

Heyelan Duyarlılık Haritası İçin Adımsal Regresyona Dayalı Faktör Seçme Yönteminin Etkinliğinin Araştırılması

(Assessment of Factor Selection Based on Stepwise Regression for Landslide Susceptibility Mapping)

Emrehan Kutluğ ŞAHİN

Gebze Teknik Üniversitesi, Harita Mühendisliği Bölümü, Gebze, Kocaeli
e.sahin@gtu.edu.tr

ÖZ

Optimum faktör seçimi ve harita üretim metodunun belirlenmesi heyelan duyarlılık haritası üretim sürecinin en önemli bileşenleri arasında yer almaktadır. Fakat, heyelan duyarlılık haritalarında kullanılan faktörler ile yöntemlerin tespitinde henüz evrensel kuralların oluşturulamadığı söylenebilir. Bu çalışmanın temel amacı, çok sayıda heyelanda etkili faktörler içerisinde optimum faktör modelinin tespiti ve farklı yöntemler kullanılarak heyelan duyarlılık haritası üretimindeki etkinliğinin araştırılmasıdır. Bu amaç doğrultusunda, adımsal regresyon (AR) metodu optimum faktör modeli tespitinde kullanılmıştır. Ayrıca, lojistik regresyon (LR) ve rastgele orman (RO) gibi harita tahmin yöntemlerinin heyelan duyarlılık haritası üretimindeki etkinliği araştırılmıştır. Üretilen duyarlılık haritalarının performanslarının değerlendirilmesi amacıyla genel doğruluk, kappa değerleri, ROC (Alıcı İşlem Karakteristikleri) eğrisi ile AUC (Eğri Altında Kalan) değerleri incelenmiştir. Ayrıca üretilen modeller arasındaki farklılıklar ve istatistiksel anlamlılıkları McNemar's testi ile analiz edilmiştir. Çalışma sonucunda, AR yöntemiyle belirlenen 6 faktörden oluşan Model-6 (NDVI (Normalized Difference Vegetation Index), Yükseklik, Bakı, Litoloji, Topoğrafik Islaklık İndeksi(Tİİ), Eğim) ile optimum faktör model kurulmuştur. En yüksek genel doğruluk (%88,27) AR algoritması tarafından belirlenen 6 faktörü içeren veri setinin (Model-6) RO algoritması yardımıyla tahmin edilmiştir. Elde edilen sonuçlar, AR ve RO yöntemi ile belirlenen optimum faktör modelinin geleneksel LR yönteminden yaklaşık %8 oranında daha üstün olduğunu ve bu farkların istatistiksel olarak anlamlı olduğunu ortaya koymuştur.

Anahtar Kelimeler: Faktör seçimi, rastgele orman, duyarlılık haritası, lojistik regresyon, adımsal regresyon

ABSTRACT

Selection of optimum factors and map production methods are main crucial components of landslide susceptibility mapping process. Unfortunately, there are no universal guidelines regarding the selection of factors in landslide susceptibility mapping. The main purpose of this study was to determine an optimal factor model within the available landslide conditioning factors and investigation of suitable prediction method for landslide susceptibility mapping. For this purpose, stepwise regression method was used to determine the optimal factor model. In addition, this study investigates the map prediction method of logistic regression and random forest for producing landslide

susceptibility map. All optimum factor models were tested using overall accuracy, kappa value, area under the ROC curve and AUC value. Also, McNemar's test statistic was also applied to test the statistical significance of performance differences for factor models. As a result of the study, the optimum factor model was established as Model-6 consisting of 6 landslide causative factors (NDVI, elevation, aspect, lithology, topographic wetness index and slope) by applying stepwise regression method. The highest overall accuracy (88.27%) was estimated by RF algorithm for the dataset including 6 factors selected with Stepwise algorithm (Model-6). Results showed that the optimum factor model determined by stepwise regression and random forest method outperformed the conventional LR method by about 8% and this level of difference was found to be statistically significant.

Keywords: Factor selection, random forest, susceptibility map, logistic regression, stepwise regression

1. GİRİŞ

Doğal afet, büyük oranda veya tamamen insanların kontrolü dışında gerçekleşen, can ve mal kaybına neden olabilen, yerel ve küresel anlamda büyük zararlar doğuran olay ya da olaylar bütünüdür. Dünyada doğal afetler, nüfus artışı ve yetersiz planlama nedeniyle gelişmekte olan ülkelerin kırsal ve kentsel alanlarında büyüyerek artan bir zafiyettir. Doğal afetler ortaya çıkış nedenlerine göre incelendiğinde deprem, heyelan, yanardağ patlamaları ve tsunami gibi jeolojik kökenli olaylar olabileceği gibi su taşkını, fırtına, kuraklık, sel, hortum ve orman yangınları gibi meteorolojik etmenler olabilmektedir. Türkiye de yaşanan doğal afetler incelendiğinde, özellikle deprem, heyelan, erozyon, sel, su taşkını, yangın ve çığ en sık görülen doğal afet türleri olduğu söylenebilir (Ergünay, 2007). Ülkemizde can ve mal kayıpları açısından en yıkıcı doğal afetlerin başında deprem ve heyelan gelmektedir.

Heyelana eğilimli alanların belirlenmesi ve mekânsal tahmini, yerel yönetimlerin afet azaltma çalışmalarında ve yatırımlarında önemli bir rol oynamaktadır. Heyelanların karakteristik özellikleri ve çevresel etkileri tematik haritalar ile ortaya konulabilmektedir. Tematik haritalar, heyelana eğilimli alanlardaki duyarlılık

seviyelerinin ölçülmesi ve mekânsal dağılımları hakkında bilgi sağlayan ve birçok risk yönetimi çalışmalarında birincil girdi olarak yaygın bir şekilde kullanılan haritalardır. Bir tematik harita türü olan heyelan duyarlılık haritaları gelecekte oluşabilecek potansiyel heyelan alanlarının tespitinde kullanılır. Bu nedenle risk yönetimi için büyük önem arz eden heyelan duyarlılık haritalarının doğru ve güncel üretilmesi yerel ölçekten küresel ölçeğe kadar birçok çalışma için en önemli esaslardan biridir.

Heyelan duyarlılık haritalarının üretilmesindeki en önemli işlem adımı, heyelana etki edebilecek faktörleri belirlenmesi ve kullanılan faktörlerin bir arada değerlendirilmesi sürecidir. Bu kapsamda topoğrafik, jeolojik ve meteorolojik veri türlerine bağlı doğal kaynaklar ile birlikte insan kaynaklı birçok faktörün bir arada değerlendirilmesini gerektirmektedir. Bu nedenle duyarlılık haritası üretiminde çok sayıda faktörün birlikte değerlendirilmesini gerektiren karmaşık durumlar söz konusu olduğundan, karar vericilerin sezgisel, deneme-yanılma veya benzer çalışmalarda neler tercih ediliyor gibi yaklaşımlara başvurmaları yanlış kararların verilmesi riskini taşımaktadır. Ayrıca, heyelan duyarlılık haritalarında kullanılan faktörlerin seçimi konusunda henüz evrensel bir uzlaşma bulunmadığı söylenebilir (Gokceoglu ve Ercanoglu, 2001). Heyelan duyarlılık haritası üretimi değerlendirmelerinde önemli bir nokta ise her bölgenin farklı yapısal özellikler göstermesidir. Bu nedenle, duyarlılık değerlendirilmelerinde kimi bölgeler için topoğrafik parametreler anlamlı bulunup değerlendirilirken kimi bölgelerde jeolojik, çevresel veya hidrolojik niteliklere sahip faktörler tercih edilmektedir. (Nefeslioglu vd., 2008). Söz konusu nedenlerden dolayı karar vericilerin heyelan duyarlılık haritası üretiminde kullanılabilecek anlamlı faktörleri tespit etme işlemlerinde kullanıcı tabanlı eğilimler dışında modern karar verme tekniği olan özellik seçimi tekniklerine başvurmaları doğru ve hızlı sonuçlara ulaşmalarında faydalı olacaktır.

