



ZEMİNLERİN KIVAM VE KOMPAKSİYON ÖZELLİKLERİNİN TAHMİNİNDE RASTGELE ORMAN REGRESYONU YÖNTEMİNİN UYGULANABİLİRLİĞİ

Said Enes NURAY¹, Hazal Berrak GENÇDAL^{2*}, Zülal AKBAY ARAMA¹

¹ İstanbul Üniversitesi – Cerrahpaşa, Mühendislik Fakültesi, İnşaat Mühendisliği Bölümü, İstanbul, Türkiye

² İstanbul Kültür Üniversitesi, Mühendislik Fakültesi, İnşaat Mühendisliği Bölümü, İstanbul, Türkiye

Anahtar Kelimeler

Likit Limit,
Kompaksiyon,
Kil Zeminler,
Rastgele Orman Regresyonu,
Karar Ağaçları,
Regresyon Yöntemi.

Öz

Bu makalede, yüksek plastisiteli kil zeminlerin kıvam limitleri ve kompaksiyon karakteristikleri arasındaki ilişki basit regresyon ve karar ağaçları tabanlı Rastgele Orman regresyon (RO) yöntemlerinin karşılaştırmalı olarak analiz edilmesi yoluyla irdelenmiştir. Zeminlerin kompaksiyon parametrelerini oluşturan maksimum kuru birim hacim ağırlık ve optimum su muhtevası değerlerinin doğrudan belirlenmesinde kullanılan standart laboratuvar deneylerin zorluğu ve uzun numune hazırlama-bekleme süreçleri içermesi nedeni ile göreceli olarak daha pratik deneyler kullanılarak bu parametrelerin tahmin edilmesi günümüzde sıklıkla uygulanan bir yöntemdir. Ayrıca, kıvam limiti deneylerinden likit limit, tüm geoteknik mühendisliği tasarımlarında uygulanan ve tatminkar sonuçlar veren bir deneydir. Bu çalışmada, yüksek plastisiteli kil zeminlere ait literatürde sunulan 387 adet kıvam limiti ve 59 kompaksiyon-kıvam limiti test çiftinin kullanılması ile oluşturulan bir veri tabanı kullanılarak iki aşamalı bir tahmin süreci yürütülmüştür. Birinci aşamada plastisite indisinin doğrudan likit limit değerinden, ikinci aşamada ise kompaksiyon parametrelerinin plastisite indisinden tahmin olasılığı araştırılmıştır. Aynı zamanda, laboratuvar deneylerinden elde edilen gerçek verilerin tutarsızlık durumları ve bu verilerin belirli bir eğilim izlememesi sebebi ile genel regresyon çalışmalarında oluşan doğruluk oranı düşüklüğüne dikkat çekilerek, bu doğruluk oranlarının Rastgele Orman regresyonu yöntemi ile nasıl yükseltilebileceği de incelenmektedir. Sonuçlarda, Rastgele Orman regresyonu yönteminin yüksek plastisiteli kil zeminlerin kıvam ve kompaksiyon özelliklerinin tahmininde başarılı olduğu ve kullanılabilir nitelikte sonuçlar sunduğu gösterilmektedir.

THE APPLICABILITY OF RANDOM FOREST REGRESSION METHOD FOR THE PREDICTION OF THE CONSISTENCY AND COMPACTION PROPERTIES OF SOILS

Keywords

Liquid Limit,
Compaction,
Clay Soils,
Random Forest Regression,
Decision Trees,
Regression method.

Abstract

Within this paper, the relationship between consistency limits and compaction characteristics of highly plastic clay soils was examined by comparative analysis of regression and Random Forest regression methods. Due to the difficulty of standard laboratory experiments that have long sample preparation-waiting processes, which are used to directly obtain the maximum dry unit weight and optimum water content values representing the compaction parameters of soils, it is relatively more applicable to estimate these parameters by the use of practical experiments is a method that is frequently applied today. In addition, the liquid limit, one of the consistency limit tests, is an experiment that is applied in all geotechnical engineering designs and gives satisfactory results. In this study, a two-stage estimation process was carried out by using a database created by using 387 consistency limit and 59 compaction-consistency limit test couples presented in the literature studies of high plasticity clay soils. In the first stage, the estimation of the plasticity index directly from the liquid limit, in the second stage, the probability of estimating the compaction parameters from the plasticity index was investigated.

* İlgili yazar / Corresponding author: h.gencdal@iku.edu.tr, +90-553-347-24-96

At the same time, this study is focused on the inconsistency of the real data obtained directly from the laboratory experiments and the low accuracy rate that occurs in the general regression studies due to the fact that these data do not follow a certain trend. It is also examined how these accuracy rates can be increased by the Random Forest regression method. Consequently, it is shown that the Random Forest regression method can be used for the estimation of the consistency and compaction properties of highly plastic clayey soils, and gives satisfactory results to use.

Alıntı / Cite

Nuray, S.E., Gençdal, H.B., Akbay Arama, Z., (2021). Zeminlerin Kıvam Ve Kompaksiyon Özelliklerinin Tahmininde Rastgele Orman Regresyonu Yönteminin Uygulanabilirliği, Mühendislik Bilimleri ve Tasarım Dergisi, 9(1), 265-281.

Yazar Kimliği / Author ID (ORCID Number)	Makale Süreci / Article Process	
S.E. Nuray, 0000-0002-2333-2687	Başvuru Tarihi / Submission Date	02.10.2020
H.B. Gençdal, 0000-0002-1041-5643	Revizyon Tarihi / Revision Date	10.12.2020
Z. Akbay Arama, 0000-0001-8185-7329	Kabul Tarihi / Accepted Date	07.01.2021
	Yayın Tarihi / Published Date	30.03.2021

1. Giriş (Introduction)

Geoteknik mühendisliği, doğası zemin olan birçok farklı ve karmaşık malzemeyi inceleyen ve bu malzemelerin davranış mekanizmalarını ve diğer yapısal sistemler ile etkileşimli davranışlarını tasarlayan çok disiplinli bir mühendislik bakış açısını içermektedir. Bu bağlamda, zeminlerin her noktada değişebilen özelliklerinin belirlenmesi ve tasarımlarda kullanılmak üzere ortak tanımlayıcı parametrelerin elde edilmesi geoteknik mühendisliğini çalışmalarının ilk adımını oluşturmaktadır. Arazi ve laboratuvar deneylerinin uygulanması ile belirlenebilen bu tasarım parametrelerinin, elde edilme hızının artırılması ve proje giderlerinin azaltılması amacı ile alternatif tahmin süreçleri günümüz mühendislik tasarımlarında odaklanılan konular arasında yerini almaktadır. Ancak tahmin edilen parametrelerin tasarım için doğrudan kullanımının gerçekliği kanıtlanabilir nitelikte olmalıdır. Bu bağlamda, gözetimli makine öğrenmesi metotlarından olan ve kolektif bir öğrenme algoritmasına sahip Rastgele Orman (RO) regresyonu sıklıkla kullanılan yöntemlerdendir (Breiman, 2001). Kolektif öğrenme sistemi, metodun bünyesinde barındırdığı karar ağaçları ile çalışması olarak açıklanabilmektedir (Akman vd., 2011). Karmaşık, yüksek boyutlu veriler, karma öznelikli ve sayısal değişkenler, Gauss dışı istatistiksel dağılımlar, doğrusal olmayan ilişkiler gibi sorunlar karşısında, RO regresyonunun arka planında bulundurduğu her bir karar ağacının verdiği tahmin sonuçlarının ortalamasını alarak her veri için bir sayısal tahmin değeri vermesi nedeni ile daha tutarlı olduğu ve bu nedenle yoğun ilgi gördüğü literatür kaynaklarında açıklanmaktadır (Liaw ve Wiener, 2002; Gunther vd., 2003; Segal, 2003; Pal, 2005).

2. Kaynak Araştırması (Literature Survey)

Son yıllarda uydu görüntüleri üzerinden en doğru bilgiye hızlı bir şekilde ulaşmak için RO regresyon yöntemi, Karar Ağaçları (KA), Yapay Sinir Ağları, Destek Vektör Makineleri, K En Yakın Komşuluk gibi regresyon yöntemleri olarak anılan, farklı öğrenme tabanlı algoritmalar geliştirilmektedir. Bu algoritmalar, aynı zamanda makine öğrenmesi yöntemleri olarak da adlandırılabilir. (Özdarıcı Ok vd., 2011). İstatistiksel yöntemlerin aksine, verinin dağılımı ile ilgili herhangi bir varsayıma dayanmadığı için parametrik olmayan makine öğrenme yöntemleri, veri temelli yöntemlerdir. Bu yöntemler, tahmin edici ve buna karşılık gelen cevaplar arasındaki ilişkileri öğrenirler (Breiman, 2001). Makine öğrenme yöntemleri, yeteri büyüklükteki veri setlerinden ve parametrelerden faydalanarak, veri için en uygun modeli girdi veriden elde edilen karar kurallarını kullanarak bulmayı amaçlarlar (Akar ve Güngör, 2012). Öğrenme tabanlı bir makine öğrenmesi aracı olan RO regresyon yöntemi, yüksek doğruluk ve hızlı bir çözüm sağlaması nedeniyle çok farklı alanlarda çalışan araştırmacıların dikkatini çekmiştir (Pal, 2003; Gislason vd., 2004; Waske vd., 2007; Akar vd., 2010). Breiman (2001), RO regresyonun çoğu durumda doğru sonuçlar verdiğini; Pal (2003) ve Gislason vd. (2004), yaptıkları çalışmalarda RO, karar ağaçları yöntemlerini karşılaştırarak, en iyi sonucu RO yönteminin verdiğini; Breiman ve Cutler (2005), RO'nun, şuan ki algoritmalar arasında doğruluğu eşsiz olan bir regresyon yöntemi olduğunu; Gislason vd., (2006), yapmış oldukları araştırmalarda, RO regresyonunun, yeterliliği ve doğruluğu ile çok kullanışlı bir regresyon yöntemi olduğunu, Waske vd. (2007), RO, En Çok Benzerlik (EÇB), Karar Ağaçları ve Hızlandırma yöntemlerini karşılaştırmış, en iyi sonucu RO yöntemi ile elde edildiğini belirtmişlerdir. Akar vd. (2010), yapmış oldukları çalışmalarında RO'nun hızlı olduğunu ve yüksek doğruluk gösterdiğini vurgulamışlardır. Bahsi geçen çalışmaların hiçbiri geoteknik mühendisliği uygulamaları veya zemin bilimi kapsamında değildir ancak RO yönteminin avantajlarını ve kullanılabilir niteliklerini gözler önüne sermektedir. Zeminlerin özelliklerini, RO kullanımı ile tahmin etmeyi amaçlayan çalışmalar son derece az sayıda ve sınırlı çalışma alanındadır. Bu bağlamda, Grimm vd. (2008), Viscara Rossel ve Behrens (2010) çalışmalarında, toprak organik karbon konsantrasyonları, kil içeriği ve pH gibi zemin özelliklerini, RO ile tahmin etmişlerdir. Singh vd. (2017), RO regresyon yöntemini kullanarak

