayıyı arama doldurma algoritması kullanılacaktır. Bunun için HUGIN paket programından yararlanılmıştır.¹³

Şekil 9'daki grafiğe göre, meme kanseri teşhisi öncesi kontrollerini yaptırap yaptırmaması, hem lenf bezlerinde atlama olup olmamasını, hem de memesinde kanserin kendisi tarafından fark edilip



ŞEKİL 9: On üç değişkenden oluşan en iyi Bayes ağ modeli.



ŞEKİL 10: Yön verilmiş döngüsel olmayan grafik için küme grafiği.

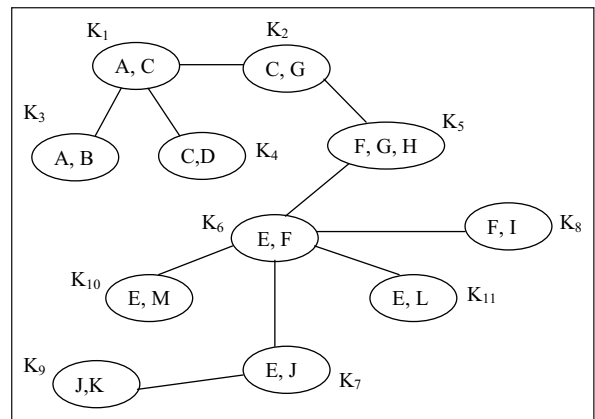
edilmemesini etkiler. Kısaca, A düğümü C ve D düğümlerine neden olmaktadır. Tedavi yöntemlerinden memenin tümünün alınıp alınmaması, hem yapay meme yaptırmak isteyip istememesi, hem de meme kanserinden önceki iki yıl içinde şiddetli bir üzüntü geçirip geçirmemesini etkiler.

İlk olarak, Bayes ağ modeli için ortak olasılık dağılımının çarpanlarına ayrılması aşağıda verilmiştir:

$$p(a, b, c, d, e, f, g, h, i, j, k, l, m) = p(a) p(b/a) p(c/a) p(d/c) p(g/c) p(h/g, f) p(i/f) p(f/e) p(j/e) p(k/j) p(m/e) p(l/e) \quad (4)$$

Şekil 9'daki Bayes ağ modeli için elde edilen kümeler Tablo 1'de verilmiştir.

TABLO 1: Şekil 9'daki Bayes ağ modeli için elde edilen kümeler.			
Küme (K)	Kümeye ait olan düğüm sayısı	Kümeye ait olan düğümler	Tablo büyüklüğü
K ₁	2	(A, C)	4
K ₂	2	(C, G)	4
K ₃	2	(A, B)	4
K ₄	2	(C, D)	4
K ₅	3	(F, G, H)	8
K ₆	2	(E, F)	4
K ₇	2	(E, J)	4
K ₈	2	(F, I)	4
K ₉	2	(J, K)	4
K ₁₀	2	(E, M)	4
K ₁₁	2	(E, L)	4
Toplam küme tablo büyüklüğü			48



ŞEKİL 11: Şekil 9 için elde edilen küme ağacı.

Şekil 9 için elde edilen küme grafiği Şekil 10'da gösterilmiştir.

Küme grafiğinden bağlantılar keyfi olarak kırıldığından dolayı, aynı yön verilmemiş grafik ile ilişkili birden fazla küme ağacı elde edilebilir. Bu küme ağaçlarından biri Şekil 11'de verilmiştir.

Ayrıca elde edilen bağlantılı kümeler, kümelerin ayıraç düğümleri ve tablo büyüklükleri Tablo 2'de verilmiştir.

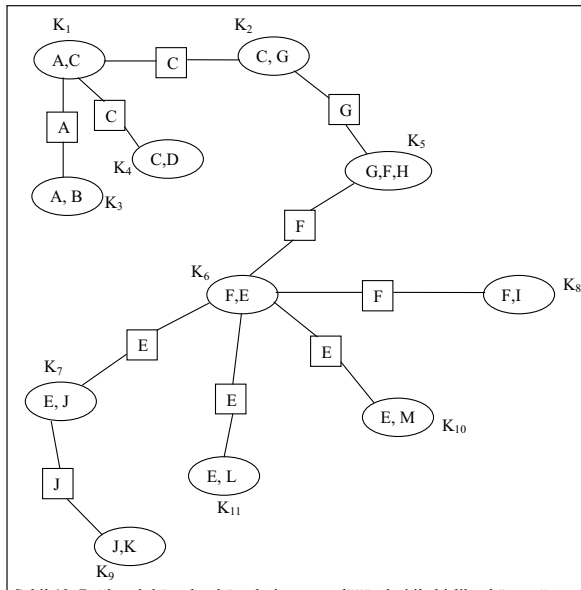
Küme ağacını ayıraç düğümleri kullanarak grafiksel olarak Şekil 12'de gösterilmiştir.

