HEYELAN DUYARLILIK ANALİZİNDE Kİ-KARE TESTİNE DAYALI FAKTÖR SEÇİMİ

Taşkın KAVZOĞLU[1], Emrehan Kutluğ ŞAHİN[2], İsmail ÇÖLKESEN[3]

[1]Prof. Dr., Gebze Yüksek Teknoloji Enstitüsü, Jeodezi ve Fotogrametri Müh. Bölümü, 41400, Gebze, Kocaeli, kavzoglu@gyte.edu.tr
[2]Araş. Gör., Gebze Yüksek Teknoloji Enstitüsü, Jeodezi ve Fotogrametri Müh. Bölümü, 41400, Gebze, Kocaeli, e.sahin@gyte.edu.tr
[3]Araş. Gör., Gebze Yüksek Teknoloji Enstitüsü, Jeodezi ve Fotogrametri Müh. Bölümü, 41400, Gebze, Kocaeli, icolkesen@gyte.edu.tr

ÖZET

Heyelan, insanların kontrolü dışında gerçekleşen mal ve can kaybına neden olan bir doğal afettir. Heyelan duyarlılık haritalarının üretilmesi ve heyelan duyarlı alanların belirlenmesi mühendisler ve karar vericiler için büyük önem arz etmektedir. Üretilecek haritaların doğrulukları can ve mal kayıplarının azaltılmasında büyük öneme sahiptir. Bu çalışmada Trabzon ili Maçka ilçesi için heyelan duyarlılık haritası üretilmesi amacıyla 9 farklı faktör katmanı (litoloji, eğim, arazi kullanımı, bakı, yükseklik, topoğrafik ıslaklık indeksi, eğim uzunluğu, eğrilik profili ve çizgisellik) kullanılmıştır. 9 faktör seti içinden ki-kare testi kullanılarak optimum 3,4,5,6,7 ve 8 faktör setleri belirlenmiştir. Tüm parametrelerin heyelan duyarlılık haritası üretimi için kullanımı, korelasyona ve gereksiz bilgiye sahip veriler nedeni ile sonuç harita üzerinde doğruluğu azaltan etkiye neden olabilir. Bu olay boyutsallık problemi veya Hughes etkisi ile açıklanmaktadır. Bu çalışmada, heyelan duyarlılık haritası üretimi için son yıllarda literatürde yaygınlıkla kullanılan lojistik regresyon metodu kullanılmıştır. Altı modelden üretilen duyarlılık haritalarının performanslarının değerlendirilmesi amacıyla genel doğruluk değerlerine bakılmıştır. Ayrıca üretilen duyarlılık haritaları arasındaki farklılıklar ve istatistiksel anlamlılıkları McNemar's testi ile analiz edilmiştir. Çalışma sonucunda ki-kare testi yardımıyla belirlenen en iyi 5 faktör kombinasyonu (eğim, litoloji, arazi örtüsü, yükseklik ve eğim uzunluğu) kullanılarak en yüksek doğruluğa (%80,33) sahip heyelan duyarlılık haritası üretilmiştir. Diğer bir taraftan bu çalışma için bakı, topoğrafik ıslaklık indeksi, eğrilik profili ve çizgisellik faktörlerinin sınırlı etkiye sahip olduğu görülmüştür. Ayrıca, McNemar's test sonuçları da, artan veri setinin istatistiksel olarak anlamlı olmadığını göstermiştir. Elde edilen sonuçlar, çok sayıda faktör ile heyelan duyarlık haritası üretiminde otomatik veya akıllı yaklaşımlar kullanarak da parametre seçimi yapılabileceğini göstermektedir.

Anahtar Sözcükler: Heyelan duyarlılık haritası, Ki-kare testi, Özellik seçimi, McNemar's testi, Lojistik regresyon

ABSTRACT

A landslide is a destructive natural disaster affecting a large number of people and properties negatively. Production of landslide susceptibility maps and identification of landslide prone areas are important for engineering and decision makers. The accuracy of the susceptibility maps is of great importance in reducing life and property losses. In this study, in order to produce landslide susceptibility map for Macka District of Trabzon province nine factor maps (lithology, slope, land cover, aspect, elevation, topographic wetness index, slope length, profile curvature and lineaments) were used. Nine factor map combinations were applied to determine the optimum 3, 4, 5, 6, 7 and 8 sets by the chi-square test. When all available parameters are used, it is high likely that correlated and redundant information is considered, which may reduce the accuracy of a resulting map. This phenomenon can be explained by curse of the dimensionality or Hughes effect. In this study, logistic regression (LR) method, which has been widely used in the literature, was applied to generate landslide susceptibility maps for all delivered models. Overall accuracy estimated to evaluate performances of the landslide susceptibility maps produced by the six models. The model with five landslide related factors (slope, lithology, land cover, elevation and slope length) determined by the Chi-square produced the highest accuracy (80.33%) of all factor combinations. On the other hand, aspect, topographic wetness index, curvature profile and lineaments were found to be the least effective ones considering all combinations tested in this study. Moreover, McNemar's test results showed that increasing the dimension of data sets was found statistically insignificant. Above findings clearly indicate that parameter selection for landslide susceptibility mapping should be performed by using some intelligent or automatic approaches when a large number of factors are available.