Heyelan duyarlılık haritası üretiminde önemli işlem adımlarından biri, sonuca etki edebilecek faktörlerin belirlenerek en uygun faktör kümesinin oluşturulması, bir diğeri ise en uygun faktör setinin kullanılacağı harita üretimi metodunun tespitidir. Bu nedenle araştırmacılar heyelan mekanizması üzerinde çalışmalarını günümüzde bile sürdürmekte (Sakellariou ve Ferentinou, 2001) ve heyelan duyarlılık haritalarının üretiminde makine öğrenme (Pham vd., 2017; Colkesen vd., 2016; Pourghasemi ve Kerle,

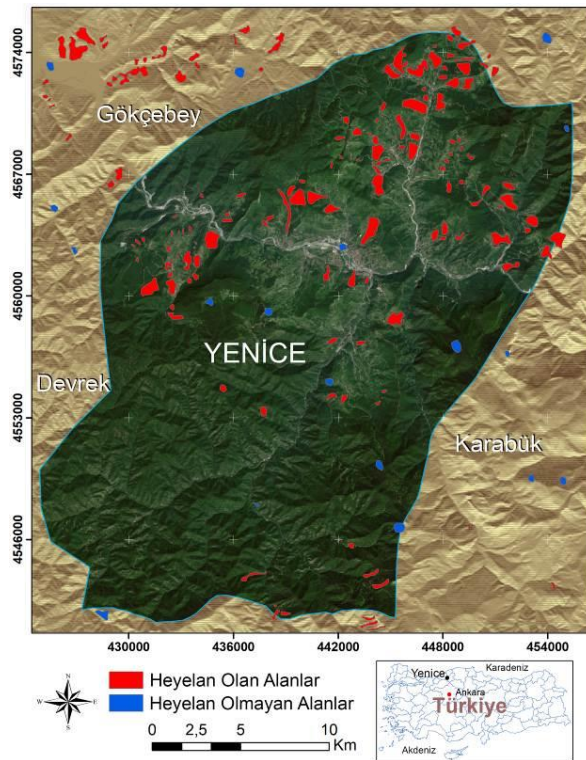
2016; Catani vd., 2013; Yılmaz, 2009), istatistiksel tabanlı metotlar (Zhang vd., 2016; Kavzoglu vd., 2015; Hong vd., 2015; Yalcin, 2008) veya kullanıcı tabanlı metotlar (Kutlug Sahin vd., 2016; Ahmed, 2015) gibi birçok farklı algoritma kullanılmaktadır. Heyelan duyarlılık haritalarının üretiminde literatürde yaygın olarak kullanılan diğeri bir yöntem ise istatistik tabanlı lojistik regresyon metodudur. (Ayalew ve Yamağishi, 2005; Akgun ve Bulut, 2007; Pradhan, 2010; Ercanoglu ve Temiz, 2011; Romer ve Ferentinou, 2016). Ayrıca son yıllarda ağaç tabanlı rastgele orman (RO) algoritması ile duyarlılık haritalarının üretilmesi konusunda araştırmalar bulunmaktadır (Catani vd., 2013; Goetz vd., 2015; Trigila vd., 2015; Youssef vd., 2016; Tsangaratos vd., 2017).

Heyelan duyarlılık haritaları üretiminde kullanılan parametrelerin birbirleri ile olan korelasyonlarının incelenmesi, gürlütlü ve/veya ilgisiz bilgiler içeren parametrelerin elenerek optimum parametre setinin belirlenmesi literatürde yeni bir yaklaşımdır (Kavzoglu vd., 2015; Bui vd., 2016; Vasu ve Lee, 2016; Liv vd., 2017). İstatistiksel faktör seçimi yöntemlerinden yaygın olarak bilinenlerinden biri olan adımsal regresyon (AR) yaklaşımı optimum faktör modelinin tespitinde kullanılan yöntemlerden biridir (Kano ve Harada, 2000; Wagner ve Shimshak, 2007; Zhang, 2016).

Bu çalışmada Karabük ili, Yenice ilçesi için üretilen heyelan duyarlılık haritasında bakı, drenaj yoğunluğu, litoloji, eğim, arazi kullanımı/örtüsü, yükseklik, NDVI, eğim uzunluğu, TII, topoğrafik pürüzlülük indeksi (TPI), sediment taşıma indeksi ve akarsu güç indeksi olmak üzere 12 farklı faktör dikkate alınmıştır. Söz konusu faktör kümesi içinde heyelan duyarlılık haritası üretiminde kullanılmak üzere optimum model tespiti için AR metodu kullanılmıştır. Optimum faktör modelinin elde edilmesinden sonra heyelan duyarlılık haritası üretiminde LR ve RO metodundan faydalanılmıştır. Her iki yöntemde de optimum model ve tüm veri setini içeren faktör grubu ile indeks haritaları üretilerek önemlilik derecelerine göre (çok yüksek, yüksek, orta, düşük ve çok düşük) beş sınıflı tematik heyelan duyarlılık haritaları oluşturulmuştur. Üretilen heyelan duyarlılık haritalarının performanslarının değerlendirilmesi amacıyla genel doğruluk, AUC ve kappa değerleri karşılaştırılmıştır. Ayrıca çalışmada optimum model ve diğeri veri seti ile üretilmiş haritaların birbirleriyle olan farklılıkları ve istatistiksel anlamlılıklarının irdelenmesi için ROC eğrisi karşılaştırması ve McNemar's testi kullanılmıştır.

2. ÇALIŞMA ALANI

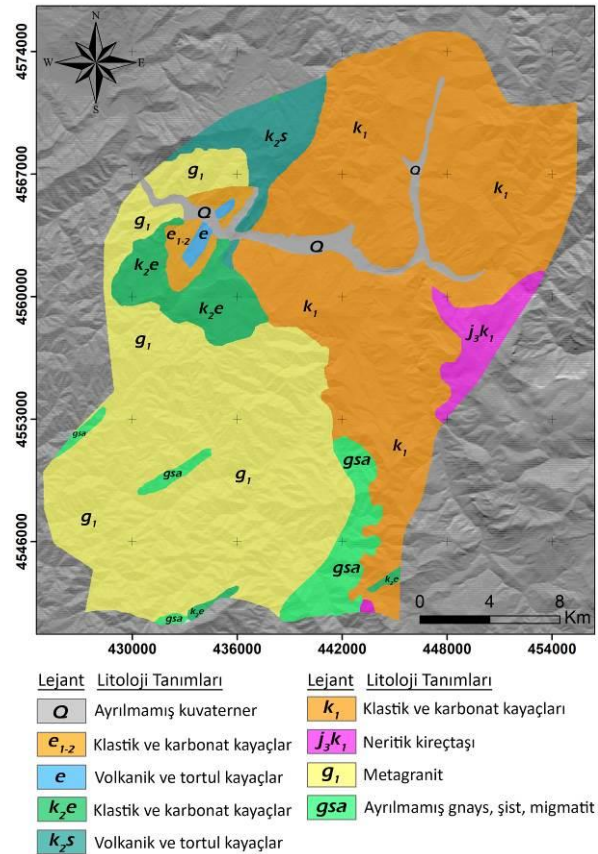
Batı Karadeniz bölgesinde yer alan Karabük ili sınırları içerisinde yer alan Yenice ilçesi çalışma bölgesi olarak seçilmiştir (Şekil 1). Yenice ilçesinin kuzeyinde Karadeniz ve Bartın ili, batısında Zonguldak ili ve doğusunda Karabük Merkez ilçesi bulunmaktadır. Çalışma alanı yaklaşık 678 km² bir alanı kapsamakta ve 31° 58' ve 32° 31' doğu-batı meridyenleri ile 41° 19' ve 41° 0' kuzey-güney paralelleri arasında yer almaktadır. Batı Karadeniz Bölgesi, Türkiye'nin önemli heyelan alanlarından biridir. Çalışma bölgesinin meteorolojik olarak yoğun yağış alan ve bununla birlikte zayıf zemin ve heyelana duyarlı kaya formasyonlarına sahip oluşu nedeniyle heyelana meyilli bir bölge konumundadır. Çalışma alanının ortalama yıllık yağış miktarı yaklaşık 1.200 milimetre ve ortalama sıcaklık yaklaşık 9°C'dir (www.karabukyenice.bel.tr). Çalışma alanındaki topoğrafik yükseklik 42 m ile 1.998 m arasında değişirken, arazi eğimi 0° ile 70° arasında değişen dik ve sarp alanları içermektedir.



Şekil 1. Çalışma alanı ve konumu

Çalışma alanının litolojik birimlerini gösteren tematik harita 1: 500.000 ölçekli tematik harita Türkiye Maden Tetkik ve Arama Genel Müdürlüğü (MTA) tarafından 1998 yılında üretilmiştir. Şekil 2'de verilen çalışma alanına ait litoloji haritasında 9

litolojik birim bulunmaktadır. Çalışma bölgesinde yaygın olarak iki tip zemin yapısı (klastik ve karbonat kayaçlar ile metagranit) mevcuttur. Bölgenin yaklaşık 288 km²'sini (yaklaşık %42'si) kapsayan Kırıntılı ve karbonatlı kayaçlardır (k₁) ve çoğunlukla marn, türbiditik akıntılar, kumtaşı, kil taşı ve silt taşlarından oluşur. Bu bölgede kumtaşı, silttaşı ve marn değişimlerinden dolayı Jura-Alt kretase süreci heyelan aktivitelerinin en sık görüldüğü sahalardır. Bununla birlikte, heyelanlar g₁, k_{2s}, gsa ve j_{3k1} oluşumlarında daha az görülmektedir.