Kümeleme algoritmasının tüm adımları sırasıyla uygulandığında Şekil 12'ye göre elde edilen sonuçlar aşağıda verilmiştir.

Meme kanseri teşhisi öncesi kontrollerini yaptırmayıp yaptırmaması (A) ve memesinde kanserin

TABLO 2: Bağlantılı kümeler, kümelerin ayıraç düğümleri ve tablo büyüklükleri.

Bağlantılı kümeler	Ayıraç düğümleri	Tablo büyüklüğü
K ₁ ve K ₂	{C}	2
K ₁ ve K ₃	{A}	2
K ₁ ve K ₄	{C}	2
K ₂ ve K ₅	{G}	2
K ₅ ve K ₆	{F}	2
K ₆ ve K ₇	{E}	2
K ₆ ve K ₈	{F}	2
K ₇ ve K ₉	{J}	2
K ₆ ve K ₁₀	{E}	2
K ₆ ve K ₁₁	{E}	2

**ŞEKİL 12:** Bağlantılı kümeler, kümelerin ayıraç düğümleri ile birlikte küme ağacı.

kendisi tarafından fark edilip edilmeme (C) düğümleri K₁ kümesine atanmıştır. Tedavi yöntemlerinden kemoterapi alıp almaması (G) düğümü K₂ kümesine atanmıştır. Lenf bezlerinde atlama olup olmama (B) düğümü K₃ kümesine atanmıştır. Memenizde kanserin mamografi ile tespit edilip edilememe (D) düğümü K₄ kümesine atanmıştır. Tedavi yöntemlerinden radyoterapi alıp almama düğümü (H) K₅ kümesine atanmıştır. Tedavi yöntemlerinden memenin tümünün alınması (E) ve memenin bir kısmının alınması (F) düğümü K₆ kü-

mesine atanmıştır. Protez meme kullanıp kullanmama (J) düğümü K₇ kümesine atanmıştır. Tedavi yöntemlerinden hormonal terapi alıp almama (I) düğümü K₈ düğümüne atanmıştır. Protez meme kullanmaktan memnun olup olmama (K) düğümü K₉ kümesine atanmıştır. Meme kanserinden önceki iki yıl içinde şiddetli bir üzüntü geçirip geçirmeme (M) düğümü K₁₀ kümesine atanmıştır. Yapay meme yaptırmak isteyip istememe (L) düğümü K₁₁ kümesine atanmıştır.

Böylece, bu kümelere ilişkin olasılık değerleri aşağıdaki gibi gösterimle yazılabilir.

$$\begin{aligned}
 \psi_1(a,b) &= p(a)p(b/a) \\
 \psi_2(a,c) &= p(c/a) \\
 \psi_3(c,g) &= p(g/c) \\
 \psi_4(g,f,h) &= p(h/g,f) \\
 \psi_5(e,f) &= p(e)p(f/e) \\
 \psi_6(e,m) &= p(m/e) \\
 \psi_7(e,l) &= p(l/e) \\
 \psi_8(e,j) &= p(j/e) \\
 \psi_9(j,k) &= p(k/j) \\
 \psi_{10}(f,i) &= p(i/f) \\
 \psi_{11}(c,d) &= p(d/c)
 \end{aligned} \tag{5}$$

Eşitlik 5'in sağ tarafındaki olasılık değerleri kullanılarak her bir potansiyel fonksiyonun sayısal değerleri Tablo 3'deki çizelgelerde verilmiştir. Bu çizelgelerde "0" her bir düğümün hayır cevabına, "1" her bir düğümün evet cevabına karşılık kullanılmıştır.