zeminin sızma hızını incelemiş, sonuçları farklı regresyon analiz yöntemleri ile karşılaştırmış ve RO ile elde edilen doğruluk payının diğer analiz sonuçlarına göre %25 daha fazla olduğunu tespit etmişlerdir. Zhang vd. (2020), ikincil konsolidasyon (Ca) indeksi tahmini için parçacık sürü optimizasyonu ve RO metoduna dayalı yeni bir hibrit akıllı model geliştirmek üzerine çalışmalar yapmışlardır. Pham vd. (2020), RO ve Parçacık Sürü Optimizasyon metodlarını kullanarak zeminin drenajsız kayma dayanımı tahmini için yeni bir hibrit yumuşak hesaplama modeli üzerinde çalışmışlardır. 127 zemin örneğinin deneysel sonuçlarını modelleri eğitmek ve doğrulamak için veri kümeleri oluşturmak için kullanmış, yeni oluşturdukları hibrit yumuşak hesaplama modelini ile zeminin kayma dayanımı tahmininde çok yüksek bir doğruluk performansı gösterdiğini elde etmişlerdir. Şahin (2018), RO ve birkaç farklı metodun heyelanda etkili faktörler içerisinde optimum faktör modelinin tespiti ve heyelan duyarlılık haritası üretimindeki etkinliğini araştırmış ve en yüksek doğruluğu RO'nun tahmin ettiğini belirtmiştir. Shukla vd. (2018), farklı zemin tipleri üzerinde yapmış oldukları çalışmada, zeminlerin sınıflandırılmasında zeminlerin mekansal dağılımını haritalamak, zemin kullanımı ve yönetim kararları hakkında bilgilendirme yapmak için, RO modelini etkin bir şekilde kullanmışlardır. Hint bölgesinin zemin sınıflandırması için RO modeli ile davranış ve performans değerlendirilmesi yapmışlardır. Ließ vd. (2011), 56 farklı zemin tipinin mekânsal dağılımını arazi parametrelerinden tahmin etmek için istatistiksel modeller uyarlayarak, analizler gerçekleştirmişler. Doğrusal regresyon analizleri yanında RO modellerinin de performansını incelemişlerdir. Dharumarajan vd. (2017), 116 zemin örneğinin saha araştırması verilerine dayanarak, zemin özelliklerinin mekânsal değişimini RO yöntemi kullanarak haritalandırmıştır. Kempinen vd. (2017), RO regresyon yönteminin de içinde olduğu farklı yöntemler ile toprağın nemini ve bunun zamansal değişiminin incelemek amacıyla çalışmalar yapmışlardır. Rastgou vd. (2020), çoklu doğrusal olmayan regresyon yaklaşımı ve RO metodu ile toprak suyu tutma eğrisinin tahmin edilmesi üzerine çalışmalar yapmışlardır. Ouedraogo vd. (2018), yeraltı suyu nitrat kirlenmesini modellemek için 250 adet veri kullanarak RO regresyonu analizleri yapmış, sonuçları farklı regresyon yöntemleri ile karşılaştırmış, hangi açıklayıcı değişkenlerin yeraltı sularında nitrat kirliliğinin oluşumunu etkilediğini belirlemiştir. RO metodunun, geleneksel bir doğrusal regresyon modelinden çok daha yüksek bir tahmin gücüne sahip olduğunu belirtmiştir.

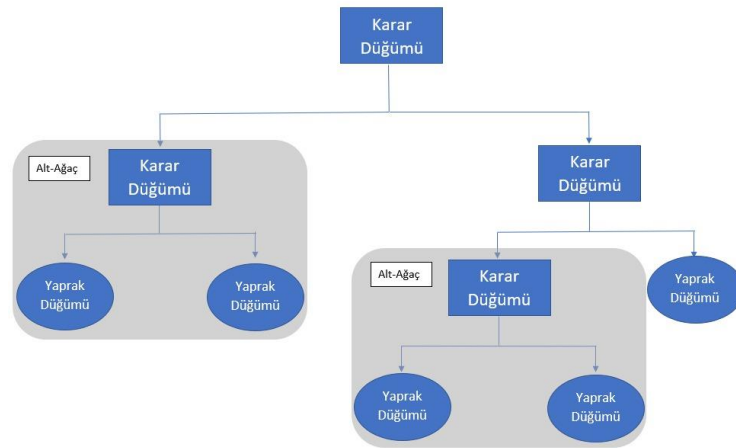
Tüm bu çalışmalar göz önüne alındığında; RO algoritmasının çözüm aracı olarak kullanıldığı tüm problemlerde; hızlı olması ve yüksek doğruluk sağlaması, regresyon analizlerinde çok iyi performans sergilemesi, öğrenmede büyük avantaj sağlaması özellikleri özellikle son yıllarda sıklıkla tercih edilmesini sağlamaktadır. Hali hazırda güncel literatürde RO'nun regresyon performansını denemeye ve diğer regresyon araçlarıyla karşılaştırmaya yönelik uygulamalar yapılmaya da devam edilmektedir. Ancak, RO'nun geoteknik mühendisliği disiplini içeriğinde olan zemin parametre tahmin süreçlerinde kullanımı yok denecek kadar azdır. Bu nedenle, bu makale kapsamında öncelikli olarak Rastgele Orman regresyonu ve arka planında çalışan Karar Ağaçları algoritmaları detaylı şekilde açıklanmaya çalışılacaktır. Akabinde, özellikle kil zemin içeren alanlarda inşa edilecek yapıların tasarımında uygulanan standart laboratuvar deneyi olan likit limit deneyi sonuçlarından faydalanarak plastisite indisi değerinin tahmini ve plastisite indisi değerinden faydalanılarak zemin kompaksiyon karakteristiklerinin RO analizleri ile elde edilmesi amaçlanmıştır. Bu amaçla, literatürde yer alan kaynaklarda yüksek plastisiteli kil zeminler için uygulanmış olan kıvam limiti deneyleri sayısallaştırılarak bir veri tabanı oluşturulmuştur. Literatürden alınan bu verilerin ilişkilendirmesinde belirli bir eğilim izlemediği ve genel olarak kullanılan regresyon analizleri ile çok düşük doğruluk veren 59 likit deney sonucunu temsil eden ilk veri seti ile, yine zeminden alınan verilerin belirli bir eğilim izlediği ve genel olarak kullanılan regresyon analizleri ile yüksek doğruluk veren 387 likit deney sonucunu temsil eden ikinci veri seti, RO regresyon yöntemi ile incelenmiş ve sonuçlar genel kullanımdaki regresyon analizleri ile karşılaştırılmıştır. Uygulanan bu çalışma, literatürdeki bilinen ve sık kullanılan veri analizi yöntemleri ile elde edilen düşük determinasyon katsayısı değerlerini (R^2) ve diğer ölçüm metriklerini de arttırmayı hedeflemiştir. Bu durum özellikle belirli bir eğilim göstermeyen veri setlerinde (lineer, polinomal, üssel vb.) oluşmaktadır ve sık kullanılan doğrusal regresyon, polinomal regresyon, üssel regresyon gibi analizler düşük determinasyon katsayısı vermektedirler. Rastgele Orman regresyonu ile dağınık veri setlerinde dahi yüksek doğruluk elde etmek mümkün olabildiğinden geoteknik alanındaki veri analizleri işlemlerine fayda sağlayabileceği öngörülmektedir. Sonuç olarak, çalışmaya uygulanabilir bir nitelik kazandırmak amacı ile RO regresyonu yöntemi ile elde edilen ilişkiler, uluslararası metrik değerlendirme parametreleri ile ifade edilerek yüksek plastisiteli kil zeminler için parametre tahmininde kullanılması amacı ile sunulmuştur.

3. Materyal ve Yöntem (Material and Method)

Bu bölümde, zeminlerin geoteknik parametre tahmin sürecinde kullanılan RO regresyonu ve RO regresyonunun arka planında çalışan karar ağaçları yöntemleri tanıtılarak analizlerde kullanılacak olan verilerin tasnif süreci özetlenmiştir. Bu amaçla, Python Programlama dilinin makine öğrenmesi kütüphanesi olan Scikit-learn kütüphanesi kullanılmıştır. Ayrıca, zeminlerin geoteknik özelliklerinin belirlenmesi amacı ile literatürden derlenen veri setinde uygulanan likit limit ve plastik limit deneyleri ASTM D-4318'e göre, zeminlerin sıkışma karakteristiklerini temsil eden standart kompaksiyon deneyleri ASTM D-698'e göre uygulanmıştır.

3.1. Karar Ağaçları (KA) (Decision Trees)

Karar ağaçları (KA), literatürde birçok sınıflandırma ve regresyon probleminin çözümünde sıklıkla kullanılan bir veri madenciliği tekniğidir (Pal ve Mather, 2003; Kavzoğlu vd., 2010; Boulila vd., 2011). Sınıflandırma ve regresyon problemlerinin çözümünde tercih edilen KA, çok aşamalı ve ardışık bir yaklaşım kullanarak karmaşık yapıdaki verileri aşamalı bir hale dönüştürür ve basit bir şekilde karar vermeyi sağlamaktadır (Safavian vd., 1991). Bir problemi oluşturan veri setlerinin yapısına göre bir ağaç yapısı şeklinde sınıflandırma ve regresyon modelleri oluşturan karar ağaçları; kategorik veya belirli sınıflara ait verilerin sınıflandırılmasında kullanıldıklarında sınıflandırma ağacı, sürekli verilerin bulunduğu regresyon problemlerinin çözümünde kullanıldıklarında regresyon ağaçları olarak adlandırılmaktadırlar (Hill ve Lewicki, 2006). Karar ağacı yapılarının oluşturulmasında, verilere ilişkin birtakım sorular sorularak karar kurallarının oluşturulması için, ağaç yapısının temel elemanı olan kök düğümünde sorular sorulmaya başlanır ve ağaç yapısının son elemanı olan yapraklara ulaşıncaya kadar ağacın büyümesi veya dallanması devam eder (Pal vd., 2003). Karar ağaç yapılarının oluşturulmasında kullanılan karar kurallarının anlaşılabilir olması yöntemin kullanımını yaygın hale getirmiştir (Kavzoğlu vd., 2010). Bir karar ağacı, kararın içeriğine bağlı olarak olası seçeneklerin her birine bir olasılık atayan karar verme ağacı olarak tanımlanabilmektedir (Magerman, 1995). Karar ağacı, her dal düğümünün çeşitli alternatifler arasında bir seçimi temsil ettiği ve her yaprak düğümünün bir kararı temsil ettiği ağaçtır. Karar ağacı genellikle karar alma hedefiyle, bilgiye ulaşmak için kullanılmaktadır ve kullanıcıların işlem yapmasını gerektiren bir kök düğümü ile başlamaktadır. Bu düğümden sonra kullanıcılar, her bir düğümü karar ağacı öğrenme algoritmasına göre bölmekte ve en sonunda ulaşılan sonuç, her dalın olası bir karar senaryosunu ve sonucunu temsil ettiği bir karar ağacı olmaktadır (Peng vd., 2009) (Şekil 1).