Şekil 2. Çalışma alanına ait litoloji haritası.

3. MATERYAL VE YÖNTEM

Bu çalışmada, Karabük Yenice ilçesine ait heyelan duyarlılık haritasının üretilmesi için ilk adım olarak AR metodu kullanılarak optimum faktörleri içeren model tespit edilmiştir. AR metodu ile belirlenen model sonrasında harita tahmin metotları (LR ve RO) ile heyelan duyarlılık haritaları üretilmiştir. Ayrıca RO metodunun performansının karşılaştırılması amacıyla literatürde yaygın olarak kullanılan LR yönteminden faydalanılmıştır. Çalışmadaki son adımda, genel doğruluk, kappa ile AUC değerleri ve ROC eğrisi yardımıyla üretilen haritaların

doğrulukları ve model performanslarına bakılmıştır. Ayrıca, üretilen tüm modellerin birbirleriyle olan farklılıkları ve istatistiksel anlamlılıkları değerlendirilerek optimum heyelan duyarlılık haritasına karar verilmiştir.

Heyelanda etkili faktörleri içeren modelin tespiti ve heyelan duyarlılık haritası üretimi için kullanılacak faktörlerin elde edilmesinde sayısal yükseklik modeli (SYM), uydu görüntüsü ve jeolojik haritalardan faydalanılmıştır. Çalışma bölgesine ait 42 m - 1998 m aralığındaki SYM verisi Harita Genel Komutanlığına ait 1/25.000'lik eş yükseklik eğrilerinden üretilmiştir. 30 metre çözünürlüklü raster formatta dönüştürülen SYM'den eğim, eğim uzunluğu, bakı, yükseklik, drenaj yoğunluğu, Tİİ, TPI, akarsu güç indeksi ve sediment taşıma indeksi haritaları elde edilmiştir. 2016 yılına ait Landsat OLI uydu verisi kullanılarak arazi örtüsü/kullanımı haritası üretilmiştir. Çalışma alanına ait litoloji haritasını elde edilmesinde 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 normalize edilmiştir. Çalışmada kullanılan

tüm faktörlerin özellikleri, temin kaynakları ve veri türüne bilgiler aşağıda Tablo 1'de verilmiştir.

Çalışma çerçevesinde genel olarak dönel kayma, ötelenmeli kayma ve toprak akması olmak üzere üç tip heyelan görülmektedir. Dönel ve ötelenmeli kaymalar sık görülürken, toprak akmaları daha az ve dik yamaçlarda görülmektedir (Ercanoğlu vd., 2008). Çalışma alanında toplam 159 adet heyelan poligonu sayısallaştırılarak CBS ortamına aktarılmıştır (Şekil 1). Poligon türündeki heyelan verileri 30x30 m'lik raster veri formatına dönüştürülmüştür. Heyelan envanteri içerisindeki heyelan olan alanlar için "1" ve heyelan olmayan alanlar içinse "0" değerleri atanmıştır. Çalışmada, AR işlemi, heyelan duyarlılık haritası üretimi işlemi ve doğrulama işlemlerinde heyelan olan alanlar haricinde olmayan alanlarda gereklidir. Heyelan olmayan alanların tespitinde eğimin 5 dereceden az olma koşulu, nehir, göl gibi düz, topoğrafik açıdan elverişsiz alanlar ve heyelana etkin faktörler olan alanlarla karşılaştırılarak frekans karşılaştırmaları incelenerek karar verilmiştir (Gómez ve Kavzoglu, 2005). Bartın-Ulus çalışma çerçevesi için düzenlenen heyelan envanter verilerine ait bilgiler Tablo 2'de sunulmuştur.

Tablo 1. Çalışmada kullanılan veri kaynak ve türleri.

Veri Özelliği	Veri	Veri Kaynağı	Veri Türü
Yüzey jeolojisi	Heyelan envanter verisi Litoloji	Jeolojik veri	Poligon
Yüzey Topoğrafya bilgisi	Yükseklik Bakı Eğim Eğim uzunluğu Topoğrafik pürüzlülük indeksi	SYM	Raster
Yüzey-su ilişkisi	Topoğrafik ıslaklık indeksi Akarsu güç indeksi Sediment taşıma indeksi Drenaj yoğunluğu	SYM	Raster
Yüzey kullanım bilgisi	Arazi kullanımı/örtüsü	Uydu görüntüsü	Raster
Bitki örtüsü	NDVI	Uydu görüntüsü	Raster

Tablo 2. Karabük-Yenice heyelan envanter bilgisi.

Envanter Türü	Adet (Poligon)	Piksel Sayısı	Alan			
			Toplam (km ²)	Ortalama (m ²)	Maksimum (m ²)	Minimum (m ²)
Olan	159	27294	245,070	158.752	770.910	3.625
Olmayan	18	3.012	2,72	150.973	282.762	33.629

a. Adımsal Regresyon (AR)

AR, çok sayıda verinin değerlendirileceği işlemlerde optimum sayıda veri setinin belirlemeye yarayan ve optimum faktör modelinin kurulduğu regresyona dayalı otomatik hesaplama yöntemidir (Neter vd., 1996; Saltelli vd., 2000). Regresyon analizlerindeki amaç, iki veya daha fazla bağımsız değişken ile bir bağımlı değişken arasındaki ilişkiyi belirlemek ve bu ilişkiyi bir modelle tanımlayabilmek amacıyla yapılır. AR analizlerinde olduğu gibi en ideali bağımsız değişkenlerin veya faktör modelinin belirlenmesinde bağımsız değişkenlerin kendi aralarındaki korelasyonun düşük ama bağımlı değişken ile korelasyonunun yüksek olması şartı aranmaktadır. AR analizi yöntemi işlem adımları incelendiğinde;

- Bağımsız değişkenlerin eşitliğe hangi sırayla sokulacağı tamamen istatistiksel bir kriterle belirlenir.
- Bağımsız değişkenler arasında en yüksek iki değişkenli korelasyona sahip olan değişken modelin ilk en önemli değişkeni olarak regresyon modeline sokulur.
- İkinci bağımsız değişken, ilk bağımsız değişkenin tahmini için hesaplanan R²'de en yüksek artışı üreten değişkendir.
- İkinci bağımsız değişken eklendikten sonra, birinci bağımsız değişkenin istatistiksel olarak anlamlı bir belirteç olup olmadığına ikinci bir istatistiksel anlamlılık testi ile bakılır ve değilse regresyon modelinden çıkarılır.
- Bu işlem, tüm bağımsız değişkenlerin regresyon denkleminde girilinceye kadar veya kalan bağımsız değişkenlerin R²'de istatistiksel olarak anlamlı bir artış sağlamayana kadar tekrarlanır.
- Sonuç model en fazla katkı sağlayan değişkenler kümesini içerir.

AR analizinin dezavantajı, modelde olmayan bütün değişkenlerin olasılık tahminlerinin her adımda tek tek hesaplanması gerekliliğidir. Söz konusu durumlarda çok değişkenli yüksek boyutlu modellemelerin hesaplanma zorluğu ve zamansal sıkıntılara neden olabilir.

b. Lojistik Regresyon (LR)

Heyelan duyarlılık haritalarının üretilmesinde en çok kullanılan çok değişkenli analiz yöntemlerinden biri LR yöntemidir. LR birçok yer bilimi araştırmalarında olduğu gibi heyelan duyarlılık analizlerinde de sıklıkla kullanılan istatistiksel tabanlı bir modeldir (Yılmaz, 2010). LR yönteminin 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 değişkenler ise modelde kullanılan ve heyelan olayına etki eden faktörlerdir. Bağımlı değişkenin ikili değişken olması durumunda, lojistik fonksiyon uygulanabilir (Atkinson ve Massari, 1998). Heyelana duyarlı alanların tespiti gibi çalışmalarda bağımlı değişken girdi verisi 0 ya da 1 olmalıdır. LR modelinin dayandığı temel lineer eşitlik aşağıdaki gibi gösterilebilir,

$$Y = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_n X_n \quad (1)$$

Bu eşitlikteki Y, 0 ve 1 aralığındaki bağımlı değişkeni, B₀ bağımsız değişkenlerin sıfır değerini aldıklarındaki bağımlı değişken değeri ya da sabiti, B₁, ..., B_n bağımsız değişkenlerin regresyon katsayılarını ve X₁, ..., X_n ise bağımsız değişkenleri göstermektedir. LR modelindeki bağımlı değişkenler aşağıdaki gibi tanımlanır.

$$\text{Logit}(p) = \ln\left(\frac{p}{1-p}\right) = 1/1 + e^{-B_0 + B_1 X_1 + B_2 X_2 + \dots + B_n X_n} \quad (2)$$

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

LR sonucu elde edilen her bir regresyon katsayısı risk faktörünün sonuca olan katkısının büyüklüğünü ifade etmektedir. Pozitif regresyon katsayısı sonuç olasılığının arttığını ve doğru orantılı bir ilişkiye sahip olduğunu gösterir. Negatif bir regresyon katsayısı ise negatif yönde yani ters orantılı bir ilişkiyi göstermektedir. Artan bir regresyon katsayısı sonuç olasılık üzerinde güçlü bir etki veya ağırlığa sahip olduğunu gösterirken, regresyon katsayısı sıfıra yaklaştığında sonuç olasılık üzerindeki etkisinin ve ağırlığının az olduğunu ifade etmektedir. LR yönteminin avantajları, bağımsız değişkenler arasında doğrusal ilişkilerin bulunmaması ve değişkenlerin eşit istatistiksel varyansları olduğu kabul etmemesidir. Bir diğer avantajı değişkenler süreklilik veya kesikli ya da kombinasyonunda nominal, eşit aralıklı veya oransal ölçüm yapabilmesidir. Ayrıca birçok araştırmacı tarafından LR yaklaşımının en uygun veri tabanlı heyelan duyarlılık haritalama yaklaşımı olduğu öne sürülmüştür (Carrara vd., 1991; Brenning, 2005).