Bu çalışmada, elde edilen küme grafiğine göre, toplam 11 tane küme elde edilmiştir. Sonuç olarak, her bir değişkenin hangi kümeye ait olduğu belirlenmiştir. Örneğin; meme kanseri teşhisi kontrollerini yaptırıp yaptırmama (A) ve lenf bezlerinde atlama olup olmaması (B) değişkenlerin meydana getirdiği ortak olasılık dağılımı, yani K₁ kümesi elde edilmiştir. Meme kanseri teşhisi kontrollerini yaptırması (A=1) ve lenf bezlerinde atlama olmasına (B=1) ilişkin elde edilen ortak olasılık dağılımı 0,099 bulunmuştur. Sonuç olarak, 74 kadın üzerinde yapılan anket çalışmasına göre meme kanseri öncesi kontrollerini yaptıran kadınlarda, lenf bezlerinde atlama olasılığı düşüktür. Ancak, meme

TABLO 3: Kümeleme algoritmasında kümelere karşılık gelen olasılık değerleri.

a	b	$\Psi_1(a,b)$	a	c	$\Psi_2(a,c)$	c	g	$\Psi_3(c,g)$	e	f	$\Psi_4(e,f)$	e	m	$\Psi_5(e,m)$
0	0	0,2803	0	0	0,5926	0	0	0,5909	0	0	0,1357	0	0	0,1428
0	1	0,4675	0	1	0,1667	0	1	0,1964	0	1	0,1111	0	1	0,4333
1	0	0,1899	1	0	0,4077	1	0	0,4091	1	0	0,7277	1	0	0,8571
1	1	0,0990	1	1	0,8333	1	1	0,8036	1	1	0,0256	1	1	0,5667
e	l	$\Psi_6(e,l)$	e	j	$\Psi_7(e,j)$	j	k	$\Psi_8(j,k)$	f	ı	$\Psi_9(f,ı)$	c	d	$\Psi_{10}(c,d)$
0	0	0,2273	0	0	0,8971	0	0	0,8071	0	0	0,8971	0	0	0,3913
0	1	0,4590	0	1	0,5455	0	1	0,1555	0	1	0,5455	0	1	0,8947
1	0	0,7727	1	0	0,1029	1	0	0,1929	1	0	0,1029	1	0	0,6087
1	1	0,5410	1	1	0,4545	1	1	0,8444	1	1	0,4545	1	1	0,1053
g	f	h	$\Psi_1(g,f,h)$											
0	0	0	0,9524											
0	0	1	0,0476											
0	1	0	0,4000											
0	1	1	0,6000											
1	0	0	0,4082											
1	0	1	0,5918											
1	1	0	0,6250											
1	1	1	0,3750											

kanseri öncesi kontrollerini yaptırmayan kadınlarda, lenf bezlerinde atlama olasılığı 0,4675 olarak elde edilmiştir.

Elde edilen modele ilişkin bazı sonuçlar aşağıda verilmiştir:

Meme kanseri teşhisi öncesi kontrollerini yaptırmayan kadınların, lenf bezlerine atlama olasılığı 0,4675 olarak elde edilmiştir. Ancak, meme kanseri teşhisi öncesi kontrollerini yaptıran kadınların, lenf bezlerine atlama olasılığı (0,0990) azalmaktadır.

Memesinde herhangi bir kitle veya tümörün kendi tarafından farkedilmediğinde, memesinde kanserin mamografide tespit edilmesine ilişkin ortak olasılık dağılımı incelendiğinde bu olasılık 0,8947 elde edilmiştir.

Tedavi yöntemlerinden memenin tümünün alındığı bilindiğinde, yapay meme yaptırmak isteyen kadınlar için olasılık 0,5410 olarak elde edilmiştir. Ancak, memenin tümünün alındığı bilindiğinde, yapay meme yaptırmak istemeyen ka-

dınlar için olasılık değeri daha yüksektir (0,7727). Tedavi yöntemlerinden memenin bir kısmının alınmaması, kemoterapi görmemesi ve radyoterapi görmeme olasılığı 0,9524 olarak elde edilmiştir. Tedavi yöntemlerinden memenin bir kısmının alınması, kemoterapi görmesi ve radyoterapi görme olasılığı (0,3750) daha düşüktür.

Kadınların memenin tümü alındığında protez meme kullananların olasılığı 0,4545'dir. Protez meme kullanıp, protez meme kullanmaktan memnun olanların olasılığı 0,8444 olarak bulunmuştur.

Tedavi yöntemlerinden memesinin tümü alınan kadınlarda, meme kanserinden önceki iki yıl içinde şiddetli bir üzüntü geçirme olasılığında 0,5667 olarak elde edilmiştir.