Şekil 1. Karar ağacı şeması (Template of decision tree) [Fidan, H.]

3.2. Rastgele Orman Regresyon Yöntemi (RO) (Random Forest Regression Method RF)

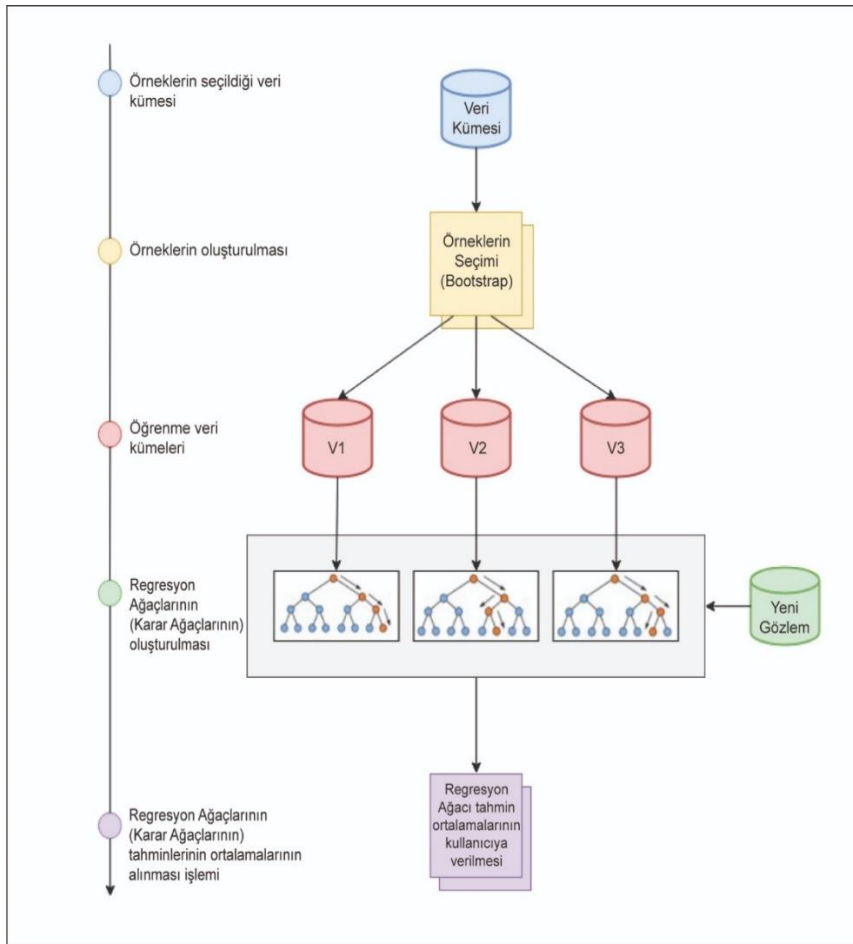
Bu araştırmada, zeminlerin geoteknik parametrelerinin belirlenmesinde, Rastgele Orman regresyonu olarak bilinen karar ağacı tabanlı bir regresyon aracı ele alınmıştır. Breiman (2001) tarafından karar ağaçlarının bir kombinasyonu olarak önerilen RO, rastgelelik özelliği eklenerek torbalama yönteminin geliştirilmiş bir versiyonu olarak kullanılmaktadır. RO, tüm değişkenler arasından en iyi dalı kullanarak her bir düğümü dallara ayırmak yerine, her bir düğümde rastgele olarak seçilen değişkenler arasından en iyisini kullanarak her bir düğümü dallara ayırır. Her bir veri seti orijinal veri setinden yer değiştirmeli olarak üretilir (Akar ve Güngör, 2012) ve rastgele özellik kullanılarak ağaçlar geliştirilir (Archer, 2008; Breiman, 2001). Hızlı, istenilen sayı kadar regresyon ağacı geliştiren ve çalıştıran RO, bir veriyi tahmin etmek için, girdi verisini ormandaki her ağaca yerleştirir. Her bir veri seti orijinal veri setinden yeniden örnekleme kullanılarak üretilir. Sonuç olarak RO regresyonu, birden fazla karar ağacını oluşturmaktadır ve daha doğru ve istikrarlı bir tahmin elde etmek için onları birleştirmektedir. Bu nedenle, RO regresyonunun doğru olarak uygulanabilmesi, Karar Ağaçları algoritmasının mantığının çözülmesi ile mümkün olabilmektedir.

RO regresyonu için kullanıcı tarafından iki parametre tanımlanması gerekmektedir. Bunlar RO regresyonu altyapısında geliştirilecek olan Karar Ağacı sayısı (N) ve her bir düğümde kullanılacak değişken sayılarıdır (m). ID3 algoritması (Scikit-learn kütüphanesi Oython içerisinde analiz yapan araç) ile RO regresyonu için Karar Ağaçları üretilir (Breiman, 2001) ve ID3 algoritması kurallarına göre entropi ile belirsizlik ölçülür ve Eşitlik 1 ile ifade edilir.

$$H(T) = -\sum_{i=1}^n p(x_i) \log p(x_i) \quad (1)$$

Burada H entropiyi, $p(x_i)$ belirli bir veri sınıfına ait grubun yüzdesini, T ise seçilen eğitim kümesi elemanlarını belirtmektedir. Entropi ne kadar yüksekse, belirsizlik o kadar yüksektir. Entropi, "0" olana kadar ağaçların dallanması devam eder veya ağaç derinliği belirli bir düzeyde seçilerek ağaç budaması yapılabilir. Rastgele Orman algoritmasının uygulanma aşamaları aşağıdaki gibidir (Akman vd. 2011), (Şekil 2):

1. Orijinal veri setinden N adet örneklem seçilir. Seçilen örneklemin belli bir kısmı oluşturulacak ağaçlar için öğrenme verisi olarak kullanılır.
2. Budanmamış regresyon ağacı aşağıdaki adımları takip ederek oluşturulur.
 - a. Öğrenme verisi setinden her düğümde bütün bağımsız değişkenler içerisinde en iyi bağımsız değişkeni seçmek yerine rastgele tane bağımsız değişken seçilir ve bunların içerisinde en iyi dallara ayıracak (bilgi kazancını en iyi şekilde sağlayacak) olanı belirlenir.
 - b. Belirlenen tahmin değişkeni için en iyi dallanma kriteri entropi veya Gini indeksi ile hesaplanır ve hesaplanan değere göre veri setini her düğümde iki alt dala ayrılır.
 - c. İlk iki adımda verilen işlemleri aşağıya doğru yaparak düğüm elde edilinceye kadar her bir düğümde tekrar edilir.



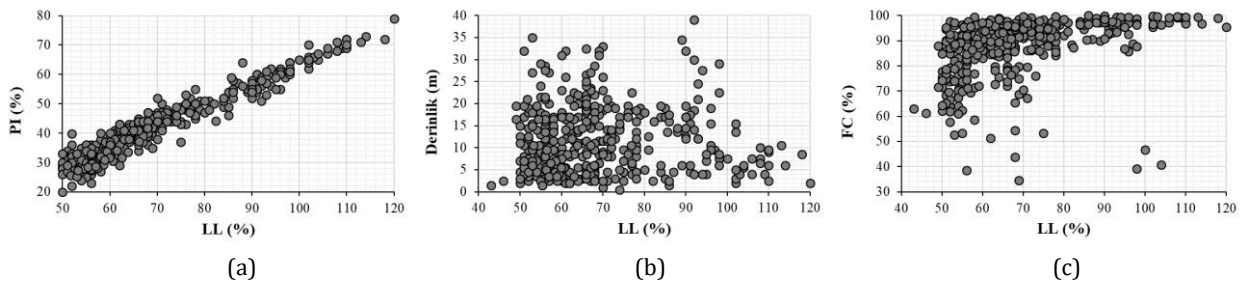
Şekil 2. RO regresyon algoritmasının çalışma prensibi (The working process of Random Forest Regression Algorithm) (Fidan, H.)

3. N tane ağacın ayrı ayrı yapmış olduğu tahminleri bir araya getirerek yeni bir tahminde bulunur. Regresyon ağaçları (Karar Ağaçları) için en son adımda çıkan ortalama tahmin final tahmini olarak seçilir, yani her gözlem için ayrı ayrı bulunan tahmin değerleri toplanır ve sayısını kullanıcının belirlediği Karar Ağacı sayısına bölünür.

Algoritmanın herhangi bir denklem veya iki boyuttan fazla düşünülmesinde bir düzlem oluşturamaması bu metodun en önemli eksikliğidir. Herhangi bir denklem üretilmediği için bilgisayar haricinde kullanımı mümkün olamazken, bu regresyon herhangi bir bilgisayar programına entegre edildiğinde başarılı sonuçlar üretebilir. Regresyonun, sadece veri girişinin olduğu aralıklarda çalışabilmesi, yani veri setindeki en küçük girdinin "x", en büyük girdinin "y" olduğu bir durumda, regresyonun "x" den küçük ve "y" den büyük girdileri tahmin edememesi de metodun diğer kısıtlayıcı tarafıdır. Ayrıca bu regresyonun algoritması, verilerin tamamının kullanılmamasına dikkat ederek, veriden rastgele kesitler alır. Rastgele alınan kesitler nedeniyle program, her çalıştırıldığında, birbirine çok yakın ama farklı doğruluklar (R^2) verir. Program, aldığı rastgele kesitler üzerinde karar ağaçları oluşturur ve regresyona bir veri girildiğinde karşılığı olan bağımlı veriyi vermek için bu karar ağaçlarının her birinde girdiyi yerine koyarak oluşturduğu karar ağaçlarındaki sonuçların ortalamasını alır. Bu regresyon türünde kullanılacak olan karar ağacı sayısı genel olarak veri sayısının karekökü ile yarısı arasındaki sayılar kadar seçilip, analizler gerçekleştirilebilir. Analizin gerçekleştirileceği ağaç sayısı, doğruluk oranı (R^2)'nin maksimum olduğu karar ağacı sayısı olarak seçilir.