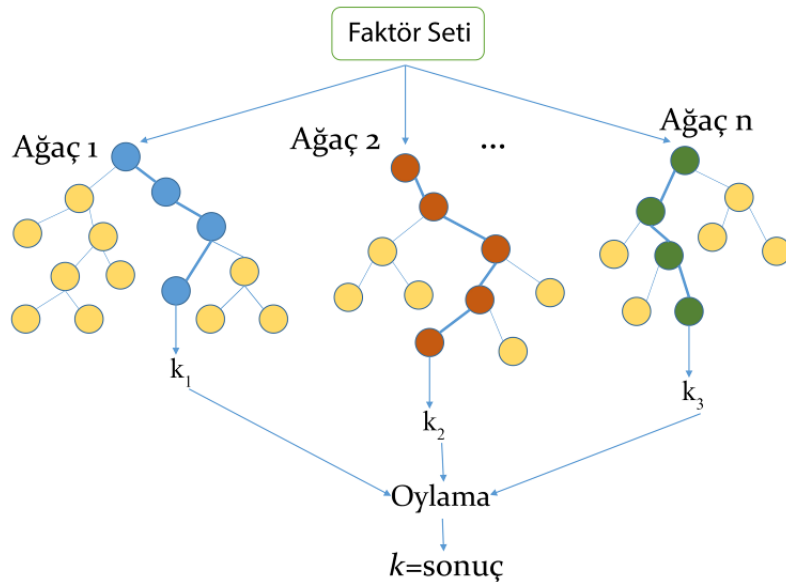
c. Rastgele Orman (RO)

RO özellik seçimi yaklaşımı Breiman (2001) tarafından geliştirilmiştir. RO algoritmasının temel prensibi, rastgele seçilmiş çok sayıda ve çok değişkenli karar ağaçlarının her birini farklı eğitim kümeleri ile eğitilmesi sonucu ortaya çıkan kararların birleştirilmesidir (Breiman, 2001). N kadar örnek ve m kadar açıklayıcı ile ormandaki

her bir ağaç yapısı aşağıdaki gibi kurulmaktadır (Reif vd., 2006):

- Tüm veri setinin değiştirilmesi ile, N kadar örneklem içinden bir eğitim seti seçilir,
- Ağaçtaki her düğüm, tüm veri setindeki M kadar özellik içerisinden rastgele m kadar özellik seçilir,
- m kadar özellik içerisinde en iyi ayrılmış kök belirlenir,
- Ağaç tamamen büyüyene kadar ikinci ve üçüncü adımlar yinelenerek devam edilir.

RO algoritması dallanma kriterlerinin belirlenmesinde veya dallanma için özneliklerin seçiminde Gini indeksi yöntemini kullanır. Bu algoritmanın işlemesi için kullanıcı tarafından belirlenmesi gereken iki parametre mevcuttur. Bu parametrelerden ilki her bir düğümde kullanılacak örneklem sayısı ve ikinci parametre ise oluşturulacak ağaç sayısıdır. Kurulan ağaç yapılındasında düğümler en iyi dallanma için seçilen özellikleri araştırır. RO algoritması içerisinde kullanıcı tarafından belirlenen ve oluşturulan ağaç sayıları n ile ifade edilir. Sonuç olarak, algoritma bu n sayıda ağaçtan elde edilen n sayıda oy arasından en fazla oya veya ağırlığa sahip olan faktörü seçer ve bu şekilde yeni veri seti belirler (Çölkese, 2015). RO algoritması ait temel çalışma prensibi Şekil 3'te gösterilmiştir.



Şekil 3. Rastgele orman algoritmasının temel çalışma prensibi.

4. UYGULAMA

Çok sayıda faktör veya özelliği içeren veri setinin duyarlılık haritası üretimi işleminde kullanılması durumunda veri içerisinde bozucu etkiye ya da yüksek korelasyona sahip özelliklerin dikkate alınmasına neden olmaktadır. Ayrıca artan faktör sayısının doğruluk üzerinde olumlu etki yaratmadığı tam tersi faktör boyutun artmasına paralel olarak belirli bir noktadan sonra harita doğruluğunun sabit kaldığı veya düştüğü bilinmektedir. Bu nedenle veri setindeki optimum faktör sayısının tespit işlemi hem işlem yükünün azaltılması hem de boyutsallık probleminin (Hughes fenomeni) önüne geçilerek harita doğruluğunun artırılmasına yardımcı olacaktır (Hughes, 1968). Faktör sayısının azaltılması özellik çıkarımı veya seçimi olarak bilinen yaklaşımlarla gerçekleştirilmektedir. Özellik seçimi neticesinde yüksek boyutlu veri setini temsil eden daha düşük boyutlu ve az korelasyona sahip bir veri seti tanımlanmaktadır. Söz konusu veri seti düşük korelasyona sahip olması nedeniyle sadece istenilen amaca yönelik yararlı bilgileri içeren seçilmiş özelliklerden oluşmaktadır. Çalışma kapsamında optimum faktörlerin tespiti ve en iyi faktör modelinin belirlenmesi amacıyla AR metodu kullanılmıştır.

a. AR ile Optimum Faktör Model Tespiti

Optimum faktör sayısının tespiti ve faktör modelinin elde edilmesi amacıyla SPSS v.20 yazılımı kullanılarak AR işlemi yapılmıştır. Optimum faktör modelinin tespitinde ihtiyaç duyulan eğitim verisi çalışma bölgesine ait heyelan olan ve heyelan olmayan alanları içeren heyelan envanter haritasından faydalanılmıştır.

Eğitim ve test verisi içerinden rastgele örneklem metodu yardımıyla envanter haritasındaki verinin %70'i eğitim verisi ve %30'u ise test verisi olacak şekilde belirlenmiştir. Karabük-Yenice çalışma bölgesi için AR metodu kullanılarak elde edilen model ve modellerin istatistiksel sonuçları aşağıda Tablo 3'de verilmiştir. Tablo 3'de verilen R mevcut ve tahmini değerler arasındaki basit doğrusal Pearson korelasyon kat sayısı, R^2 (belirleme katsayısı) örnek regresyon denkleminin göreceli etkinliğini belirlemede kullanılan korelasyon katsayısının karesi ve düzeltilmiş R^2 ise belirleme katsayısını hatadan arındırılmış şeklidir. Kurulan regresyon modeline göre (Tablo 3) tüm faktör setinin bağımlı değişken üzerindeki ortak etkileri incelenmiştir. AR işleminde, eklenen ya da çıkarılan faktörlerin bağımlı değişken üzerinde anlamlılığına bakılır ve bu işlem hiçbir faktörün modele anlamlı bir şekilde katkı edemeyeceği duruma (modele eklenecek anlamlı faktör kalmayana kadar) kadar devam eder. Bu çalışmada ise Model-6 dışında kalan faktörlerin %95 güven aralığı için *p-value* değerleri 0,05'in üzerinde çıktığı ve bağımlı değişken üzerindeki etkisinin anlamlı olmadığı tespit edilmiştir. Böylece en az sayıda (Model-6) faktör yardımı ile model kurulmuştur. Optimum model analizde, bağımlı değişken üzerinde katkısı az olan drenaj yoğunluğu, arazi örtüsü/arazi kullanımı, eğim uzunluğu, akarsu güç indeksi, sediment taşıma indeksi ve TPI faktörleri optimum model dışında kalmıştır. Sonuç olarak 12 faktör setinden oluşan veri setinin yarısı elenerek 6 adet faktörden oluşan Model-6 optimum faktör setine sahip model olarak kabul edilmiştir.

Tablo 3. AR analizi sonucu elde edilen modeller.