TARTIŞMA VE SONUÇ

Elde edilen sonuçlara göre, kadınların meme kanseri, kendi kendine meme muayenesi konusunda bilgi düzeyleri düşük olduğu gözlenmiştir. Sağlık personeli eğitim kurumları ile beraber meme kan-

seri erken tanı ve muayene yöntemleri hakkında kadınları bilinçlendirmeye yönelik eğitim programlarının verilmesinin önemi üzerinde durulmuştur.¹⁴ Meme kanserinin önceden teşhis edilmesi, ortaya çıkacak bazı sonuçları azaltmaktadır. Meme kanserinde mamografi ile kitle ya da tümörün tespit edilebilmesi de yüksek olarak bulunmuştur. Meme kanseri teşhisi kontrollerini önceden yaptıran kadınlarda, memenin tümünün alınması, tedavi yöntemlerinden kemoterapi, radyoterapi veya hormonal terapi alma olasılıkları

düşmektedir. Meme kanserine yakalanmadan önceki 2 yıl içinde şiddetli bir üzüntü geçirmeleri de oldukça önemli bulunmuştur. Tedavi yöntemlerinden memesinin tümü alınan kadınlarda, protez meme kullanmama olasılığı yüksek iken, yapay meme kullanmak isteyen kadınların olasılığı daha yüksek bulunmuştur.

Sonuç olarak, Bayes ağ modelleri, tıp ile ilgili alanlar da oldukça önem taşımaktadır. Bayes ağ modellerinde neden-etki değişkenlerinin belirlenmesi de önem taşımaktadır.

KAYNAKLAR

1. Spiegelhalter DJ, Dawid AP, Lauritzen SL, Cowell RG. Bayesian analysis in Expert systems. *Stat Sci* 1993;8(3):219-83.
2. Heckerman D, Geiger D, Chickering D.M. Learning Bayesian Networks: The combination of knowledge and statistical data. *Machine Learning* 1995;20(3):197-243.
3. Akyolcu N, Altun Uğraş G. [Breast self-examination: how important is it in early diagnosis?]. *Journal of Breast Health* 2011;7(1): 4-10.
4. Gadewadikar J, Kuljaca O, Agyepong K, Sarigul E, Zheng Y, Zhang P. Exploring Bayesian networks for medical decision support in breast cancer detection. *African Journal of Mathematics and Computer Science Research* 2010;3(10):225-31.
5. Velikova M, Ferreira NC, Lucas P. Bayesian network decomposition for modeling breast cancer detection. *Proceedings of the 11th Conference on Artificial Intelligence in Medicine (AIME)*, LNA 4594. Amsterdam: AIME; 2007. p.346-50.
6. Sierra B, Inza I, Larrañaga P. Medical bayes networks. In: Brause RW, Hanisch E, eds. *Medical Data Analysis Lecture Notes in Computer Science* 1933. 1st ed. Frankfurt: Springer; 2000. p.4-14.
7. Cooper F, Herskovits E. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning* 1992;9(4):309-47.
8. Lauritzen SL, Spiegelhalter DJ. Local computations with probabilities on graphical structures and their application to expert systems. *J R Statist Soc B* 1988;50(2):157-224.
9. Neapolitan RE. *Learning Bayesian Networks. Parameter Learning: Binary Variables*. 1st ed. London: Prentice Hall; 2004. p.286-364.
10. Castillo E, Gutierrez JM, Hadi AS. Exact propagation in probabilistic network models. *Expert Systems and Probabilistic Network Models*. 1st ed. New York: Springer-Verlag; 1997. p.317-87.
11. Buntine W. Operations for learning with graphical models. *J Artif Intell Res* 1994; 2(1):159-225.
12. Xiang Y. Belief updating and cluster graphs. *Probabilistic Reasoning in Multiagent Systems: A Graphical Models Approach*. 1st ed. Cambridge: Cambridge University Press; 2002. p.59-37.
13. Jensen FV. *Causal and Bayesian Networks, Building Models. An introduction to Bayesian Networks*. 1st ed. London: UCL Press Ltd.; 1996. p.50-5.
14. Koçyiğit O, Erel S, Kismet K, Kılıçoğlu B, Sabuncuoğlu MZ, Akkuş MA. [Knowledge about breast cancer, mammography and breast self examination in women who applied to outpatient clinics: a study conducted in the city center]. *Nobel Medicus* 2011;7 (2):19-25.