Keywords: Landslide susceptibility maps, Chi-Square test, Feature selection, McNemar's test, Logistic regression

1. GİRİŞ

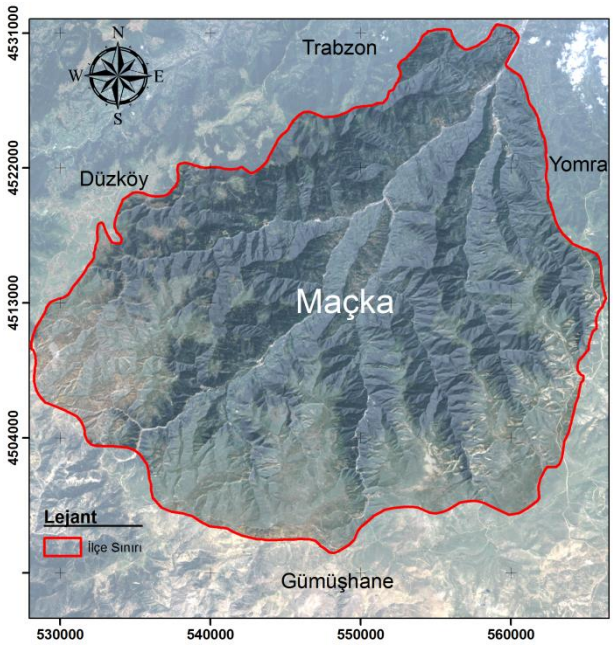
Heyelan, yamaç dengesinin bozulması sonucu yer çekiminin etkisiyle arazinin bir bölümünün yamaç eğilimi doğrultusunda hareket ederek şekil ve yer değiştirmesidir (Öztürk, 2002). Heyelan olayı etkili olduğu bölge üzerinde büyük çaplı deformasyonlar ile yerleşim yerleri, tarım arazileri, kara ve demir yolları üzerinde can ve mal kaybına yol açabilirler. Bu nedenle insan hayatının güvenliği sağlamak ile birlikte bölge ve ülke ekonomisine yaratacağı olumsuz etkileri önlemek amacıyla olası heyelan alanlarının tespiti önemlidir. Heyelan duyarlılık analizleri ile olası heyelan alanlarının tespitini yaparak tehlikeli ve riskli alanlarının belirlemeye ve heyelanın etkilerini azaltılması amaçlanmaktadır. Heyelan duyarlılık haritalarının üretilmesi, heyelana etki edebilecek faktörlerin belirlenerek kullanılan faktörlerin bir arada değerlendirilmesini içeren bir analiz sürecidir.

Heyelan duyarlılık haritalarının üretimi iki önemli adımı içermektedir. Bunlardan birincisi üretilecek duyarlılık haritasında heyelana etki edebilecek faktörlerin belirlenerek en uygun faktör kümesinin oluşturulması, bir diğeri ise harita üretiminde kullanılacak metodun tespitidir. Duyarlılık haritalarının üretiminde makine öğrenme (Kavzoğlu vd., 2012; Pourghasemi et al., 2013; Kavzoglu et al., 2014), istatistiksel tabanlı metotlar (Mancini et al., 2010; Regmi et al., 2014;Thiery et al., 2014) ve kullanıcı tabanlı metotlar (Yalcin, 2008; Şahin, 2012; Shahabi et al., 2014;) gibi birçok farklı algoritmalar kullanılmaktadır. Herhangi bir yaklaşımla üretilecek heyelan duyarlılık haritası için kullanılacak özellikler veya faktörlerin sayısı, niteliği ve içerdiği bilgi sonuç haritanın üretim doğruluğuna etkisi büyüktür. Fakat üretilecek duyarlılık haritasında çok fazla sayıda faktör kullanmak ve kullanılacak bu faktörlerin hepsinin harita doğruluğuna her zaman iyi yönde katkı vereceğini düşünmek yanlıştır. Özellikle faktörler içerisindeki gürültü veya çalışma için istenilen bilgi dışındaki verileri içermesi doğruluk artırmak dışında düşmesine ya da etki etmemesine neden olabilmektedir. İlişkisiz veri kümesinin yaratacağı sonuç "boyutsallık çıkmazı (curse of dimensionality)" veya "Hughes fenomeni" etkisine yani parametre sayısının artmasına karşın doğruluğun düşmesine sebep olur (Hughes, 1968). Faktör sayısında indirgeme işlemi yapılarak çalışma için en uygun faktör kümesi elde edilirken gürültü içeren verilerden kaçınılmış olunur. Ayrıca büyük boyutlu çalışmalarda faktör indirgemesiyle işlem süreci, depolama alanı ve bilgisayar hafızasında tasarruf ile bilgisayarın üzerindeki işlem yükü de hafifletilmiş olacaktır.

Bu çalışmada Trabzon ili Maçka ilçesi için üretilecek heyelan duyarlılık haritasında litoloji, eğim, arazi kullanımı, bakı, yükseklik, topoğrafik ıslaklık indeksi, eğim uzunluğu, eğrilik profili ve çizgisellik gibi 9 farklı faktör katmanı dikkate alınmıştır. Söz konusu faktör kümesi içinde heyelan duyarlılık haritası üretimin kullanılmak üzere 9 faktör seti ki-kare testi kullanılarak optimum 3,4,5,6,7 ve 8 faktör içeren modeller oluşturulmuştur. Optimum faktör setlerinin elde edilmesinden sonra heyelan duyarlılık haritası üretiminde literatüre de yaygın bir kullanıma sahip olan lojistik regresyon metodundan faydalanılmıştır (Dai vd., 2002; Bai vd., 2011; Kavzoglu et al., 2014). Üretilen tematik haritalar kendi içerisinde önemlilik derecelerine göre sıralandırılarak çok yüksek, yüksek, orta, düşük ve çok düşük olmak üzere 5 farklı dereceye ayrılarak heyelan duyarlılık haritası oluşturulmuştur. Üretilen her bir duyarlılık haritalarının doğruluklarının tespiti için genel doğruluklarına bakılmıştır. Ayrıca çalışmada ki-kare özellik seçimi metoduyla belirlenen alt küme sonuçlarının birbirleriyle olan farklılıklar ve istatistiksel anlamlılıklarının irdelenmesi için McNemar's testi kullanılmıştır.