Model Özeti				
Model	R	R^2	Düzeltilmiş R^2	Standart Tahmin Hatası
1	0,676 ^a	0,458	0,457	0,369
2	0,756 ^b	0,571	0,570	0,328
3	0,774 ^c	0,598	0,597	0,318
4	0,786 ^d	0,618	0,616	0,310
5	0,794 ^e	0,630	0,628	0,305
6	0,795 ^f	0,632	0,629	0,305
Model-1: NDVI		Model-4: NDVI, Yükseklik, Bakı, Litoloji		
Model-2: NDVI, Yükseklik		Model-5: NDVI, Yükseklik, Bakı, Litoloji, Tİİ		
Model-3: NDVI, Yükseklik, Bakı		Model-6: NDVI, Yükseklik, Bakı, Litoloji, Tİİ, Eğim		

R-kare istatistiğinin minimum değeri 0 ve maksimum değer ise 1'dir. R^2 değeri ne kadar yüksek olursa modelinde o kadar iyi olduğunu gösterir. R^2 istatistik değerler sonuçlarına göre, 0,2 değerine kadar ki R^2 değeri modelin kabul edilebilir olduğunu, 0,2 ve 0,4 değer aralığı iyi dereceli bir modeli ve 0,5 değeri üzerinde ise modelin ise faktörlerin modelle çok iyi derecede uyduğunu gösterir niteliktedir (Wulf, 2014). AR sonucu elde edilen istatistiksel değerler Model-6 için irdelendiğinde, R^2 değeri 0,632 ve düzeltilmiş R^2 değeri ise 0,629 olarak bulunmuştur. Elde edilen R^2 değerine göre bağımlı değişken (heyelan envanter verisi) ile bağımsız değişkenler (heyelanda etkili faktörler) arasında orta kuvvette bir ilişki olduğunu göstermektedir. Çalışmada elde edilen bir diğer değer olan düzeltilmiş R^2 değerine bakıldığında ise, Model-6 ile oluşturulan regresyon modelinde yer alan bağımsız değişkenlerin, bu çalışmada örnek olarak seçilen diğer heyelanda etkili faktörlerin birçoğunu kapsadığını ve açıklama gücünün %63 olduğunu göstermektedir. Bir diğer istatistik sonuç standart tahmin hatasıdır ve bu hatanın küçük olması elde edilen parametrelerin tahmini açısından önemlidir ve bu hatanın "0" yakın çıkması istenir. Tahmin modeller için elde edilen standart tahmin hataları incelendiğinde Model-6'nın 0,305 ile en az hata katsayısına sahip olduğu görülmüştür.

b. LR Metodu ile Heyelan Duyarlılık Haritası Üretimi

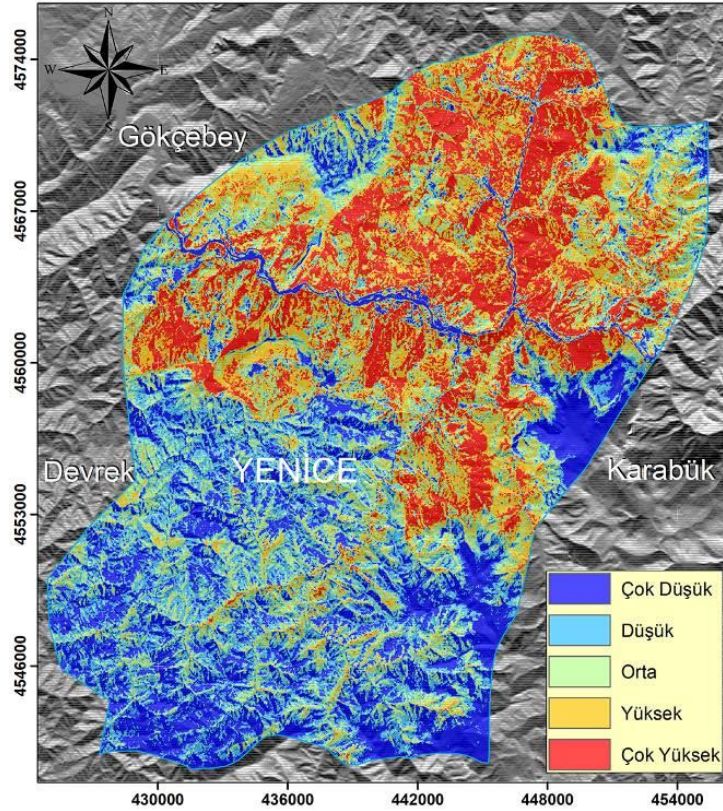
Heyelan duyarlılık haritalarında istatistiksel seçime dayalı AR yönteminin performansının incelenmesi ve heyelan duyarlılık haritaları üretimi için LR faydalanılmıştır. Çalışmada AR analizin vermiş olduğu optimum faktör setine sahip Model-6 kullanılarak heyelan duyarlılık haritası üretilecektir. LR yöntemi literatürde heyelan duyarlılık haritası üretimi çalışmalarında

sıklıkla kullanılan bir yöntem olmakla birlikte ayrıca literatüre önerilen yeni yöntemlerin performanslarının karşılaştırılmasında "benchmark" olarak kullanılan bir yöntemdir. Bu çalışmada LR yöntemi Model-6 için uygulanmış ve regresyon analiz sonuçları olan B regresyon katsayısı, SH standart hata, Wald testi değeri ve Sig anlamlılık değerleri Tablo 4'te verilmiştir.

Model-6 LR analiz sonuçlarına göre (Tablo 4) yükseklik faktörü hariç diğer faktörler pozitif B katsayılarına sahiptir. Pozitif katsayıya sahip olan faktörlerin olasılık değerlerinin pozitif yönlü arttığının bir göstergesidir. LR katsayıları incelendiğinde, NDVI faktörü tüm faktörler içerisinde +21,368 katsayısına sahip en etkili faktör olduğu tespit edilmiştir. Bir diğer istatistiksel değer olan Wald değerinin 2'den büyük değerler alması anlamlı kabul edilmekle beraber, Wald değeri büyüdükçe p anlamlılık düzeyi küçülmektedir (p anlamlılık düzeyi $>0,05$). Model-6 içerisindeki heyelan duyarlılığına etki eden faktörlerin Wald istatistik değerlerine bakıldığında, tüm faktörlerin Wald değeri 2'den büyük çıktığı ve modeldeki tüm faktörlerin anlamlı oldukları tespit edilmiştir. Tablo 4'deki $Exp(B)$ değerleri incelendiğinde 1 değerine uzak olan NDVI, bakı, litoloji ve eğim faktörlerin bağımlı değişken değişimine önemli etkisi olduğu, Tİİ faktörünün ise 1'e en yakın değer olduğu içinde bağımlı değişken üzerinde en etkisiz faktör olduğu tespit edilmiştir. Ayrıca Tablodaki "Sig" kolonu ise faktörlerin anlamlılık düzeyini göstermekte olup SPSS'te varsayılan olarak $p<0,05$ olarak çalışmaktadır. Modeldeki faktörlerin Sig değerleri sıfıra ne kadar yakın olursa anlamlılık düzeyleri o kadar fazla demektir. Böylelikle Tablo 4'deki faktörlerin Sig değerleri incelendiğinde tüm faktörlerin istatistiksel olarak anlamlı oldukları görülmektedir.

Tablo 4. Lojistik regresyon analizi sonucu elde edilen model sonuçları.

	B	S.H.	Wald	df	Sig	Exp(B)
NDVI	21,368	1,196	318,968	1	0,000	1,905
Yükseklik	-0,004	0,000	193,404	1	0,000	0,996
Litoloji	0,117	0,025	22,284	1	0,000	1,124
Bakı	0,366	0,048	57,937	1	0,000	1,442
Eğim	-0,046	0,010	19,255	1	0,000	0,955
Tİİ	0,000	0,000	17,591	1	0,000	1,000
Sabit	-8,541	0,700	148,779	1	0,000	0,000

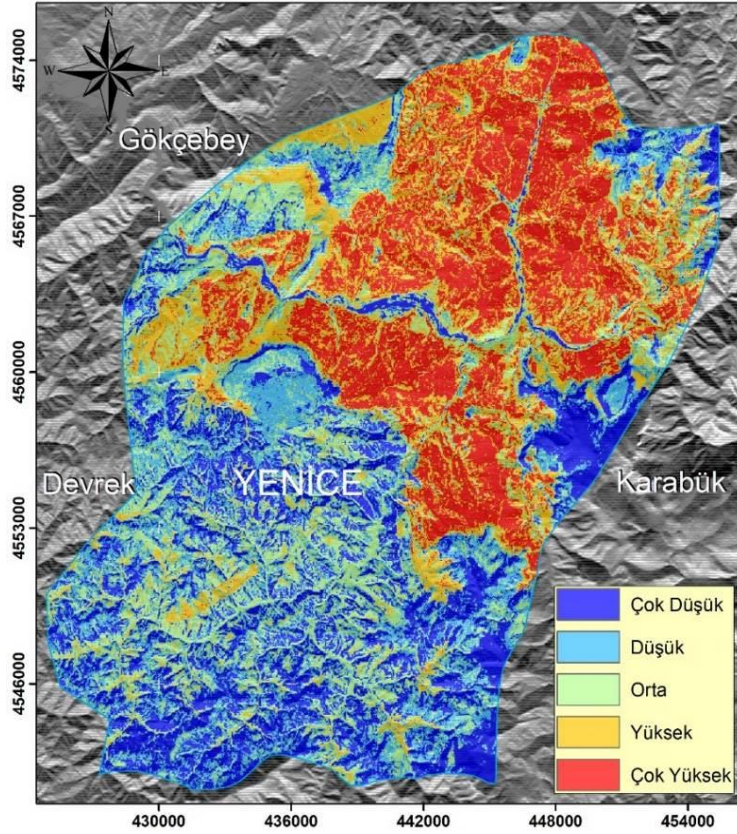


Şekil 4. Model-6 veri seti için LR metodu ile üretilmiş heyelan duyarlılık haritası.