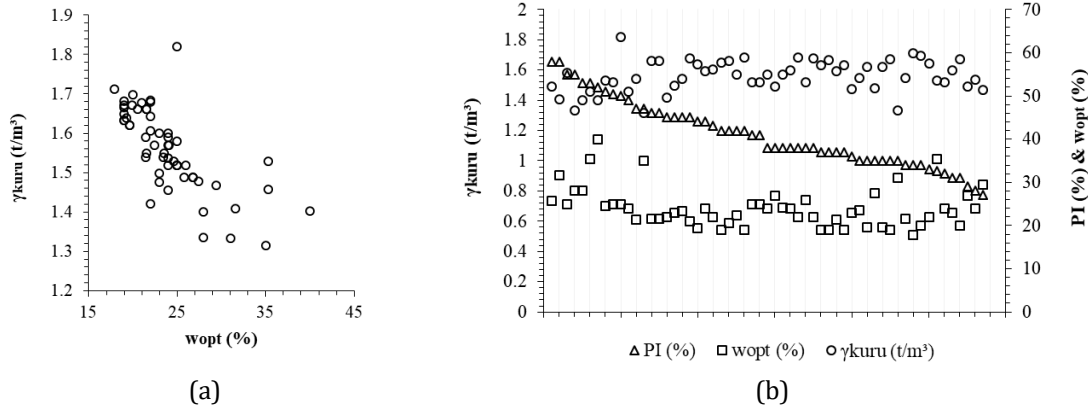
3.3. Verilerin Sınıflandırılması (Classification of Datas)

Kıvam limitleri, zeminlerin geoteknik özellikleri açısından en kolay elde edilebilir parametrelerdendir. Likit, plastik ve rötre limitleri şeklinde ayrıştırılarak, zeminlerin fazları arasındaki sınırları temsil eden kıvam limitleri; özellikle ince daneli zeminlerin sınıflandırılmasında doğrudan kullanılan önemli parametreler olup dolaylı yollardan da zeminlerin mukavemet ve rijitlik parametrelerinin tahmininde kullanılmaktadırlar. Kıvam limitlerinin ilgili olduğu uygulamaların başında, her hafriyat ve dolgu alanında uygulanan kompaksiyon işlemleri de gelmektedir. Kompaksiyon belirli bir sıkıştırma enerjisi uygulanarak zemin içerisinde bulunan hava hacmini azaltmak ve bu yolla zeminin sıklık derecesini ve yoğunluğunu arttırmak amacı ile uygulanan mekanik işlemlere verilen addır. En köklü zemin iyileştirme işlemleri arasında başı çeken kompaksiyon uygulamalarının denetim kriteri, sıkıştırılan zeminin optimum su muhtevası ve maksimum kuru birim hacim ağırlığının elde edilmesidir. Bu sıkıştırma özelliklerinin tespiti, karmaşık ve zaman alan bir süreci içerir. Bu nedenle, kıvam limiti deneyleri gibi kolay uygulanan deneylerden faydalanarak sıkıştırma özelliklerini tahmin etme mantığının kullanılması sıklıkla tercih edilmektedir. Literatürde etkin kullanımı gözlemlenen yöntemlerden regresyon analizleri; Günaydın (2009), Thompson ve White (2008), Matteo vd. (2009), Canillas ve Saloke (2001), Omar vd. (2003), Khuntia vd. (2014) ve Viji vd. (2013), Akbay Arama vd. (2020) tarafından, yapay sinir ağları Günaydın (2009), Viji vd. (2013), Sinha ve Wang (2008) ve Ardakani ve Kordnaej (2019) tarafından geoteknik parametre tahminlerinde kullanılmaktadır. Bu çalışma kapsamında, sadece yüksek derecede plastik inorganik killer (CH) için yapılan kıvam limitleri ve kompaksiyon deneyleri dikkate alınarak, literatürde kabul gören kaynaklardan 357 adet kıvam limiti deneyleri ve 59 adet standart Proctor deneyi verisi toplanarak geniş bir veri tabanı oluşturulmuştur. (Dewoolkar ve Huzjak, 2005; Mehta ve Sachan, 2017; Benson vd., 1994; Benson vd., 1999, Singhal vd., 2005). Çalışma, plastisite indeksi değeri ve sıkıştırma özelliklerinin elde edilmesi amacıyla iki farklı aşamaya ayrılmıştır. Çalışmanın ilk aşamasında, likit limit deneyleri değerleri ile basit regresyon ve RO regresyon analizleri kullanılarak doğrudan plastisite indeksi değerine ulaşılması ve gerçek laboratuvar deneylerinin sonuçları ile kontrol edilmesi amaçlanmıştır. Çalışmanın ikinci aşamasında ise tekrar basit regresyon ve RO regresyon uygulamaları kullanılarak yüksek plastik killer için belirlenen optimum su içeriği ve maksimum kuru birim ağırlık değerlerinin elde edilmesinin mümkün olup olmadığının kontrol edilmesi amaçlanmıştır. Hem kıvam hem de sıkışma davranışını en uygun belirleme yoluna ulaşmak için söz konusu çözüm teknikleri arasında karşılaştırmalar yapılmıştır. Aynı zamanda, zeminlerin kıvam özelliklerinin yapılan regresyon analizleri sonucunda belirli bir davranış eğilimi takip ettiği ve bu sebeple yüksek doğruluk veren verilerin 1. veri setini oluşturduğu düşünülmüştür. Bu veri setinde değişkenler kendi aralarında birleştirilerek tek ve çok değişkenli regresyon ile RO regresyonu yöntemi kullanılarak analizler yapılmıştır. Bu veri seti, yüksek plastisiteli killi zeminler için Şekil 3a'da verildiği gibi likit limit ve plastisite indeksi grubunu ve aynı zamanda çok değişkenli analizler için Şekil 3b ve Şekil 3c'de sırası ile verildiği gibi derinlik ve ince dane oranını içermektedir.



Şekil 3. Birinci veri setinde; likit limit, plastik limit ve ince dane oranı değişimi (The change of liquid limit and fine content rate for the first data set)

İkinci veri setinde ise, zeminden alınan verilerin belirli bir davranış eğilimi izlemediği ve genel olarak kullanılan regresyon analizleri ile çok düşük doğruluk veren 59 adet deney sonucunu temsil etmektedir (Şekil 2). İkinci veri seti uygulamalarında da yine değişkenler kendi aralarında birleştirilmiş ve tek ve çok değişkenli regresyon ile RO regresyonu yöntemi kullanılarak analizler yapılmıştır. Bu veri seti, killi zeminler için likit limit ve plastik limit deney sonuçlarından elde edilen plastisite indeksi (PI) ile kompaksiyon deneylerinden elde edilen optimum su muhtevası (w_{opt}) ve kuru birim hacim ağırlık (γ_{kuru}) parametrelerini içermektedir. İkinci veri seti sırası ile Şekil 4'te likit limit, plastisite indeksi ve ince dane oranı değişimi şeklinde ve Şekil 4a'da optimum su muhtevası-maksimum kuru birim hacim ağırlığı değişimi ve Şekil 4b'de optimum su muhtevası-maksimum kuru birim hacim ağırlığı-plastisite indeksi değişimi ile sunulmuştur.



Şekil 4. İkinci veri setinde; maksimum kuru birim hacim ağırlık, optimum su muhtevası ve plastisite indeksi değişimi (The change of maximum dry unit weight, optimum water content and plasticity index for the second data set)

Söz konusu iki veri seti arasındaki temel fark, killi zeminler için uygulanan regresyon analizlerinde verdikleri doğruluk oranıdır. Birinci veri seti uygulanan doğrusal regresyon analizleri sonucunda yüksek doğruluk vermektedir. Bu ayrıştırmadaki asıl amaç, doğrusal regresyon ile yüksek doğruluk veren veri setlerinde de RO metodunun sonuçlarını gözlemleyebilmektir. Bu bağlamda uygulanan analizlerde, birinci veri seti için, regresyon ve RO analizlerinde ilk adımda likit limit ve plastisite indeksi arasındaki ilişkiyi elde etmek; ikinci adımda ise plastisite indeksi değerini hem deney numunesinin alındığı derinliği (h) hem de hesaplanan likit limit değeri açısından elde etmek amaçlanmıştır. İkinci veri seti için; regresyon ve RO regresyonu analizleri, ilk adım plastisite indeksi ile kompaksiyon parametresi olan optimum su muhtevası ilişkisini elde etmek için yapılmıştır. İkinci adım ise plastisite indeksi ile diğer bir kompaksiyon parametresi olan maksimum kuru birim hacim ağırlık ilişkisini elde etmek için yapılmıştır. İkinci veri seti analizlerinde ayrıca üçüncü ve dördüncü adım da öngörülerek, plastisite indeksi ile kompaksiyon parametreleri olan optimum su muhtevası ve maksimum kuru birim hacim ağırlık ilişkisi entegre bir şekilde de göz önüne alınmıştır.

4. Değerlendirme ve Sonuçlar (Evaluations And Results)

Zeminlerin kıvam ve kompaksiyon özelliklerinin tahmininde RO regresyonu yönteminin uygulanabilirliğinin denetlenmesi amacı ile ayrıştırılan veri setlerinde yapılan analizler sonucunda elde edilen çıktılar aşağıda maddeler halinde sunulmuştur.

4.1. Veri Seti Analiz 1 (The Analysis of Data Set 1)

Toplam 387 veriye sahip olan birinci veri seti (Şekil 3) analizleri dahilinde üç farklı analiz türü uygulanmaktadır. Bu analiz türleri basit doğrusal regresyon ve çoklu doğrusal regresyon ve RO regresyonu yöntemi şeklindedir. Bu bölümdeki çalışmaların temel amacı, rastgele bir eğilime uyum gösteren verilerde Rastgele Orman metodunun etkisini LL değerlerinin kullanımı ile PI değeri tahmin sürecinde gözlemlemektir.

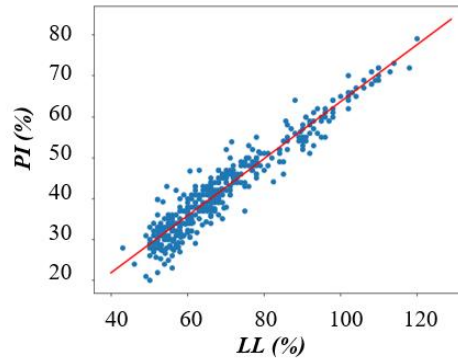
4.1.1. PI değerinin LL ile tahmini (LL: Bağımsız, PI: Bağımlı) (The Prediction of PI via LL (LL:Independent, PI: Dependent))

1. Basit regresyon analizleri kullanımı (The usage of simple regression analysis) :

PI değerinin tek değişken olan LL değerine bağlı olarak tahmininde basit doğrusal regresyon analizleri kullanılmıştır. Analizler için 387 adet likit limit deneyi kullanılmış ve PI ile LL arasındaki ilişki Eşitlik (2) ile elde edilmiştir.

$$PI = 0.696*LL - 6.077 \quad (2)$$

Kıvam limitleri deneylerinin gerçek değerlerine göre Eşitlik (2) 'in hesaplanması ile elde edilen sonuç eğrisi Şekil 5'te verilmiştir. Şekil 5'te PI ile LL arasındaki doğrusal bir ilişki eğilimi bulunduğu aşikârdır.