2. ÇALIŞMA ALANI VE VERİ

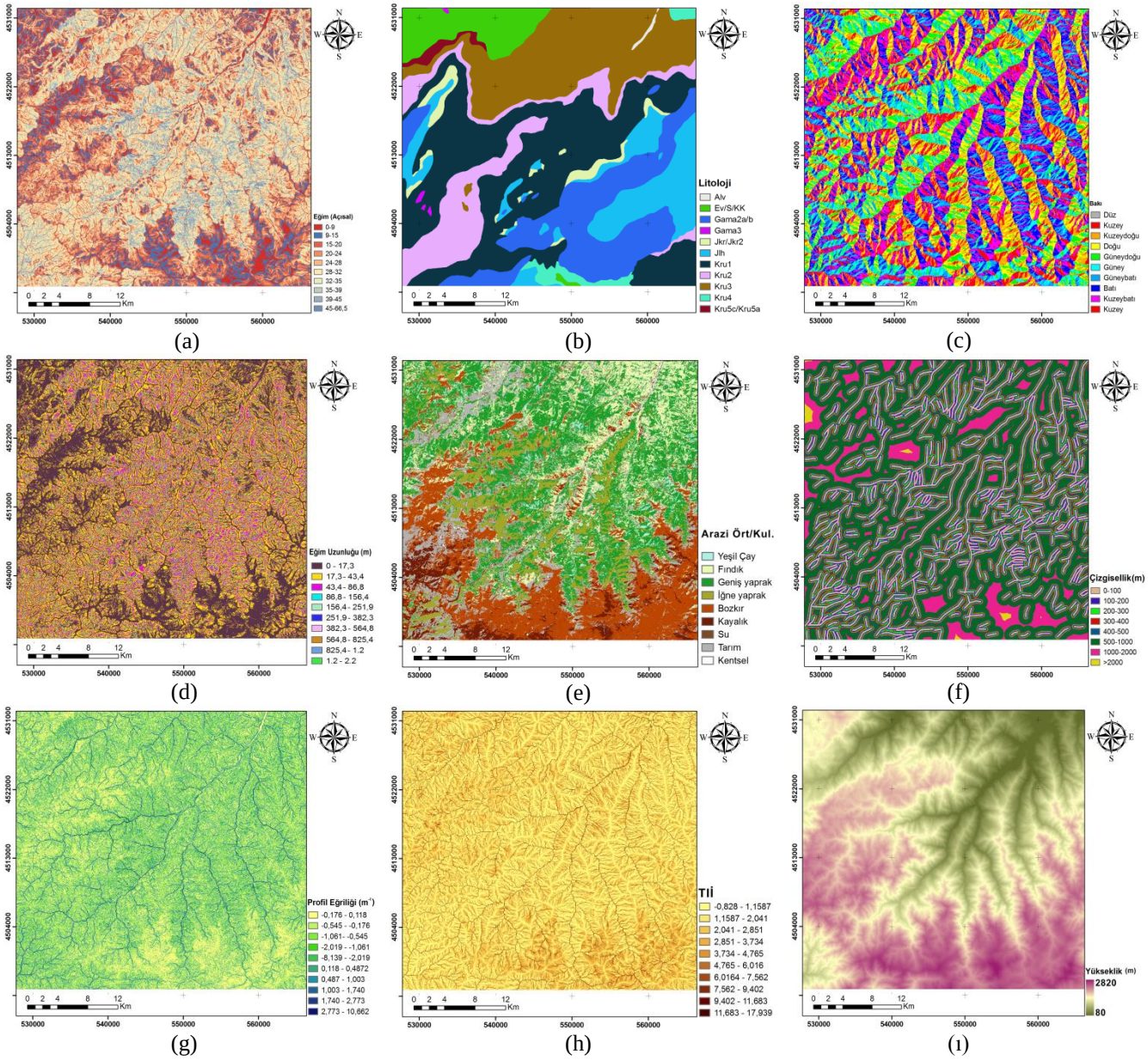
Çalışma alanı olarak seçilen Trabzon ili Maçka ilçesi yaklaşık 855 km^2 bir alanı kaplamaktadır. Doğu Karadeniz bölgesinin sahip olduğu karakteristik meteorolojik ve topoğrafik koşullar nedeniyle bölgede çok sayıda heyelan afeti gerçekleşmektedir. Yıllık ortalama 2.200 milimetrelik yağış alan bölgenin sahip olduğu yüksek eğimli yapısı heyelan riski için en önemli etki faktörüdür. Bölgede yer yer heyelan olayları görülse de bilinen en önemli heyelan kaynaklı afet 1998 yılında Çatak köyünde gerçekleşmiş ve 64 kişi hayatını kaybetmiştir.



Şekil 1. Çalışma alanı olarak seçilen Trabzon ili Maçka ilçesi.

Çalışma bölgesine ait Maden Teknik Arama (MTA) Enstitüsü tarafında üretilmiş heyelan envanter haritası mevcuttur. Söz konusu envanter haritası meydana gelmiş 51 poligon içermektedir. Poligonlar uygulama için 30 m lik raster formatta dönüştürülmüştür. Ayrıca yine uygulamada kullanılmak üzere 30 adet heyelan olmayan alan tespit edilmiştir. Heyelan

olmayan alanların tespitinde eğimin 5 dereceden az ve nehir, göl gibi düz ve topoğrafik açıdan heyelana elverişsiz alanlar dikkate alınmıştır (Gomez and Kavzoglu, 2005). Heyelan duyarlılık haritası üretiminde kullanılacak faktörler için sayısal yükseklik modeli (SYM), uydu görüntüsü ve jeolojik haritalardan faydalanılmıştır. Çalışma bölgesine ait SYM verisi Harita Genel Komutanlığına ait 1/25.000'lik eş yükseklik eğrilerinden üretilmiştir. 30 metre çözünürlüklü raster formatta dönüştürülen SYM'den eğim, bakı, yükseklik, topoğrafik ıslaklık indeksi, eğim uzunluğu, eğrilik profili ve çizgisellik haritaları elde edilmiştir. 2010 yılına ait Landsat TM uydu verisi kullanılarak arazi örtüsü haritası ve 1998 yılında MTA tarafından üretilen 1/100.000'lik litoloji verisi kullanılmıştır. Tüm faktör haritalar uygulama süresince analiz işlem kolaylığı ve veriler arası standartlaştırma için tekrar sınıflandırılarak alt sınıflara ayrılmıştır. Çalışmada kullanılan tüm faktörler ve alt sınıflara ait bilgiler aşağıda Şekil 2'de verilmiştir.