LR metodu kullanılarak Model-6 veri seti ve tüm faktörleri kapsayan Model-12 için heyelan duyarlılık haritası üretilmiştir. Duyarlılık haritalarının tematik harita gösterimi ve değerlendirilmesi safhasında tüm harita eşit dağılım (quantile) esasına dayanılarak 5 sınıfta (çok yüksek, yüksek, orta, düşük ve çok düşük) bölümlendirilmiştir (Şekil 4). LR yöntemiyle üretilmiş Model-6 ait heyelan duyarlılık haritasındaki sınıfların test alanları ile karşılaştırılması ile elde edilen yüzdesel dağılımları incelendiğinde, %38 çok yüksek, %35'u yüksek, %15'i orta, %7'ü düşük ve %5 ise çok düşük duyarlı alanları içerdiği görülmüştür. Elde edilen tematik haritanın yüksek ve çok yüksek riskli alanlardaki toplam doğruluk değeri dikkate alındığında ise yaklaşık %73 olduğu görülmektedir. Çalışma bölgesinde meydana gelen heyelan dağılımı incelendiğinde, yüksek duyarlılık alanların kuzey ve orta kesimlerde ve düşük duyarlılık alanların ise güney kesimlerde yoğunlukta olduğu görülmektedir.

c. RO Metodu ile Heyelan Duyarlılık Haritası Üretimi

RO metodu ise son yıllarda karmaşık yapıdaki birçok sınıflandırma ve regresyon probleminin çözümünde kullanılan bir veri madenciliği tekniğidir. Birçok regresyon probleminin çözümünde kullanılan rasgele orman yöntemi bu çalışmada kurulan modellerin performanslarının tespitinde ve çalışma alanları için optimum heyelan duyarlılık haritası üretimi amacıyla kullanılmıştır. RO algoritmasında ağaç oluşumu için her bir düğümde kullanılan ağaç sayısı kullanıcı tarafından belirlenmesi gereken parametrelerdir. Bu nedenle bu çalışmada tüm modeller için ağaç sayısı geliştirilmiş hata grafiğine bakılarak 250 olarak belirlenmiştir. Regresyon ağacı oluşturulması ve duyarlılık haritasının üretilmesi işlemleri MATLAB ortamında yazılmış programlarla gerçekleştirilmiştir. RO algoritmasının Model-6 için uygulanarak çalışma bölgesi için 5 sınıf aralığında sınıflandırılmış tematik harita elde edilmiştir (Şekil 5). RO metodu kullanarak hem AR metodunun belirlendiği Model-6 ve tüm faktörleri kapsayan Model-12 için heyelan duyarlılık haritaları üretilerek karşılaştırılmıştır.



Şekil 5. Model-6 veri seti için RO metodu ile üretilmiş heyelan duyarlılık haritası.

RO yöntemi Model-6 ile üretilmiş heyelan duyarlılık haritasındaki sınıfların test alanları ile karşılaştırılması ile elde edilen yüzdesel dağılımları incelendiğinde, %54 çok yüksek, %29'u yüksek, %8'i orta, %3'ü düşük ve %6 ise çok düşük duyarlı alanlara denk gelmiştir. Elde edilen tematik haritanın yüksek ve çok yüksek riskli alanlardaki toplam doğruluk değeri dikkate alındığında ise yaklaşık %83 olduğu görülmektedir. Çalışma bölgesinde meydana gelen heyelan dağılımı incelendiğinde, yüksek duyarlılık alanlarının özellikle çalışma alanının kuzeyinden orta kesimlere kadar yoğunlukta olduğu görülmektedir. Heyelan dağılımına paralel olarak duyarlılık haritasında bu bölgelerde yoğunlaştığı açıkça görülmektedir.

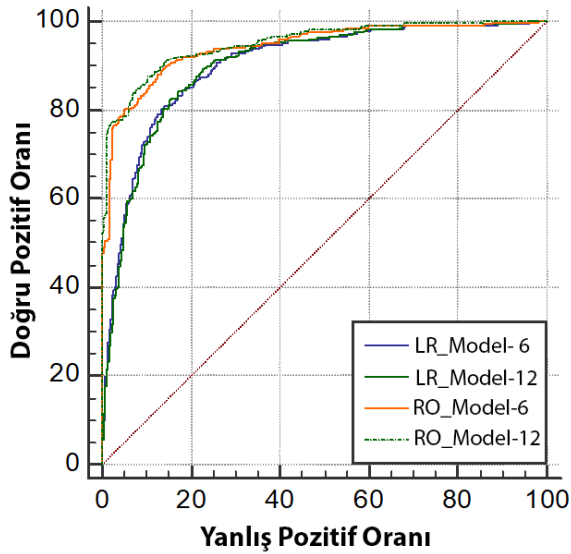
ç. Model Performansları

LR ve RO algoritmaları kullanılarak üretilen heyelan duyarlılık haritalarının performanslarının analiz edilmesi amacıyla genel doğruluk, ROC (Alıcı İşlem Karakteristikleri, Receiver Operating Characteristic) eğrisi, AUC (Eğri Altında Kalan, Area Under Curve) ile kappa değerlerine bakılmıştır. Üretilen heyelan duyarlılık haritalarının performans değerlendirilmelerinde

hata matrislerinde faydalanılmıştır ve hata matrisi hesaplamasında tematik haritalardaki çok yüksek ve yüksek sınıfları "1" ve orta, düşük ve çok düşük sınıflar ise "0" olarak değerlendirilmiştir. Otomatik faktör seçimi yaklaşımının heyelan duyarlılık haritası üretimindeki etkilerinin karşılaştırılması amacıyla AR yaklaşımıyla elde edilen Model-6'nın ve tüm faktörlerinde bulunduğu Model-12 (bakı, drenaj yoğunluğu, litoloji, eğim, arazi kullanımı/örtüsü, yükseklik, NDVI, eğim uzunluğu, TII, TPI, sediment taşıma indeksi ve akarsu güç indeksi) için doğruluk ve performans karşılaştırma analizleri yapılmıştır.

ROC eğrileri model performansını test etmek için birçok alanda uygulanır. ROC eğrileri ile iki sınıf arasındaki ayrımı ve sınıflandırıcının performansını görselleştirmek mümkündür (Swets, 1988). ROC eğrileri analizlerde test ve eğitim verilerinden faydalanan ve işlemlerde ayırt ediciliği göstermekle beraber, farklı modellerin performans açısından karşılaştırılmasında, eğri altında kalan alana (AUC) gereksinim olur. AUC bir olayın meydana gelme ya da gelmeme durumunu güvenli bir biçimde tahmin etme yeteneğini tanımlayarak oluşturulan modelin kalitesini ifade etmektedir (Akgün ve Türk, 2010).

AUC değerinin 1'e yakınlığı ideal durumu gösterirken, 0,5'e yakın bir değer ise söz konusu modelin doğruluğunun oldukça az olduğunu işaret etmektedir. Bu çalışmada, AR yaklaşımını belirlediği Model-6 ve tüm faktör kümesini içeren Model-12'nin model performansları Şekil 6'da verilen ROC eğrisi ve AUC değer sonuçlarına bakılarak incelenmiştir. AUC değerleri incelendiğinde sırasıyla LR metodu Model-6 için 0,903 ve Model-12 için 0,902 olarak bulunmuştur. RO algoritması AUC değerleri ise Model-6 için 0,945 ve Model-12 için 0,952 elde edildiği görülmektedir.



Şekil 6. Model-5 ve Model-12 ile üretilmiş heyelan duyarlılık haritasına ait ROC eğrisi.