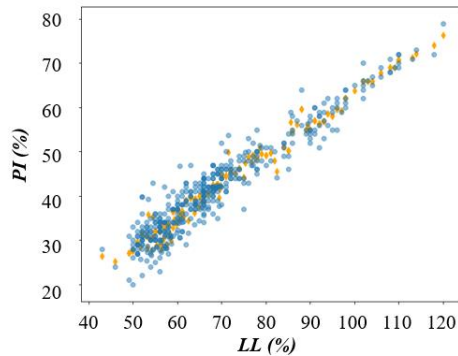


Şekil 5. Birinci veri seti için basit doğrusal regresyon ile PI-LL ilişkisi (The relationship between PI-LL via linear regression (Data set 1))

Bu veri dağılımı ile R^2 (determinasyon katsayısı) değeri 0.91 olarak elde edilmiş, SSE (Sum of Squared Errors) değeri 4692.55, MSE (Mean Squared Errors) değeri 11.015, RMSE (Root Mean Squared Error) değeri 3.318 ve MAE (Mean Absolute Error) değeri 2.51 şeklinde hesaplanmıştır. Bu sonuçlar, PI-LL ilişkisinden yüksek bir doğruluk elde edildiğini ve veri başına düşen hatanın oldukça düşük olduğunu göstermektedir.

2. RO regresyonu yöntemi kullanımı (The usage of random forest regression method):

Basit regresyon analizi ile aynı veri seti kullanılarak RO regresyonu ile analizler yenilenmiştir. Şekil 6'daki turuncu noktalar gerçek verileri, mavi noktalar RO regresyonu yönteminin tahmin ettiği verileri temsil etmektedir. Bu veri dağılımından R^2 değeri 0.935, SSE değeri 3433.46 ve MAE değeri 2.13 şeklinde hesaplanmıştır. RO regresyon yöntemi kullanılarak yapılan analizler, yüksek doğrulukta doğrusal davranış gösteren bir veri setine bile RO regresyonu uygulandığında, doğruluğun 0.91 değerinden 0.935 değerine kadar arttığı, SSE'nin ise 4692.55 değerinden 3433.46 değerine gerilediği ve MAE'nin 2.51 değerinden 2.13 değerine gerilediği gözlemlenmiştir.



Şekil 6. Birinci veri seti için RO regresyonu yöntemi ile PI-LL ilişkisi (The relationship between PI-LL via Random Forest (RF) regression method (Data set 1))

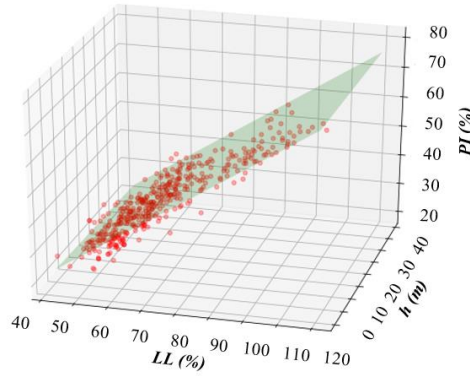
4.1.2. PI değerinin LL ve h ile tahmini (LL, h: Bağımsız, PI: Bağımlı) (The Prediction of PI via LL and h (LL, h: Independent, PI: Dependent))

1. Çoklu regresyon analizleri kullanımı (The usage of multi-linear regression analysis) :

PI değerinin LL ve h değerine bağlı olarak tahmininde çoklu doğrusal regresyon analizleri uygulanmıştır. Analizler için 387 adet likit limit deneyi kullanılmış ve PI ile LL, h arasındaki ilişki Eşitlik (3) ile elde edilmiştir.

$$PI = 0.6965*LL - 0.0413*h - 5.60 \quad (3)$$

Kıvam limit deneylerinin gerçek değerlerine göre Eşitlik (3) 'in hesaplanması ile elde edilen sonuç eğrisi Şekil 7'de verilmiştir. Bu veri dağılımından R^2 değeri 0.912, SSE değeri 4650 ve MAE değeri 2.52 şeklinde hesaplanmıştır.

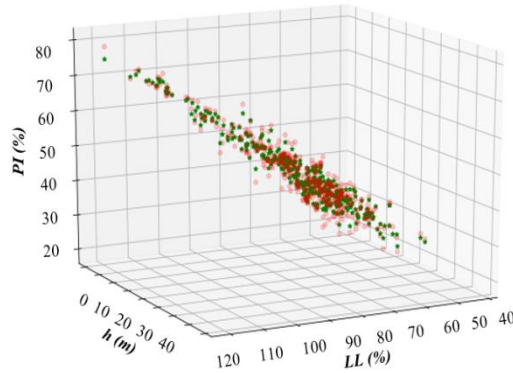


Şekil 7. Birinci veri seti için çoklu doğrusal regresyon analizi ile LL-h-PI ilişkisi (The relationship between PI-LL-h via multi-linear regression (Data set 1))

PI değerinin tahmin edilmesinde kullanılan tekli ve çoklu doğrusal analizler karşılaştırıldığında, aralarındaki farkın sadece derinlik değişkeninin de dikkate alınması olduğu görülmektedir. Ancak derinlik değişkeninin PI tahmin sürecine eklenmesi, tahmin doğruluğuna kayda değer bir katkı sağlayamamıştır. Bu durum, çoklu doğrusal ilişki denkleminde ortaya çıkan h değeri çarpanının sayısal değerinin küçüklüğünden de direk olarak görülebilmektedir.

2. RO regresyonu yöntemi kullanımı (The usage of random forest regression method):

Çoklu regresyon analizi ile aynı veri seti kullanılarak analizler, RO regresyon sistemi ile yenilenmiştir. PI, h ve LL değeri arasındaki ilişki Şekil 8'de verilmiştir ve bu veri dağılımından R^2 değeri 0.98, SSE değeri 1025.3 ve MAE değeri 1.17 şeklinde hesaplanmıştır. RO regresyon yöntemi kullanılarak yapılan analizler, doğrusal davranış gösteren bir veri setine bile RO regresyonu uygulandığında, doğruluğun 0.912 değerinden 0.98 değerine kadar arttığı, SSE'nin ise 4650.56 değerinden 1025.3 değerine gerilediği ve MAE'nin 2.52 değerinden 1.17 değerine gerilediği gözlemlenmiştir.



Şekil 8. Birinci veri seti için RO regresyonu yöntemi ile LL-h-PI ilişkisi (The relationship between PI-LL-h RF regression method (Data set 1))

Gerçekleşen bu durum RO metodunun parametreler arasındaki ilişkiyi genel kullanımdaki regresyonlara göre çok daha iyi yorumladığının bir göstergesidir. RO metodunda analizdeki bağımsız değişken sayısı artırıldıkça analizin doğruluk değerlerinin o derece artacağı söylenebilmektedir.

4.2. Veri Seti Analiz 2 (The analysis of data set 2)

Toplam 59 veriye sahip olan ikinci veri seti (Şekil 2) analizleri dâhilinde dört farklı analiz türü uygulanmaktadır. Bu analiz türleri basit doğrusal regresyon, çoklu doğrusal regresyon, ikinci dereceden polinomal regresyon ve RO regresyon yöntemi şeklindedir. Bu bölümdeki çalışmaların temel amacı, rastgele bir eğilime uyum göstermeyen verilerde Rastgele Orman Metodunun etkisini PI değerlerinin kullanımı ile kompaksiyon parametrelerinin tahmin sürecinde gözlemlemektir.

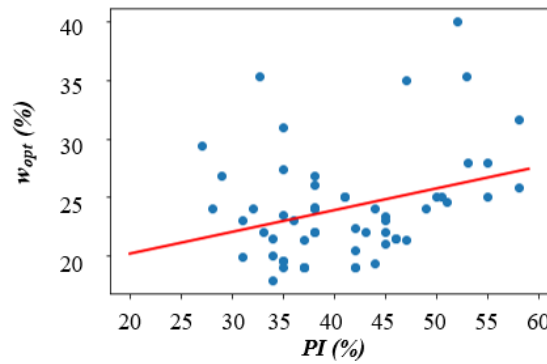
4.2.1. w_{opt} değerinin PI ile tahmini [PI: Bağımsız, w_{opt} : Bağımlı] (The prediction of w_{opt} via PI (PI: Independent, w_{opt} : Dependent))

1. Basit regresyon analizleri kullanımı (The usage of simple regression analysis) :

w_{opt} değerinin tek değişken olan PI değerine bağlı olarak tahmininde basit doğrusal regresyon analizleri uygulanmıştır. Analizler için 59 adet plastisite indisi değeri kullanılmıştır. Bu değerler doğrudan sahadan alınan zemin numunelerine laboratuvar koşullarında uygulanan LL ve PL deneylerinin sonucunda elde edilen ve LL ile PL arasındaki fark ile hesaplanan verileri temsil etmektedir. Basit doğrusal regresyon analizleri ile elde edilen ilişki Eşitlik (4) ile temsil edilmekte olup ilgili veri dağılımı Şekil 9'da sunulmuştur.

$$w_{opt} = 0.186*PI + 16.433 \quad (4)$$

Bu ilişkiyi ifade eden en uygun denklemin R^2 değeri 0.0969 olarak elde edilmiştir. Analizler sonucunda SSE 1084 ve MAE 3.26 olarak hesaplanmıştır.



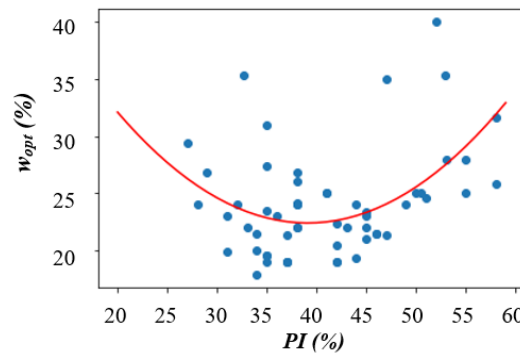
Şekil 9. İkinci veri seti için basit doğrusal regresyon analizi ile PI- w_{opt} ilişkisi (The relationship between PI- w_{opt} via linear regression (Data set 2))

w_{opt} ile PI arasındaki ilişkiyi tahmin etmek amacı ile yapılan basit doğrusal regresyon analizinin çok düşük tahmin yaklaşımı sergilediği açıkça görülmektedir.