Şekil 2. Heyelan duyarlılık haritası üretiminde kullanılan faktörler (a) Eğim, (b) Litoloji, (c)Bakı, (d) Eğim uzunluğu, (e) Arazi örtüsü/kullanımı, (f) Çizgisellik, (g) Profil eğriliği, (h) Topogragik ıslaklık indeksi, (ı) Yükseklik

3. ÇALIŞMADA KULLANILAN YÖNTEMLER VE UYGULAMA

3.1 Özellik Seçimi

Özellik seçimi, sınıflandırma veya regresyon işlemleri için kullanılacak özelliklerin (örn. faktör veya parametre) tüm özellik kümesi içinden doğruluğu etkileyecek ilgisiz verilerinin elenerek kümeyi en iyi temsil eden alt kümenin belirlenmesi işlemidir (Webb and Copsey, 2011). Özellik seçimi genel olarak iki ana sebep olan doğruluk ve ölçeklenebilirlik amacıyla kullanılır. Sınıflandırma veya regresyon algoritmaların kullanılan veri setinin tamamının işleme sokulması elde edilecek işlem sonucunun da doğruluğunun yüksek olacağı varsayılır. Fakat veri seti içindeki bazı özellikler işlem performansını olumsuz etkileyecek gürültüye sahiptir ve bu nedenle bu özellikleri veri seti içinden

silinmesi işlem sonucunun doğruluğunun artmasında etkili olabilmektedir. Diğer taraftan algoritmalarda kullanılacak veri boyutunun azaltılması da işlem gücü, hafıza ihtiyacı, ağ bant genişliği ve depolama alanı gibi işlem süreci üzerinde etkili konularda zaman tasarrufu sağlar.

Özellik seçimi metotları sarmalama (wrapper) ve filtreleme olarak bilinen iki gruba ayrılır. Filtreleme metotları özelliklerin ilişkilerinin belirlenmesinde istatistiksel testlere başvurulur ve belirlenen bir eşik değer üzerinde kalan özellikler seçilir. Sarmalama metotlarında ise en iyi özellik seçiminde sınıflandırma metotları kullanılır ve sınıflandırma sonuçlarına göre en iyi özellik kümesi belirlenir. Gereksiz özelliklerin elenmesi ve daha iyi sonuçların elde edilmesi amacıyla, filtreleme metotlarında her özellik bağımsız olarak değerlendirilirken, sarmalama metotlarında tüm veri seti değerlendirilir.

Filtreleme tabanlı özellik seçimi yöntemleri, çeşitli değerlendirme ve dizme yöntemleriyle faktörlerin sıralayıp belirli bir eşik değerinin üzerinde değer alan özelliklerin seçilmesi esasında çalıştıkları için özellik sıralama yöntemleri olarak da bilinirler. Filtreleme metotları sarmalama metotlarla karşılaştırıldığında daha hızlı ve özelliklerin bağımsız olarak yorumlanması daha kolay olan metotlardır (Camps-Valls et al., 2011). Literatüre de en çok kullanılan filtreleme tabanlı öznitelik seçimi algoritmaları arasında ki-kare istatistiği, bilgi kazanımı, kazanım oranı, simetrik belirsizlik, Relief fonkisyonu ve One-R metodu yer alır (Novakovic et al., 2011). Heyelan duyarlılık haritası için kullanılan veri setinin boyut indirgemesi için test ve eğitim alanlarından faydalanan ki-kare öznitelik seçimi yöntemi kullanılmıştır.

3.2 Ki-kare Özellik Seçimi Yöntemi

Ki-kare testi (x^2) iki değişken arasındaki ilişkinin bağımlı veya bağımsız olduğunu belirlemeye yarayan ayrık veriler için kullanılan bir hipotez test yöntemidir. Ki-kare istatistiğine dayalı özellik seçimi metodu iki adımı içermektedir. Yöntemin ilk kısmında özelliklerin sınıflara göre ki- kare istatistikleri hesaplanır. İkinci kısımda serbestlik derecesi ve belirlenen önemlilik seviyesine göre ki-kaynaşımı (chi-merge) prensibi ile ki-kare değerlerine bakılarak veri seti içerisindeki tutarsız özelliklerin bulunana kadar art arda özelliklerin ayrıştırılmasıdır (Liu and Setiono, 1995). Ki-kaynaşımı algoritması 1992 yılında Randy Kerber tarafından yazılmış ve 1995 yılında Huan Liu ve Rudy Setiono tarafından ise yeniden düzenlemiştir. Veri kümesi içinde yer alan bir faktör için hesaplanan ki-kare değeri, o faktörün sınıf içerisindeki bağımlılığını ölçmektedir. Sıfır değerine sahip bir faktör o küme içinde bağımsız olduğunu gösterir. Yüksek bir ki-kare değerine sahip olan faktör, veri kümesi için daha tanımlayıcıdır. Ki-kare değerinin hesaplanmasında kullanılan genel eşitlik aşağıda verilmiştir.

$$x^2 = \sum_{i=1}^{2} \sum_{j=1}^{k} \frac{\left(A_{ij} - E_{ij}\right)^2}{E_{ij}} \tag{1}$$

Bu eşiklikte k sınıf sayısı, A_{ij} gözlenen frekans değeri (i satır, j sütün)ve E_{ij} ise beklenen (teorik) frekans değeridir.