Şekil 6'da verilen ROC eğrileri ve AUC değerleri her iki metot için göz önünde bulundurulduğunda RO algoritmasının LR algoritmasından daha iyi performans gösterdiği söylenebilir. Fakat harita tahmin metotları kendi içerisinde değerlendirildikleri takdirde modellerin birbirleriyle büyük performans farkları içermediği görülmektedir. Böylelikle sadece 6 veri seti ile oluşturulan modelin tüm veri seti ile oluşturulan modelle aynı performansı gösterdiği tespit edilmiştir. Elde edilen bu sonucun istatistiksel olarak da yorumlanabilmesi için ROC eğrilerinin karşılaştırıldığı ve istatistiksel anlamlılıklarının incelendiği bir analiz yapılmıştır (Tablo 5).

LR metodu için Model-6 ve Model-12'te ait ROC eğrisi performanslarının istatistiksel farklılıklarının ayırt edilmesi ve tespiti amacıyla %95 güven aralığı düzleminde karşılaştırılmıştır. Buna göre Model-6 ve Model-12'nin ROC eğrisi model performansları karşılaştırıldığında %95 güven aralığı için p değeri 0,874 olarak

bulunmuştur. Bu da teknik açıdan iki modelin güven aralığı eşik değeri olan 1,96'dan küçük çıktığı için istatistiksel olarak anlamsız oldukları, yani iki modelinde aynı oldukları görülmüştür (Tablo 5). RO algoritması ile elde edilen harita modellerinin ROC eğrisi sonuçları karşılaştırıldığında ise modeller arasındaki değer farklarının istatistiksel açıdan anlamsız oldukları fakat LR ile üretilen modellerden farklı olduğu görülmektedir. Sonuç olarak, RO algoritması ile elde edilen Model-6'ya ait heyelan duyarlılık haritası LR haritalarından istatistiksel olarak farklı olduğu ve daha iyi model performansı gösterdiği istatistiksel olarak tespit edilmiştir.

Tablo 5. ROC eğrilerinin ikili karşılaştırma sonucu elde edilen istatistiksel sonuçları.

	LR Model-6	LR Model-12	RO Model-6	RO Model-12
LR Model-6	-	0,874	6,030	7,004
LR Model-12		-	5,962	6,910
RO Model-6			-	1,265
RO Model-12				-

Not: Koyu olarak gösterilen ve 1,96'dan daha yüksek test değerleri, %95 güven düzeyinde iki model arasında istatistiksel olarak önemli bir fark olduğunu göstermektedir.

Heyelan duyarlılık haritalarının model performanslarının değerlendirilmesinde bir diğer doğruluk test araçları olan genel doğruluk ve kappa değerleri incelenmiştir. Çalışma bölgesi için üretilen duyarlılık haritalarına ait genel doğruluk ve kappa değerleri Tablo 6'da verilmiştir.

Tablo 6. Heyelan duyarlılık haritaları için elde edilen genel doğruluk ve kappa değerleri.

	Genel Doğruluk	Kappa
LR Model-6	80,87	0,602
LR Model-12	81,23	0,624
RO Model-6	88,27	0,758
RO Model-12	88,10	0,762

Elde edilen test değerleri incelendiğinde LR metodu ile üretilmiş haritalardaki modeller

arasında yaklaşık %0,36 ve RO için ise yaklaşık %0,17 gibi çok düşük doğruluk farkları olduğu görülmüştür. Bu nedenle üretilen duyarlılık haritaları içerisindeki optimum modelin tespiti için McNemar's istatistik anlamlılık testi kullanılmıştır (Tablo 7). Hesaplanan istatistik kritik değerinin %95 güven aralığı için 3,84'ten büyük olması üretilen tematik harita doğrulukları arasındaki farkın istatistiksel olarak anlamlı olduğu, değilse anlamsız olduğu söylenebilir.

Tablo 7. Heyelan duyarlılık haritası modellerinin McNemar's testi kullanılarak istatistiksel karşılaştırılması.

		McNemar's (z=3,84)	Anlamlılık (%95)
LR Model-6	/ LR Model-12	3,375	Anlamsız
LR Model-6	/ RO Model-6	40,347	Anlamlı
LR Model-6	/ RO Model-12	43,457	Anlamlı
LR Model-12	/ RO Model-6	30,009	Anlamlı
LR Model-12	/ RO Model-12	34,454	Anlamlı
RO Model-6	/ RO Model- 12	0,003	Anlamsız

McNemar's testi sonuçları incelendiğinde RO algoritması kullanılarak üretilen heyelan duyarlılık haritalarından Model-6 ve Model-12 arasındaki genel doğruluk farkının anlamsız olduğu görülmüştür. Böylelikle AR yaklaşımı ile belirlenen faktörlerin (NDVI, Yükseklik, Bakı, Litoloji, Tlİ, Eğim), Karabük ili Yenice ilçesi için üretilen heyelan duyarlılık haritalarında kullanılacak optimum faktörler olabileceği tespit edilmiştir. Ayrıca LR metodu içinde benzer sonuçlar elde edilerek Model-6 ile üretilmiş duyarlılık haritasının optimum sonuçlar verdiği ve Model-12 ile aynı özelliklere sahip bir tematik harita olduğu görülmektedir.

5. SONUÇLAR

Bu çalışmada, Karabük ilinin Yenice ilçesi için heyelana duyarlı alanların tespiti ve sınıflandırılması amacıyla heyelan duyarlılık analizleri gerçekleştirilmiş ve heyelan duyarlılık haritaları üretilmiştir. Çalışmada heyelan duyarlılık haritası üretiminde kullanılan faktör sayısı ve söz konusu faktörlerin kalitesini belirlemek ve heyelan duyarlılık haritası üretimindeki etkilerinin araştırılması amacıyla

otomatik faktör modeli kurulumuna dayalı AR yöntemi kullanılmıştır. Ayrıca, seçilen faktörler yardımıyla heyelana duyarlı alanların tahmin edilmesi ve heyelan duyarlılık haritası üretiminde RO ve LR yöntemleri kullanarak güncel makine öğrenme tabanlı yaklaşım ile geleneksel yaklaşımların performansları karşılaştırılmalı olarak analiz edilmiştir. Çalışma alanı için 12 adet heyelanda etkin faktör belirlenmiştir. Söz konusu faktörler AR işlemi ile değerlendirildiğinde 6 faktör (NDVI, Yükseklik, Bakı, Litoloji, Tlİ ve eğim) setine dayalı Model-6'nın optimum faktör modeli olduğu AR analizi ile tespit edilmiştir. Model-6 ve tüm faktörleri içeren Model-12 RO ve LR yöntemleri kullanılarak heyelan duyarlılık haritaları üretilmiştir. Üretilen heyelan duyarlılık haritalarının, genel doğruluk ve kappa değerleri ile ROC eğrisi ve AUC değerlerine bakılarak doğruluk analizleri gerçekleştirilmiştir. Bunlara ilave olarak, farklı modeller ile elde edilen doğruluk farklılıklarının anlamlılıkları McNemar's testi ile analiz edilerek yorumlanmıştır. Üretilen duyarlılık haritaları içerisinde en iyi performansı değerini Model-12 ve RO algoritması ile üretilmiş heyelan duyarlılık haritasını olduğu görülmüştür. Fakat Model-6 ile Model-12 arasındaki doğruluk farklarının anlamlılıkları hem ROC eğrisi hem de McNemar's testi ile analiz edildiğinde bu iki modelin aralarındaki farkların anlamsız oldukları görülmüştür. Sonuç olarak, AR işlemi ile elde edilmiş optimum faktör modeli olan Model-6 RO algoritması ile yaklaşık %88,27 genel doğruluğa, 0,758 kappa değerine ve 0,945 AUC değerine sahip optimum model olduğu tespit edilmiştir. Ayrıca çok sayıda faktör kullanılan çalışmalarda otomatik seçim yaklaşımları ile optimum faktör setinin tespitinin mümkün olduğu ve sadece seçilen faktörler ile üretilen heyelan duyarlılık haritalarının genel doğruluk ve model performansının yüksek olduğu istatistiksel anlamlılık testleri ile de ortaya konulmuştur. Elde edilen bu sonuç ile, otomatik faktör modeli yaklaşımının kullanıldığında Karabük-Yenice çalışma alanı için %50 oranda veri seti boyutunun azaltılabileceği ve böylelikle büyük ve çok faktörlü çalışmalarda veri boyutunun azaltılmasının; işlem gücü, hafıza ihtiyacı, ağ bant genişliği ve depolama alanı gibi işlem süreci üzerinde etkili konularda zaman tasarrufu sağlayacağı öngörülmektedir.