2. İkinci dereceden polinomal regresyon analizleri kullanımı (The usage of multi-linear regression analysis):

w_{opt} değerinin tek değişken olan PI değerine bağlı olarak tahmininde ikinci derece polinomal regresyon analizleri kullanılmıştır. Şekil 10'da söz konusu analizler sonucunda elde edilen veri dağılımı gösterilmiştir. Bu analizler ile elde edilen ilişki Eşitlik (5) ile temsil edilmekte olup ilgili veri dağılımı Şekil 10'da sunulmuştur. Şekil 10'da verilen ilişkiden elde edilen R^2 değeri 0.238 şeklindedir. Analizler sonucunda SSE değeri 913.8 ve MAE değeri 2.87 olarak hesaplanmıştır.

$$w_{opt} = 0.0266*PI^2 - 2.08*PI + 63.048 \quad (5)$$

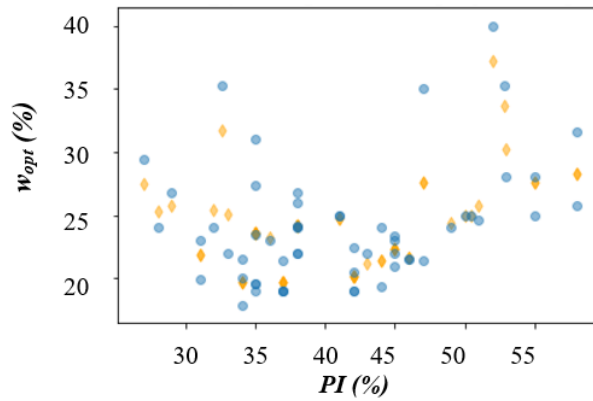


Şekil 10. İkinci veri seti için 2. dereceden polinomal regresyon analizi ile PI ile w_{opt} ilişkisi (The relationship between PI- w_{opt} via quadratic polynomial regression (Data set 2))

Basit doğrusal regresyona göre daha iyimser sonuç veren polinomal regresyon analizlerinin de, uygulamalarda kullanılması muhtemel olmayan derecede az güvenilir sonuç verdiği söylenebilmektedir.

3. RO regresyonu yöntemi kullanımı (The usage of random forest regression method):

RO regresyonu yönteminin kullanılması ile elde edilen PI ve w_{opt} değeri arasındaki ilişki Şekil 11'de verilmiştir ve bu veri dağılımından R^2 değeri 0.711, SSE değeri 346.6 ve MAE değeri 1.76 şeklinde hesaplanmıştır. RO regresyon yöntemi kullanılarak yapılan analizler, dağınık davranış gösteren bir veri setine bile RO regresyonu uygulandığında, doğruluğun 0.0969 değerinden 0.711 değerine kadar arttığı, SSE'nin ise 1084 değerinden 346.6 değerine gerilediği ve MAE'ın 3.26 değerinden 1.76 değerine gerilediği gözlemlenmiştir.



Şekil 11. İkinci veri seti için RO regresyonu yöntemi ile PI- w_{opt} ilişkisi (The relationship between PI- w_{opt} via RF regression method (Data set 2))

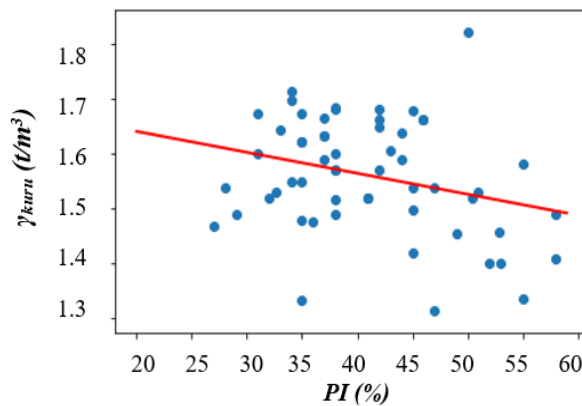
4.2.2. γ_{kuru} değerinin PI ile tahmini [PI: Bağımsız, w_{opt} : Bağımlı] (The prediction of γ_{dry} via PI (PI: Independent, γ_{dry} : Dependent))

1. Basit regresyon analizleri kullanımı (The usage of simple regression analysis):

γ_{kuru} değerinin tek değişken olan PI değerine bağlı olarak tahmininde basit doğrusal regresyon analizleri uygulanmıştır. Analizler için yine 59 adet plastisite indisi değeri kullanılmıştır. Basit doğrusal regresyon analizleri ile elde edilen ilişki Eşitlik (6) ile temsil edilmekte olup ilgili veri dağılımı Şekil 12'de sunulmuştur.

$$\gamma_{kuru} = -0.0038*PI + 1.716 \quad (6)$$

Şekil 12'de verilen ilişkiden R^2 değeri 0.079 olarak elde edilmiştir. Analizler sonucunda SSE değeri 0.56 ve MAE değeri 0.08 olarak hesaplanmıştır. SSE ve MAE değerlerinin sayısal olarak düşük değerlerde kalmasının temel nedeni, kullanılan γ_{kuru} 'nun sayısal değerinin küçüklüğünden kaynaklanmaktadır. Söz konusu yanıltıcı durum, regresyon analizi sonuçlarının bir bütün olarak yorumlanması gerektiğini göstermektedir. Analizler sonucunda elde edilen değerler, önerilen denklemin uygulamalarda kullanılmayacak nitelikte bir ilişki sunduğunu ispatlamaktadır.

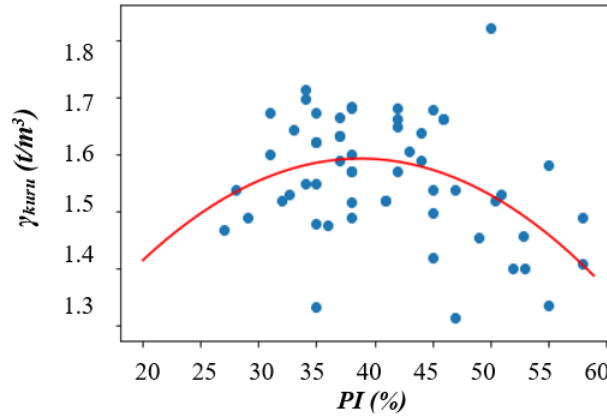


Şekil 12. İkinci veri seti için basit doğrusal regresyon analizi ile PI- γ_{kuru} ilişkisi (The relationship between PI- γ_{dry} via linear regression (Data set 2))

2. İkinci dereceden polinomal regresyon analizleri kullanımı (The usage of quadratic multi-linear regression analysis):

γ_{kuru} değerinin tek değişken olan PI değerine bağlı olarak tahmininde ikinci derece polinomal regresyon analizleri uygulanmıştır. Şekil 13'te söz konusu analizler sonucunda elde edilen veri dağılımı gösterilmiştir. Bu analizler ile elde edilen ilişki Eşitlik (7) ile temsil edilmekte olup ilgili veri dağılımını Şekil 13'te sunulmuştur. Şekil 13'te verilen ilişkiden R^2 değeri 0.179 olarak elde edilmiştir. Analizler sonucunda SSE değeri 0.5 ve MAE değeri 0.073 olarak hesaplanmıştır.

$$\gamma_{kuru} = -0.0005*PI^2 + 0.039*PI + 0.8361 \quad (7)$$

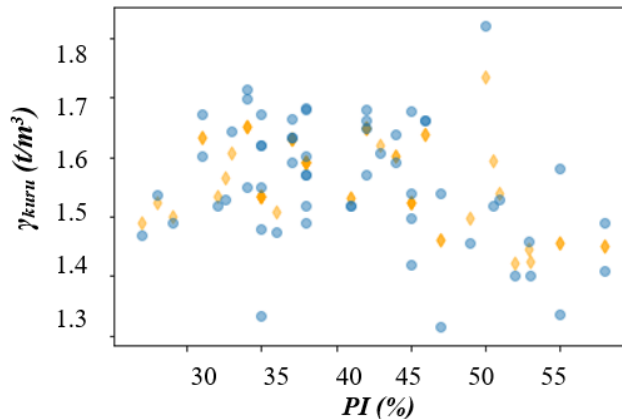


Şekil 13. İkinci veri seti için 2.dereceden polinomal regresyon analizi ile PI- γ_{kuru} ilişkisi (The relationship between PI- γ_{dry} via quadratic polinomial regression (Data set 2))

İkinci derece polinomal regresyon analizlerinde de basit doğrusal regresyon modelinde olduğu gibi, uygulanabilir bir ilişki elde edilemeyeceği görülmektedir.

3. RO regresyonu yöntemi kullanımı (The usage of random forest regression method):

RO regresyonu yönteminin kullanılması ile elde edilen PI ve γ_{kuru} değeri arasındaki ilişki Şekil 14'te verilmiştir. Şekil 14'teki turuncu noktalar RO metodunun tahmin ettiği verileri temsil etmektedir. Bu veri dağılımından R^2 değeri 0.571 olarak, SSE değeri 0.261 ve MAE değeri 0.05 şeklinde hesaplanmıştır. R^2 değerindeki bu artış ile RO metodu γ_{kuru} ile PI arasındaki ilişkide en iyi ifade eden seçenek olmuştur.



Şekil 14. İkinci veri seti için RO regresyonu yöntemi ile PI- γ_{kuru} ilişkisi (The relationship between PI- γ_{dry} via RF regression method (Data set 2))

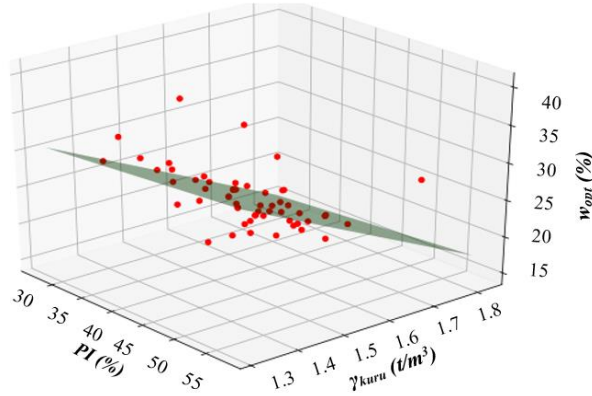
4.2.3. γ_{kuru} - PI - w_{opt} ilişkisinin tahmini (PI, γ_{kuru} : Bağımsız, w_{opt} : Bağımlı) (The prediction of w_{opt} , γ_{dry} via PI (PI: Independent, γ_{dry} : Dependent))

1. Çoklu regresyon analizleri kullanımı (The usage of multi-lineer regression analysis):

γ_{kuru} -PI - w_{opt} ilişkisinin tahmininde ikinci veri seti için çoklu doğrusal regresyon analizleri kullanılmıştır ve bu parametreler arasındaki ilişki Eşitlik (8) ile elde edilmiştir.

$$w_{opt} = 0.07*PI - 30.48*\gamma_{kuru} + 68.75 \quad (8)$$

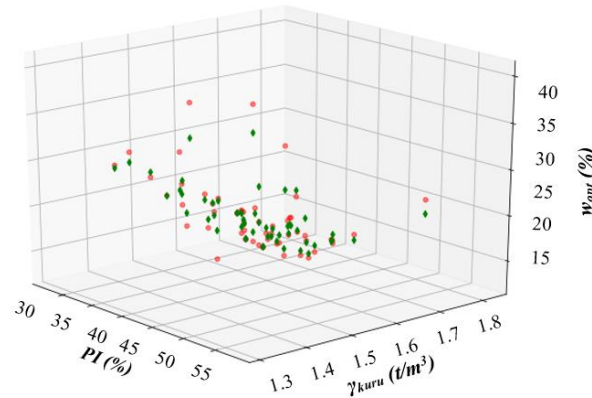
Kıvam limit deneylerinin gerçek değerlerine göre Eşitlik (8) 'in hesaplanması ile elde edilen sonuç eğrisi Şekil 15'te verilmiştir. Bu veri dağılımından R^2 değeri 0.53 olarak, SSE değeri 561.5 ve MAE değeri 2.11 şeklinde hesaplanmıştır. Çoklu değişken kullanılarak yapılan bu analiz, sadece doğrusal olarak yapılan γ_{kuru} -PI ile PI - w_{opt} ilişkilerinin regresyon analizlerinden daha tutarlı bir sonuç vermektedir.