3.3 Logistik Regrasyon

Lojistik regresyon (LR) birçok yer bilimi araştırmalarında olduğu gibi heyelan duyarlılık analizlerinde de sıklıkla kullanılan istatistiksel tabanlı bir modeldir (Yilmaz, 2010). Lojistik regresyon metodunun temel amacı, bağımlı bir değişken ile bağımsız birçok değişken arasında çok değişkenli bir regresyon ilişkisi kurmayı sağlamaktır (Lee, 2005). Bu regresyon modelinde bağımlı değişken heyelan olayının olması ya da olmaması durumunu gösterirken, bağımsız değişler ise modelde kullanılan ve heyelan olayına etki eden faktörlerdir. Lojistik regresyon modelinin dayandığı temel lineer eşitlik aşağıdaki gibi gösterilebilir,

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n \tag{2}$$

Bu eşitlikteki Y, 0 ve 1 aralığındaki bağımlı değişkeni, b_0 bağımsız değişkenlerin sıfır değerini aldıklarındaki bağımlı değişken değeri ya da sabiti, $b_1,\ldots,b_n$ bağımsız değişkenlerin regresyon katsayılarını ve $x_1,\ldots,x_n$ ise bağımsız değişkenleri göstermektedir. LR modelindeki bağımlı değişkenler aşağıdaki eşitlikle tanımlanır.

$$Logit(\mathrm{p}) = \ln\left(p/(1-p)\right) = 1/1 + e^{-b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n} \tag{3}$$

Eşitlikteki p bağımlı değişkenin 0 ile 1 arasında değer alabilen olasılık değeri ve $\left(p/(1-p)\right)$ ise tanımlanan olabilirlik oranıdır.

3.4 McNemar's Test

Farklı sayıda faktör sayısını içeren veri setlerinin lojistik regresyon analizi sonucunda elde edilen doğruluklarının karşılaştırılmasında öncelikli olarak genel sınıflandırma doğrulukları göz önüne alınmıştır. Bununla birlikte faktör

sayısındaki artışla birlikte hesaplanan doğruluklar arasındaki farkların anlamlılığı McNemar testi kullanılarak istatistiksel olarak analiz edilmiştir. Ki-kare dağılımını esas alan McNemar testi hesaplamalarda 2x2 boyutlu bir hata matrisi kullanmaktadır (Foody, 2004). Eşitlik 4 yardımıyla hesaplanan istatistik değer ki-kare tablo değerinden (%95 güven aralığında) büyük olduğunda iki doğruluk arasındaki farkın istatistiksel olarak anlamlı olduğu söylenebilir. Diğer bir ifadeyle doğruluklar arasındaki bu anlamlı farklılık, iki regresyon sonucunun birbirinden istatistiksel olarak farklı olduğunun bir göstergesidir.

$$\chi^2 = \frac{\left(n_{ij} - n_{ji}\right)^2}{n_{ij} + n_{ji}} \tag{4}$$

Eşitlikte n_{ij}, *i*. regresyon modeli tarafından hatalı *j*. regresyon modeli tarafından doğru sınıflandırılan piksel sayısını gösterirken, n_{ji}, *j*. regresyon modeli tarafından hatalı *i*. regresyon modeli tarafından doğru sınıflandırılan piksel sayısını ifade etmektedir.

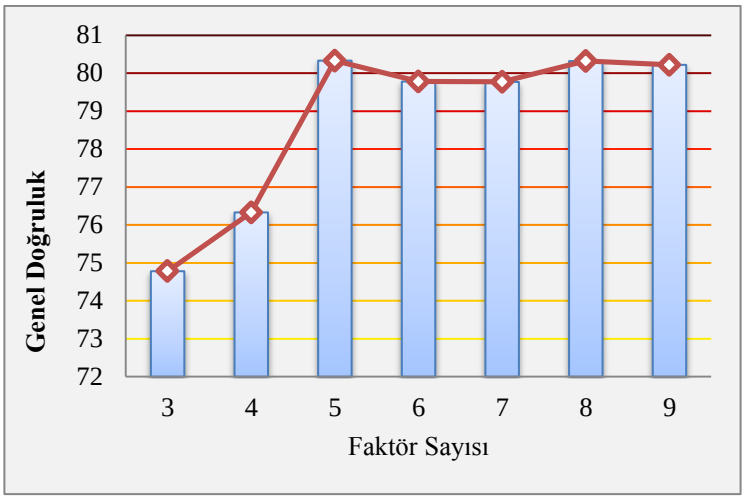
4. UYGULAMA

Heyelan duyarlılık haritası üretiminde kullanılacak 9 adet faktör analizler öncesi yeniden sınıflandırılarak alt sınıflarına ayrılmıştır. Bu işlem için Coğrafi Bilgi Sistemleri platformu kullanılmış, sürekli veri türündeki (örn. Eğim, yükseklik, eğim uzunluğu gibi) faktörlerin aralıklarının tespitinde ArcGIS yazılımı içerisinde bulanan Jenks (Doğal aralıklandırma) sınıflandırıcısından faydalanılmıştır. Ana faktör kümesi içerisindeki en iyi alt kümenin belirlenmesinde Matlab ortamında yazılmış ki-kare özellik seçimi metodu uygulanmıştır. Özellik seçimi algoritmasında optimum parametre setinin belirlenmesinde gerekli olan eğitim ve veri verisi, heyelan envanter haritası içerisinde bulunan 51 olan ve 30 olmayan heyelan alanlarından rastgele 450 eğitim ve 350 test alanından oluşan piksel verisi kullanılmıştır. Ki-kare testi sonrası elde edilen en iyi faktör sıralaması ve ağırlıkları aşağıda Tablo 1'de verilmiştir. Tabloda yüksek ağırlık değerleri alan faktörlerden eğim, yükseklik, arazi örtüsü/kullanımı ve litoloji literatürde yapılan çalışmalarda da en sık kullanılan faktörler arasındadır (Hasekiogullari and Ercanoglu, 2012). Elde edilen bu bulgu heyelan duyarlılık haritası üretimi için önerilen yaklaşımın başarısını destekler niteliktedir. Bunun dışında ki-kare özellik seçimi sonucunda yüksek ağırlık değeri alan eğim uzunluğu diğer faktörlere göre daha az sıklıkta kullanılmasına rağmen bu çalışma alanı için söz konusu faktörün önemli bir etken olduğu görülmüştür.