KAYNAKLAR

Ahmed, B., (2015), **Landslide susceptibility modelling applying user-defined weighting and data-driven statistical techniques in Cox's Bazar Municipality, Bangladesh**, Natural Hazards, 79(3), 1707-1737.

- Akgun, A., Turk, N., (2010), **İki ve Çok Değişkenli İstatistik ve Sezgisel Tabanlı Heyelan Duyarlılık Modellerinin Karşılaştırılması: Ayvalık (Balıkesir, Kuzeybatı Türkiye) Örneği**, Jeoloji Mühendisliği Dergisi, 34, 85-112.
- Akgun, A., Bulut, F., (2007), **GIS-based landslide susceptibility for Arsin-Yomra (Trabzon, North Turkey) region**, Environmental Geology 51(8), 1377-1387.
- Atkinson, P. M., Massari, R., (1998), **Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy**, Computers & Geosciences, 24(4), 373-385.
- Ayalew, L., Yamagishi, H., (2005), **The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan**, Geomorphology 65(1-2), 15-31.
- Breiman, L., (2001), **Random forests**, Machine Learning, 45(1), 5-32.
- Brenning, A., (2005), **Spatial prediction models for landslide hazards: review, comparison and evaluation**, Natural Hazards and Earth System Sciences, 5(6), 853-862.
- Bui, D. T., Tuan, T. A., Klempe, H., Pradhan, B., Revhaug, I., (2016), **Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree**, Landslides 13(2), 361-378.
- Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., Reichenbach, P., (1991), **Gis Techniques and Statistical-Models in Evaluating Landslide Hazard**, Earth Surface Processes and Landforms, 16(5), 427-445.
- Catani, F., Lagomarsino, D., Segoni, S., Tofani, V., (2013), **Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues**, Natural Hazards and Earth System Sciences 13(11), 2815-2831.
- Colkesen, I., Sahin, E. K., Kavzoglu, T., (2016), **Susceptibility mapping of shallow landslides using kernel-based Gaussian process, support vector machines and logistic regression**, Journal of African Earth Sciences 118: 53-64.
- Çölkesen, İ., (2015), **Yüksek Çözünürlüklü Uydu Görüntüleri Kullanarak Benzer Spektral Özelliklere Sahip Doğal Nesnelerin Ayırt Edilmesine Yönelik Bir Metodoloji Geliştirme**, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, Doktora Tezi.
- Ercanoğlu, M., Kasmer, O., Temiz, N., (2008), **Adaptation and comparison of expert opinion to analytical hierarchy process for landslide susceptibility mapping**, Bulletin of Engineering Geology and the Environment, 67(4), 565-578.
- Ercanoğlu, M., Temiz, F. A., (2011), **Application of logistic regression and fuzzy operators to landslide susceptibility assessment in Azdavay (Kastamonu, Turkey)**, Environmental Earth Sciences 64(4), 949-964.
- Ergünay, O., (2007), **Türkiye'nin Afet Profili**, TMMOB Afet Sempozyumu, Ankara.
- Goetz, J. N., Brenning, A., Petschko, H., Leopold, P., (2015), **Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling**, Computers & Geosciences, 81, 1-11.
- Gómez, H., Kavzoglu, T., (2005), **Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela**, Engineering Geology, 78(1-2), 11-27.
- Gökceoğlu, C., Ercanoğlu, M., (2001), **Heyelan duyarlılık haritalarının hazırlanmasında kullanılan parametrelere ilişkin belirsizlikler**, Yerbilimleri, 23, 189-206.
- Hong, H. Y., Pradhan, B., Xu, C., Tien Bui, D., (2015), **Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines**, Catena, 133, 266-281.
- Hughes, G. F., (1968) **On the mean accuracy of statistical pattern recognizers**, IEEE Transactions on Information Theory, 14(1), 55-63.

- Kano, Y., Harada, A., (2000), **Stepwise variable selection in factor analysis**, Psychometrika, 65(1), 7-22.
- Kavzoglu, T., Sahin, E. K., Colkesen, I., (2015), **An assessment of multivariate and bivariate approaches in landslide susceptibility mapping: a case study of Duzkoy district**, Natural Hazards, 76(1), 471-496.
- Kavzoglu, T., Sahin, E. K., Colkesen, I., (2015), **Selecting optimal conditioning factors in shallow translational landslide susceptibility mapping using genetic algorithm**, Engineering Geology, 192, 101-112.
- Kutlug Sahin, E., Ipbuker, C., Kavzoglu, T., (2017), **Investigation of automatic feature weighting methods (Fisher, Chi-square and Relief-F) for landslide susceptibility mapping**, Geocarto International, 32(9), 956-977.
- Lee, S., (2005), **Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data journals**, International Journal of Remote Sensing, 26(7), 1477-1491.
- Li, L., Liu, R., Pirasteh, S., Chen, X. H., He, L., Li, J., (2017), **A novel genetic algorithm for optimization of conditioning factors in shallow translational landslides and susceptibility mapping**, Arabian Journal of Geosciences, 10(9), doi:ARTN 209 10.1007/s12517-017-3002-4.
- Nefeslioglu, H. A., Duman, T. Y., Durmaz, S., (2008), **Landslide susceptibility mapping for a part of tectonic Kelkit Valley (Eastern Black Sea region of Turkey)**, Geomorphology, 94(3-4), 401-418.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., Wasserman, W., (1996), **Applied linear regression models**, Richard D. Irwin, Inc., Burr Ridge, Illinois.
- Pham, B. T., Bui, D. T., Dholakia, M. B., Prakash, I., Pham, H. V., (2016), **A Comparative Study of Least Square Support Vector Machines and Multiclass Alternating Decision Trees for Spatial Prediction of Rainfall-Induced Landslides in a Tropical Cyclones Area**, Geotechnical and Geological Engineering, 34(6), 1807-1824.
- Pourghasemi, H. R., Kerle, N., (2016), **Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran**, Environmental Earth Sciences 75(3), 185.
- Pradhan, B., (2010), **Application of an advanced fuzzy logic model for landslide susceptibility analysis**, International Journal of Computational Intelligence Systems, 3(3), 370-381.
- Reif, D. M., Motsinger, A. A., McKinney, B. A., Crowe, J. E., Moore, J. H. (2006), **Feature selection using a random forests classifier for the integrated analysis of multiple data types**, Proceedings of the 2006 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, 171-178.
- Romer, C., Ferentinou, M., (2016), **Shallow landslide susceptibility assessment in a semiarid environment - A Quaternary catchment of KwaZulu-Natal, South Africa**, Engineering Geology, 201, 29-44.
- Sakellariou, M. G., Ferentinou, M. D., (2001), **GIS-Based Estimation of Slope Stability**, Natural Hazards Review, 2(1), 12-21.
- Saltelli, A., Chan, K., Scott, E. M., (2000), **Sensitivity analysis**, UK, Wiley&Sons Ltd.
- Swets, J. A., (1988), **Measuring the Accuracy of Diagnostic Systems**, Science, 240(4857), 1285-1293.
- Trigila, A., Iadanza, C., Esposito, C., Scarascia-Mugnozza, G., (2015), **Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy)**, Geomorphology, 249, 119-136.
- Tsangaratos, P., Ilija, I., Hong, H. Y., Chen, W., Xu, C., (2017), **Applying Information Theory and GIS-based quantitative methods to produce landslide susceptibility maps in Nancheng County, China**, Landslides 14(3), 1091-1111.
- Vasu, N. N., Lee, S. R., (2016), **A hybrid feature selection algorithm integrating an extreme learning machine for landslide susceptibility modeling of Mt. Woomyeon, South Korea**, Geomorphology, 263, 50-70.

- Wagner, J. M., Shimshak, D. G., (2007), **Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives**, European Journal of Operational Research, 180(1), 57-67.
- Wulf, A.J., (2014), **An Optional European Contract Code in the Institutional Competition between European Contract Laws, Institutional Competition between Optional Codes in European Contract Law: A Theoretical and Empirical Analysis**. Springer Fachmedien Wiesbaden, Wiesbaden, 198.
- 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(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(4), 821-836.
- Yilmaz, I., (2009), **A case study from Koyulhisar (Sivas-Turkey) for landslide susceptibility mapping by artificial neural networks**, Bulletin of Engineering Geology and the Environment, 68(3), 297-306.
- Youssef, A. M., Pourghasemi, H. R., Pourtaghi, Z. S., Al-Katheeri, M. M., (2016), **Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia**, Landslides 13(5), 839-856.
- Zhang, G. F., Cai, Y. X., Zheng, Z., Zhen, J. W., Liu, Y. L., Huang, K. Y., (2016), **Integration of the Statistical Index Method and the Analytic Hierarchy Process technique for the assessment of landslide susceptibility in Huizhou, China**, Catena 142, 233-244.
- Zhang, Z. H., (2016), **Variable selection with stepwise and best subset approaches**, Annals of Translational Medicine 4(7), 136