Şekil 15. İkinci veri seti için çoklu doğrusal regresyon ile PI- γ_{kuru} - w_{opt} ilişkisi (The relationship between PI- γ_{dry} - w_{opt} via multi-linear regression (Data set 2))

2. RO regresyonu yöntemi kullanımı (The usage of random forest regression method):

RO regresyonu yönteminin kullanılması ile elde edilen PI- γ_{kuru} - w_{opt} değerleri arasındaki ilişki Şekil 16'da verilmiştir. Şekil 16'daki kırmızı noktalar gerçek verileri ve yeşil noktalar RO metodunun tahmin ettiği değerleri temsil etmektedir. Bu veri dağılımından R^2 değeri 0.912 olarak, SSE değeri 105.46 ve MAE değeri 0.88 şeklinde hesaplanmıştır. R^2 değerindeki bu artış ile RO metodu γ_{kuru} -PI- w_{opt} arasındaki ilişkide en iyi ifade eden seçenek olmuştur.



Şekil 16. İkinci veri seti için RO regresyonu yöntemi ile PI- γ_{kuru} - w_{opt} ilişkisi (The relationship between PI- γ_{dry} - w_{opt} via RF regression method (Data set 2))

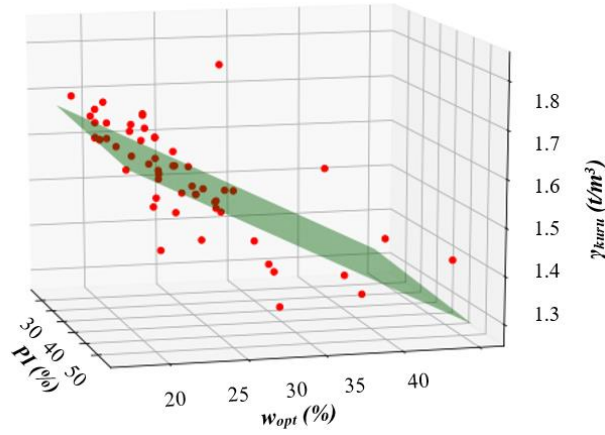
4.2.4. γ_{kuru} - PI - w_{opt} ilişkisinin tahmini (PI, w_{opt} : Bağımsız, γ_{kuru} : Bağımlı) (The prediction of w_{opt} , γ_{dry} via PI (PI, w_{opt} :Independent, γ_{dry} : Dependent))

1. Çoklu regresyon analizleri kullanımı (The usage of multi-linear regression analysis):

γ_{kuru} -PI - w_{opt} ilişkisinin tahmininde ikinci veri seti için çoklu doğrusal regresyon analizleri kullanılmıştır ve bu parametreler arasındaki ilişki Eşitlik (9) ile elde edilmiştir.

$$\gamma_{kuru} = -0.00086*PI - 0.0158*w_{opt} + 1.976 \quad (9)$$

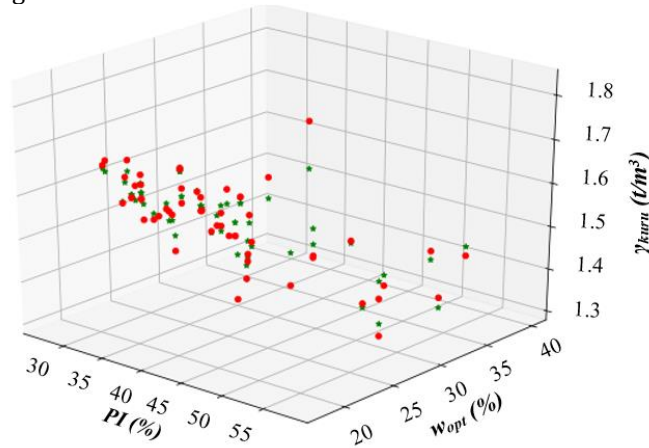
Kıvam limit deneylerinin gerçek değerlerine göre Eşitlik (9) 'un hesaplanması ile elde edilen sonuç eğrisi Şekil 17'de verilmiştir. Bu veri dağılımından elde edilen R^2 değeri 0.523 olarak, SSE değeri 0.291 ve MAE değeri 0.052 şeklinde hesaplanmıştır.



Şekil 17. İkinci veri seti için çoklu doğrusal regresyon ile PI- w_{opt} - γ_{kuru} ilişkisi (The relationship between PI- γ_{dry} - w_{opt} via multi-linear regression (Data set 2))

2. Rastgele Orman regresyon yöntemi kullanımı (The usage of random forest regression method):

RO regresyon yönteminin kullanılması ile elde edilen PI- γ_{kuru} - w_{opt} değerleri arasındaki ilişki Şekil 18'de verilmiştir. Şekil 18'deki kırmızı noktalar gerçek verileri ve yeşil yıldızlar RO metodunun tahmin ettiği değerleri temsil etmektedir. Bu veri dağılımından R^2 değeri 0.925 olarak, SSE değeri 0.046 ve MAE değeri 0.021 şeklinde hesaplanmıştır. Bu değerler, RO regresyonu yönteminin çoklu doğrusal regresyon analizine oranla uygulanabilir nitelikte yaklaşım verdiğini göstermektedir.



Şekil 18. İkinci veri seti için RO regresyonu yöntemi ile PI- w_{opt} - γ_{kuru} ilişkisi (The relationship between PI- γ_{dry} - w_{opt} via RF regression method (Data set 2))

4. Sonuç ve Tartışma (Results and Discussion)

Bu çalışma kapsamında, literatür kaynaklarından derlenen ve içeriğinde geoteknik mühendisliğinin temel deneylerinden kıvam limitleri ve kompaksiyon deneyi sonuçlarını içeren geniş bir veri tabanı hazırlanarak Rastgele Orman regresyonu metodunun kıvam ve kompaksiyon parametreleri arasında ilişki kurmadaki başarısı araştırılmıştır. Bu bağlamda, RO metodunun zemin parametrelerini tahmindeki başarısının ölçülmesi amacı ile basit doğrusal, çoklu doğrusal ve ikinci derece polinomal regresyon analizleri de yapılarak karşılaştırmalı yorumlamalar ile çalışma sonuçlandırılmıştır.

Makale kapsamında uygulanan analizler sonucunda aşağıda maddelerle özetlenen çıkarımlar yapılmıştır:

1. Uygulanan tüm incelemeler RO metodunun, kullanımı sıklıkla tercih edilen diğer standart analiz yöntemlerinden çok daha iyi sonuçlar verdiğini ortaya koymuştur. RO regresyonu, bu çalışmada yapılan analizler içerisinde iyi diğer analiz türüne göre, R^2 değerlerinde ortalama %70, MSE değerlerinde %80, RMSE değerlerinde %56, MAE değerlerinde %62, SSE değerlerinde ise %75 oranında iyileşme sağlamıştır. Bu durumun en önemli nedeni, RO regresyonunun kollektif bir makine öğrenmesi algoritması olmasıdır. Bu bağlamda RO regresyonu sayısını kullanıcının belirlediği karar ağaçlarının tahmin ettiği değerleri toparlayıp ortalamasını alır ve bu şekilde doğruluğu yüksek olan çıktılar verir. Ayrıca, RO regresyonu,

sıklıkla kullanılan diğer analiz yöntemlerindeki gibi herhangi bir lineer, polinomal, üssel, düzlemsel vb. denklemlere bağlı çalışmamaktadır. Bu tür çözümler, denklemin verdiği sonuçlar ile sınırlı kalmaktayken RO regresyonu ise marjinal verilere de ulaşabilir.

2. Doğrusal ilişkiye sahip olan veri setlerinde RO kullanıcıya, zemin parametreleri tahmininde çok yüksek uygulanabilirlik seviyesi sunmaktadır. Ancak, RO metodunun olumsuz özelliği, diğer regresyon metodlarında olduğu şekilde herhangi bir denkleme sahip olmamasıdır. RO regresyonu metodunun amacı, veriyi olduğu noktada tahmin etmektir. Bu nedenle pratik kullanıma uygunluğu yetersiz olarak yorumlanabilmekte olup çok dağınık bir veri setinin bulunduğu durumlarda ise RO metodunun üstünlüğü görülebilmektedir.
3. Yapılan analizlerde bağımsız değişken sayısının artması RO metodundaki doğruluk seviyesini arttırmaktadır. Ancak, bu durumda bağımsız değişkenlerin veri ile ilişkisi de sorgulanması gereken önemli bir konudur.

Sonuç olarak, bu çalışma kapsamında RO regresyon metodu ile yapılan analizlerin; CH türü zeminlerin sahip olduğu PI ile LL veya LL-h ilişkisi ile PI-wopt veya PI- γ kuru veya PI-wopt- γ kuru ilişkilerini temsil etme yeteneğinin yüksek olduğunu ve yeterli geoteknik araştırma bulunamayan durumlarda, veri tahmin sürecinde kıvam ve kompaksiyon özelliklerinin tespit edilmesi amacı ile uygulanabilir nitelik taşıdığını göstermektedir.

Teşekkür (Acknowledgement)

Bu çalışma, İstanbul Üniversitesi-Cerrahpaşa Bilimsel Araştırma Projeleri Birimi tarafından BYP-2020-34856 ve FBA-2020-34051 projeleri ile desteklenmektedir.

Çıkar Çatışması (Conflict of Interest)

Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir. No conflict of interest was declared by the authors.