Tablo 1. Ki-kare testi sonucu elde edilen optimum faktör sıralaması ve ağırlıkları

Faktör	Ağırlıkları
Eğim	0,3263
Yükseklik	0,1822
Eğim uzunluğu	0,1528
Arazi örtüsü/kullanımı	0,1463
Litoloji	0,0753
Çizgisellik	0,0637
Profil Eğriliği	0,0317
Tİİ	0,0141
Bakı	0,0074

Duyarlılık haritasının üretilmesi işlemleri MATLAB ortamında yazılmış lojistik regresyon algoritması kullanılarak üretilmiştir. Lojistik regresyon yöntemi ile modellerin elde edilmesinde bağımsız değişkenler 3,4,5,6,7 ve 8 risk faktörünün ki-kare testi sonrası sıralaması dikkate alınmıştır. Çalışma bölgesi için üretilen tematik haritalar CBS platformu üzerinde kendi içerisinde önemlilik derecelerine göre sıralandırılarak çok yüksek, yüksek, orta, düşük ve çok düşük olmak üzere 5 farklı dereceye ayrılarak heyelan duyarlılık haritası oluşturulmuştur. Yapılan çalışmada genel doğruluğun değerlendirilmesinde literatürde de olduğu gibi yüksek ve çok yüksek riskli alanlar dikkate alınmıştır. Çalışma bölgesi için farklı veri kümesi ile oluşturulmuş heyelan duyarlılık haritalarının genel doğrulukları aşağıda Şekil 3'de verilmiştir. Her bir duyarlılık haritasına ait genel doğruluklar elde edildiğinde en yüksek doğruluğu %80,33 ile 5 veri setinden oluşan (eğim, yükseklik, eğim uzunluğu, arazi örtüsü/kullanımı ve litoloji) modellin verdiği görülmüştür. Diğer bir taraftan ise 3 veri seti ile oluşturulan modelin %74,78 ile en düşük genel doğruluğa sahip olduğu tespit edilmiştir. Ki-kare özellik seçimi sonrası elde edilen modeller sonrası üretilmiş heyelan duyarlılık haritalarının doğruluklarının kullanılan faktör sayısının artmasına karşılık belirli bir seviyeye kadar artış gösterdiğini fakat bu noktadan sonra faktör sayısındaki artışın doğruluk üzerinde herhangi bir etkiye sahip olmadığı ve az bir oranda olsa doğrulukta düşüş olduğu görülmüştür. Bu durum literatürde boyutsallık problemi ya da Hughes Fenomeni ile bağdaşmaktadır.



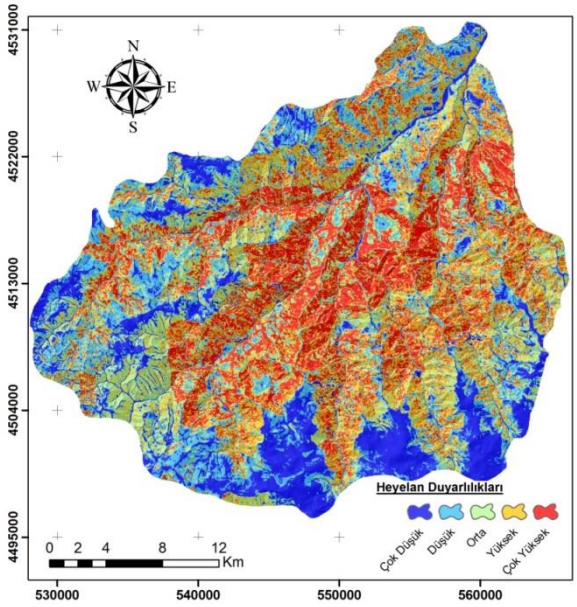
Şekil 3. Optimum veri setleri sonrası üretilen heyelan duyarlılık haritalarının genel doğrulukları.

Modellerin performanslarının karşılaştırılmasına yönelik hesaplanan McNemar's test sonuçları Tablo 2'de verilmiştir. Tabloda %95 güven aralığındaki kritik değerin (3,84) altında kalan istatistik değerler koyu olarak gösterilmiştir. Tablo incelendiğinde ki-kare testi ile seçilen 5 faktörden sonra veri setine eklenecek yeni faktörler harita doğruluğunda anlamlı değişimler ortaya çıkarmamaktadır. Diğer bir ifadeyle 5 faktör kullanılarak üretilen duyarlılık haritasının doğruluğu ile daha fazla sayıda (6,7,8 ve 9) faktörün dikkate alınmasıyla üretilen harita doğrulukları istatistiksel olarak aynıdır. İstatistiksel sonuçlar lojistik regrasyon yönteminin performansının faktör sayısının artmasıyla beraber belirli bir seviyeye kadar yükseldiğini, bu kritik noktadan sonra yöntemin performansının değişmediğini destekler niteliktedir. Diğer bir taraftan 3 ve 4 faktör setleri ile üretilen haritalarını hem kendi içinde hem de diğer haritalardan istatiksel olarak farklı performanslar sergilediği açıkça görülmektedir. Sonuç olarak çalışma bölgesi ve kullanılan veri setleri için seçilen 5 faktörün (eğim, yükseklik, eğim uzunluğu, arazi örtüsü/kullanımı ve litoloji) kullanımı ile üretilen duyarlılık haritasının istatistiksel olarak istenilen doğrulukta ve anlamlılıkta olduğunu göstermektedir. McNemar's test sonuçlarından elde edilen diğer bir bulgu yüksek korelasyona ve bozucu etkiye sahip olduğu düşünülen çizgisellik, profil eğriliği, Tİİ ve bakı faktörlerinin bu çalışma alanı için doğruluk üzerinde de etkiye sahip olmadığıdır. Çalışma sonuçları literatürde ki çalışmalar ile karşılaştırıldığında bakı faktörü hariç diğer faktörlerin duyarlılık haritaları üretiminde de en az kullanılan faktörlerden olduğu görülmüştür (Hasekiogullari and Ercanoglu, 2012). Bakı faktörünün bu çalışma alanı için en az ağırlığa sahip olması ve doğruluk üzerinde etkisi olmayışı diğer literatür çalışmaları ile örtüşmeyen bir sonuçtur. Elde edilen bu sonuç bakı faktörünün diğer faktörler ile karşılaştırıldığında en yüksek korelasyona yani benzer özelliklere sahip faktör olduğu ve bu nedenle en az etkiye sahip olan faktör olmasıyla açıklanabilir.