Kaynaklar (References)

- Akar, Ö., Güngör, O., 2012. Rastgele Orman Algoritması Kullanılarak Çok Banlı Görüntülerin Sınıflandırılması. Jeodezi ve Jeoinformasyon Dergisi. 1 (2), 139-146. DOI: 10.9733/jgg.241212.1t.
- Akar, Ö., Güngör, O., Akar, A., 2010. Rastgele Orman Sınıflandırıcısı ile Arazi Kullanım Alanlarının Belirlenmesi. III. Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Sempozyumu, Gebze, Kocaeli, 142-152.
- Akbağ Arama, Z., Akin, M.S., Nuray, S.E., Dalyan, İ., 2020. Estimation of Consistency Limits of Fine-Grained Soils via Regression Analysis: A Special Case for High and Very High Plastic Clayey Soils in Istanbul. International Advanced Researchers and Engineering Journal (10.09.2020-Kabul edilmiş makale).
- Akman, M., Genç, Y., Ankaralı, H., 2011. Random Forests Yöntemi ve Sağlık Alanında Bir Uygulama. Türkiye Klinikleri Biyoistatistik Dergisi. 3(1) 36-48.
- Archer, K.J., 2008. Empirical Characterization of Random Forest Variable Importance Measure, Computational Statistical Data Analysis, Computational Statistics & Data Analysis, 52(4), 2249-2260.
- Ardakani, A., Kordnaeij, A., 2019. Soil Compaction Parameters Prediction Using GMDH-Type Neural Network and Genetic Algorithm. European Journal of Environmental and Civil Engineering 23:4, 449-462.
- ASTM Standard D 4318. Standard Test Methods for Liquid Limit, Plastic Limit and Plasticity Index of Soils.
- ASTM Standard D 698-12. Standard Test Methods for Laboratory Compaction Characteristics of Soil Using Standard Effort.
- Benson, C.H., Zhai, H., Wang, X., 1994. Estimating Hydraulic Conductivity of Compacted Clay Liners. Journal of Geotechnical Engineering Vol. 120, No. 2.
- Benson, C.H., Daniel, D.E., Boutwell, G.P., 1999. Field Performance of Compacted Clay Liners. Journal of Geotechnical and Geoenvironmental Engineering Vol. 125, No.5.
- Breiman, L., 2001. Random Forests, Machine Learning, 2001 Kluwer Academic Publishers, 45(1), 5-32.
- Breiman, L., Cutler, A., 2005. Random Forest, http://www.stat.berkeley.edu/~g/RandomForests/cc_home.htm.
- Boulila, W., Farah, I.R., Ettabaa, K.S., Solaiman, B., Ben Ghezala, H., 2011. A Data Mining Based Approach To Predict Spatiotemporal Changes in Satellite Images. International Journal of Applied Earth Observation and Geoinformation, 13, 386-395.
- Canillas, C.E., Saloko M.V., 2001. Regression Analysis of Some Factors Influencing Soil Compaction. Soil & Tillage Research 61, 167-178.
- Dewoolkar, M.M., Huzjak, R.J., 2005. Drained Residual Shear Strength of Some Claystones from Front Range, Colorado.
- Dharumarajan, S., Hegdea, R., Singh, S.K., 2017. Spatial Prediction of Major Soil Properties Using Random Forest Techniques - A Case Study In Semi-Arid Tropics Of South India. Geoderma Regional 10 (2017) 154-162.
- Fidan, H., 2020. Random Forest (Rastgele Orman) Algoritması Temelli Süreç İzleme Yönteminin Ambulatuvar Kan Basıncı İzlemede Hipertansiyonun Erken Tanısı İçin Kullanımı. Yüksek Lisans Tezi, Muğla Sıtkı Kocaman Üniversitesi.

- Grimm, R., Behrens, T., Märker, M., Elsenbeer, H., 2008. Soil Organic Carbon Concentrations and Stocks on Barro Colorado Island-Digital Soil Mapping Using Random Forests Analysis. *Geoderma* 146, 102–113.
- Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2004. Random Forest Classification of Multi-Source Remote Sensing and Geographic Data. *IEEE International Geoscience and Remote Sensing Symposium IGARSS '04 Proceedings*, 2, 1049 – 1052.
- Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random Forest For Land Cover Classification. *Pattern Recognition Letters*, 27, 294–300.
- Gunaydin, O., 2009. Estimation of Soil Compaction Parameters by Using Statistical Analyses and Artificial Neural Networks. *Environment Geology* 57:203–215.
- Gunther, E.C., Gerwien, R.W., Heyes, M.P., Bento, P., Stone, D.J. 2003. Prediction of Clinical Drug Efficacy by Classification of Drug-Induced Genomic Expression Profiles in Vitro. *Proceedings of The National Academy of Sciences*. 100(16) 9608-9613. <https://doi.org/10.1073/pnas.1632587100>.
- Hill, T., Lewicki, P., 2006. *Statistics Methods and Applications. A Comprehensive Reference For Science, Industry and Data Mining*, StatSoft, Tulsa, OK.
- Kavzaoglu, T., Sahin, E.S., Colkesen, I., 2012. Heyelan Duyarlılığının İncelenmesinde Regresyon Ağaçlarının Kullanımı: Trabzon Örneği (Assessment of Landslide Susceptibility Using Regression Trees: The Case of Trabzon Province)
- Kemppinen, J., Niittynen, P., Riihimäki, H., Luoto, M., 2018. Modelling Soil Moisture In A High-Altitude Landscape Using LiDAR and Soil Data. *Earth Surface Processes And Landforms* 43, 1019–1031. DOI: 10.1002/esp.4301.
- Khuntia, S., Muftaba, H., Patra, C., Farooq, K., Sivakugan, N.&Das, B.M., 2014. Prediction of Compaction Parameters of Coarse Grained Soil Using Multivariate Adaptive Regression Splines (MARS). *International Journal of Geotechnical Engineering* 9:1, 79-88.
- Liaw A, Wiener M., 2002. Classification and Regression By Random Forest, *R News*, 2(3).
- Ließ, M., Glaser, B., Huwe, B., 2011. Functional Soil-Landscape Modelling To Estimate Slope Stability In A Steep Andean Mountain Forest Region. *Geomorphology* 132 (3–4), 287–299.
- Magerman, D.M., 1995. Statistical Decision-Tree Models for Parsing. *DecisionTree Modeling*. 276–283.
- Matteo, L.D, Bigotti, F., Ricco, R., 2009. Best-Fit Models to Estimate Modified Proctor Properties of Compacted Soil. *ASCE*.
- Mehta, B., Sachan, A., 2017. Effect of Mineralogical Properties of Expansive Soil on Its Mechanical Behavior. *Geotechnical Geology Engineering* 35:2923–2934.
- Omar, M., Shanableh, A., Basma, A., Barakat, S., 2003. Compaction Characteristic of Granular Soils in United Arab Emirates. *Geotechnical and Geological Engineering* 21: 283-295.
- Ouedraogo, I., Defourny, P., Vanclooster, M., 2019. Application of Random Forest Regression and Comparison of Its Performance To Multiple Linear Regression In Modeling Groundwater Nitrate Concentration At The African Continent Scale. *Hydrogeology Journal*, 27:1081–1098. <https://doi.org/10.1007/s10040-018-1900-5>.
- Özdarıcı, Ok A., Akar, Ö., Güngör, O., 2011. Rastgele Orman Sınıflandırma Yöntemi Yardımıyla Tarım Alanlarındaki Ürün Çeşitliliğinin Sınıflandırılması. *TUFUAB 2011 VI. Teknik Sempozyumu, Antalya, Türkiye*, ss.1-7.
- Pal, M., 2003. Random Forest For Land Cover Classification. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS '03 Proceedings*, 6, 3510-3512.
- Pal, M., 2005. Random Forest Classifier For Remote Sensing Classification. *International Journal of Remote Sensing*. 26(1) 217–222. <https://doi.org/10.1080/01431160412331269698>.
- Pal, M., Mather, P.M., 2003. An Assessment of The Effectiveness of Decision Tree Methods For Land Cover Classification. *Remote Sensing Of Environment*, 86, 554-565.
- Peng, W., Chen, J., Zhou, H., 2009. An Implementation of Decision Tree Learning Algorithm.
- Pham, T.B., Qi, C., Ho, L.S., Thoi, T.N., Ansari, N.A., Nguyen, M.D., Nguyen, H.D., Ly, H.B., Le, H.V., Prakash, I., 2020. A Novel Hybrid Soft Computing Model Using Random Forest and Particle Swarm Optimization for Estimation of Undrained Shear Strength of Soil. *Sustainability*, 12, 2218; DOI:10.3390/su12062218.
- Rastgou, M., Bayat, H., Mansoorizadeh, M., Gregory, A.S., 2020. Estimating The Soil Water Retention Curve: Comparison of Multiple Nonlinear Regression Approach and Random Forest Data Mining Technique. *Computers and Electronics in Agriculture* 174, 105502.
- Sahin, K.E., 2018. Heyelan Duyarlılık Haritası İçin Adımsal Regresyona Dayalı Faktör Seçme Yönteminin Etkinliğinin Araştırılması. *Harita Dergisi*. Sayı 159.
- Safavian, S.R., Landgrebe, D., 1991. A Survey of Decision Tree Classifier Methodology. *IEEE Transactions On Systems Man and Cybernetics*, 21, 660-674.
- Segal, M.R., 2003. *Machine Learning Benchmarks and Random Forest Regression*. Center for Bioinformatics and Molecular Biostatistics, UC San Francisco. 18(3), 1-14.
- Shukla, G., Garg, R.D., Srivastava, H.S. & Garg, K.P., 2018. An Effective Implementation and Assessment of A Random Forest Classifier As A Soil Spatial Predictive Model. *International Journal of Remote Sensing*, 39:8, 2637-2669, DOI: 10.1080/01431161.2018.1430399.
- Singh, B., Sihag, B., Singh, K., 2017. Modelling of Ompact of Water Quality On Infiltration Rate of Soil By Random Forest Regression Model. *Earth Syst. Environ.* 3:999–1004. DOI: 10.1007/s40808-017-0347-.
- Singhal, S., Houston, S.L., Houston, W.N., 2015. Swell Pressure, Matric Suction, and Matric Suction Equivalent for Undisturbed Expansive Clays. *Can. Geotechnical Journal* 52: 356–366.
- Sinha, S.K., Wang, M.C., 2008. Artificial Neural Network Prediction Models for Soil Compaction and Permeability. *Geotechnical Geology Engineering* 26:47-64.
- Thompson, M.J., White, D.J., 2008. Estimating Compaction of Cohesive Soils from Machine Drive Power. *ASCE*.
- Viji, V.K., Lissy, K.F., Sobha, C. & Benny M.A., 2013. Predictions on Compaction Characteristics of Fly Ashes Using Regression Analysis and Artificial Neural Network Analysis. *International Journal of Geotechnical Engineering* 7:3, 282-291.
- Viscara Rossel, R.A., Behrens, T., 2010. Using Data Mining To Model and Interpret Soil 650 Diffuse Reflectance Spectra. *Geoderma* 158 (1–2), 46–54.

- Waske B., Heinzl V., Braun M., Menz G., 2007. Random Forests for Classifying Multi-Temporal Sar Data, Proc. 'Envisat Symposium Montreux, Switzerland, <http://envisat.esa.int/envisatsymposium/proceedings/sessions/3D3/461589wa.pdf>.
- Waske, B., Braun, M., 2009. Classifier Ensembles For Land Cover Mapping Using Multitemporal SAR Imagery. ISPRS Journal of Photogrammetry and Remote Sensing 64 (2009) 450_457.
- Zhang, P., Yin, Z., Jin, Y., Chan, T.H.T., 2020. A Novel Hybrid Surrogate Intelligent Model For Creep Index Prediction Based On Particle Swarm Optimization and Random Forest. Engineering Geology 265, 105328.