Tablo 2. McNemar's Test Sonuçları (3,84)

Faktör Seti	3	4	5	6	7	8	9
3	-	10,51	39,85	30,90	28,36	32,70	31,06
4		-	24,67	18,04	16,27	20,81	19,38
5			-	**2,33**	**1,59**	**0,01**	**0,11**
6				-	**0,01**	**1,44**	**1,01**
7					-	**3,36**	**2,03**
8						-	**2,25**
9							-

Ki-kare özellik seçim ile elde edilen sıralama sonrası en iyi doğruluğun elde edildiği 5 faktör kullanılarak üretilen duyarlılık haritası Şekil 4'de gösterilmiştir. Heyelan duyarlılık haritası sonrası olası heyelan alanlarının risk dağılımı incelendiğinde çok yüksek ve yüksek risk değerlerinin Maçka ilçesinin orta ve kuzey doğu kısımlarında baskın olduğu görülmektedir. Duyarlılık haritasının güney kısımlarının büyük bir bölümünde heyelan risk derecelendirmesinin çok düşük ve düşük olduğu görülmüştür. Yapılan analizler sonrası Maçka ilçesinin yaklaşık %44'lük (yaklaşık 377 km^2) gibi büyük bir alanının çok yüksek ve yüksek risk sahasının içerisinde olduğu tespit edilmiştir. Ayrıca diğer bir analiz sonucu ise, tüm ilçenin sadece %34'lük bir alanın (yaklaşık 295 km^2) olası çok düşük ve düşük riskte heyelan bölgesi olduğu görülmektedir.



Şekil 3. Optimum 5 faktör setine dayalı heyelan duyarlılık haritası

5. SONUÇLAR

Ülkemiz için heyelan afeti depremden sonra gelen ve büyük ölçüde can ve mal kayıplarına neden olan doğal bir afettir. Heyelan riskinin önceden tespiti ve olası etkilerinin en aza indirgenmesi amacıyla üretilecek heyelan duyarlılık haritaları karar vericiler için en önemli veri altlığıdır. Bu yüzden üretilecek duyarlılık haritalarının mümkün oldukça doğru ve güvenilir olması gerekmektedir. Günümüze kadar heyelan duyarlılık haritalarının üretimi süreci ve kullanılan katmalar literatürde birçok araştırmacı tarafından çalışma konusu olmuştur. Fakat duyarlılık incelemelerinde kullanılan parametrelerin duyarlılık haritası üzerindeki etkileri ve en optimum parametre tespiti yapılarak harita üretimi genellikle araştırmacılar tarafından üzerinde durulmayan bir konu başlığıdır. Gelişen teknolojik imkanlar ve algoritmik çözümler heyelan duyarlılık haritası üretimi gibi çalışmalarda araştırmacılara onlarca faktör kullanımı imkanı sağlamaktadır. Özellikle SYM ve uydu görüntülerinden türetilen verilerden duyarlılık analizlerinde kullanılmak üzere çok sayıda veri elde edilebilmektedir. Söz konusu durumda yüksek boyutlu veri kümelerinin hepsini herhangi bir indirgeme analiziyle incelenmeden tamamının kullanılması elde edilecek doğruluklar üzerinde olumsuz etkilere neden olduğu görülmektedir. Bu nedenle, çok katmanlı veri kümesi içerisindeki birçok özellik veya faktörden, veriyi daha iyi ifade edebilecek şekilde bir alt küme bulmayı amaçlayan yaklaşımla yani özellik seçimi ile veri indirgemesi yapılmalıdır. Özellik seçimi algoritması kullanılarak çok boyutlu veri kümesinin sahip olduğu bilgi miktarlarına ve ayırma derecelerine bakılarak en yüksek bilgiyi ve en iyi ayırıcılığa sahip faktörlerin seçilmesi ve kalanların da elenmesi mümkün olmaktadır.

Bu çalışmada Trabzon ili Maçka ilçesi için üretilecek heyelan duyarlılık haritasında 9 faktörden oluşan veri kümesi dikkate alınmıştır. Mevcut veri kümesi ki-kare özellik seçimi algoritması kullanılarak duyarlılık haritası analizi için en uygun faktörden en azına kadar sıralandırılmıştır. Sıralama sonrası 3,4,5,6,7 ve 8'li veri gruplarından oluşan modeller kurulup lojistik regresyon analizine sokularak heyelan duyarlılık haritaları üretilmiştir. Üretilen haritalar içerisinde en iyi genel doğruluğu %80,33 ile 5 adet faktörden (eğim, yükseklik, eğim uzunluğu, arazi örtüsü/kullanımı ve litoloji) oluşan model vermiştir. Geri kalan 4 faktörün (bakı, topoğrafik ıslaklık indeksi, eğrilik profili ve çizgisellik) regresyon analizinde kullanılmasının McNemar's test sonuçlarına göre harita doğruluğunda anlamlı değişimler ortaya çıkarmadığı, diğer bir ifadeyle elde edilen sonuca istatistiksel olarak etki etmediği görülmüştür. Sonuç olarak, tüm veri setleri incelendiğinde genel doğruluğun 3 faktör kullanılarak üretilen modelden 5 faktör ile üretilen modele kadar artış sergilediği, sonrasında ise az miktarda düşüş sergilese de genel olarak yakın doğruluklar verdiği görülmüştür. Elde edilen bu sonuçlar faktör sayılarının artışın bir noktadan sonra heyelan duyarlılık haritasının doğruluğu üzerinde bir katkı sağlamadığıdır. Bu durumda belli bir boyut artışından sonra diğer faktörlerin problemin çözümünde farklı ve ayırt edici özellikler sunmaması bu nedeniyle boyutsallık probleminin oluştuğu söylenebilir.

KAYNAKLAR

Bai, S.B., Lu, G.N., Wang, J.A., Zhou, P.G., Ding, L.A., 2011, GIS-based Rare Events Logistic Regression for Landslide-Susceptibility Mapping of Lianyungang, China, *Environmental Earth Sciences*, 62, 139-149.

Camps-Valls, G., Tuia, D., Gomez-Chova, L., Jimenez, S., Malo, **J.,** 2011, *Remote Sensing Image Processing, Synthesis Lectures on Image,Video, and Multimedia Processing*, (Morgan and Claypool).

Dai, F.C., Lee, C.F., 2002, Landslide Characteristics and Slope Instability Modeling Using GIS, Lantau Island, Hong Kong, *Geomorphology*,42, 213-228.

Foody, G.M., 2004, Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy, *Photogrammetric Engineering and Remote Sensing*, 70(5), 627-633.

Gomez, H., Kavzoglu, T., 2005, Assessment of Shallow Landslide Susceptibility Using Artificial Neural Networks in Jabonosa River Basin, Venezuela, *Engineering Geology*, 78, 11–27.

Hasekiogullari, G.D. & Ercanoglu, M. 2012. A New Approach to Use AHP in Landslide Susceptibility Mapping: A Case Study at Yenice (Karabuk, NW Turkey). *Natural Hazards*, 63, 1157-1179.

Hughes, G.F., 1968, On The Mean Accuracy of Statistical Pattern Recognizers, *IEEE Transactions on Information Theory*, 14, 55-63.

Kavzoglu T., Kutlug Sahin E., Colkesen I., 2014, Landslide Susceptibility Mapping Using GIS-Based Multi-Criteria Decision Analysis, Support Vector Machines, And Logistic Regression, *Landslides*, 11 (3):425-439.

Kavzoglu,T., Kutluğ Şahin, E. ve Çölkesen, İ., 2012, Heyelan Duyarlılığının İncelenmesinde Regresyon Ağaçlarının Kullanımı: Trabzon Örneği, *Harita Dergisi*, 147(3), 21-33.

Kerber, R., 1992. ChiMerge: Discretization of Numeric Attributes, *Proceedings, Ninth National Conference on Artificial Intelligence, pages 123-128.*

Lee, S., 2005, Application of Logistic Regression Model and Its Validation for Landslide Susceptibility Mapping Using GIS and Remote Sensing Data, *International Journal of RemoteSensing*, 26, 1477-1491.

Liu, H., Setiono, R., 1995, Chi2: Feature Selection And Discretization of Numeric Attributes, *Seventh International Conference on Tools with Artificial Intelligence, Proceedings*:388-391

Liu, H., Setiono, R.,1995. Chi2: Feature Selection and Discretization Of Numeric Attributes. *Proceedings of the IEEE 7th International Conference on Tools with Artificial Intelligence (TAI'95)*, 388-391, IEEE Press, Washington, DC.

Novakovic, J., Strbac, P., Bulatovic, D., 2011, Toward Optimal Feature Selection Using Ranking Methods and Classification Algorithms", *Yugoslav Journal of Operations Research*, vol. 21, pp. 119-135.

Pourghasemi, H. R., A. G., 2013, Jirandeh, B. Pradhan, C. Xu, and C. Gokceoglu, Landslide Susceptibility Mapping Using Support Vector Machine and GIS at The Golestan Province, Iran, *Journal of Earth System Science*, 122, no. 2: 349-69.

Regmi, A. D., Devkota, K. C., Yoshida, K., Pradhan, B., Pourghasemi, H. R., Kumamoto, T., Akgun, A.,, 2014, Application Of Frequency Ratio, Statistical Index, and Weights-Of-Evidence Models and Their Comparison in Landslide Susceptibility Mapping in Central Nepal Himalaya, *Arabian Journal of Geosciences*, 7, no. 2: 725-42.

Kutluğ Şahin, E., 2012, *CBS Tabanlı Çok Kriterli Karar Analizi Yöntemi Kullanılarak Heyelan Duyarlılık Haritasının Üretilmesi: Trabzon İli Örneği, Yüksek Lisans Tezi*, Gebze Yüksek Teknoloji Enstitüsü, Kocaeli.

Theory 14, 55–63.Shahabi, H., Khezri, S., Bin Ahmad, B., Hashim, M., 2014, Landslide Susceptibility Mapping at Central Zab Basin, Iran: A Comparison Between Analytical Hierarchy Process, Frequency Ratio and Logistic Regression Models, *Catena* 115: 55-70.

Thiery, Y., Maquaire, O., Fressard, M., 2014, Application of Expert Rules in Indirect Approaches for Landslide Susceptibility Assessment, *Landslides* 11, no. 3: 411-24.

Webb, A.R., Copsey, K.D., 2011, *Statistical pattern recognition, Third Edition*, (Chichester: John Wiley and Sons).

Yalcin, A., 2008, Gis-Based Landslide Susceptibility Mapping Using Analytical Hierarchy Process and Bivariate Statistics in Ardesen (Turkey): Comparisons Of Results And Confirmations, *Catena*, 72, no. 1: 1-12.

Yilmaz, I., 2010. Comparison Of Landslide Susceptibility Mapping Methodologies for Koyulhisar, Turkey: Conditional Probability, Logistic Regression, Artificial Neural Networks, And Support Vector Machine, *Environmental Earth Sciences*, 61, 821–836.