

[1279]

# VERİ MADENCİLİĞİ YÖNTEMLERİ KULLANILARAK ÇOK-BANTLI UYDU GÖRÜNTÜ VERİLERİNİN SINIFLANDIRILMASI İÇİN ALGORİTMA SEÇİMİ

Hamza EROL<sup>1</sup>

<sup>1</sup>Prof. Dr., Çukurova Üniversitesi, İstatistik Bölümü, 01330, Adana, [herol@cu.edu.tr](mailto:herol@cu.edu.tr)

## ÖZET

*Veri madenciliği yöntemlerinde uzaktan algılanmış çok bantlı uydu görüntülerinin sınıflandırılması için algoritma seçimi çok önemlidir. Bu çalışmada, WEKA veri madenciliği yazılımı kullanılarak uzaktan algılanmış çok bantlı uydu görüntülerinin sınıflandırılması amacıyla algoritma seçimi için bir strateji önerilmiştir. Algoritma seçimi için önerilen stratejinin çalışma prensibi, WEKA veri madenciliği yazılımının: veriyi açıklama ortamı - explorer ve algoritma seçimi ortamı - experimenter kullanılarak uzaktan algılanmış çok bantlı uydu görüntü verisinin sınıflandırılması uygulaması üzerinde açıklanmıştır.*

**Anahtar Sözcükler:** Algoritma seçimi, Veri sınıflandırma, Veri kümeleme, Veri madenciliği, Çok bantlı görüntü verisi.

## ABSTRACT

### ALGORITHM SELECTION FOR CLASSIFICATION OF REMOTELY SENSED MULTISPECTRAL IMAGE DATA USING DATA MINING METHODS

*Algorithm selection is very important for classification of remotely sensed multispectral image data in data mining methods. A strategy of algorithm selection will be given for classification of remotely sensed multispectral image data using WEKA data mining software. The working principle of the proposed strategy of algorithm selection will be presented on classification of multispectral image data with different data mining classification algorithms using explorer and experimenter environments of WEKA data mining software.*

**Keywords:** Algorithm selection, Data classification, Data clustering, Data mining, Multispectral image data.

## 1. GİRİŞ

Uzaktan algılanmış çok bantlı uydu görüntü verileri büyük veridir. Bu verileri işlemek ve sınıflandırmak amacıyla veri madenciliği yöntemlerini kullanma ve uygulama durumu ortaya çıkmaktadır. Veri madenciliği yöntemlerindeki gelişmeler ve gelişmelerin veri madenciliği yazılımlarıyla güncellenmesi büyük verinin işlenmesine ve sınıflandırılmasına olanak sağlamıştır. Veri madenciliği yöntemleri son zamanlarda uzaktan algılanmış görüntülerin analizinde kullanılmaya başlanmıştır.

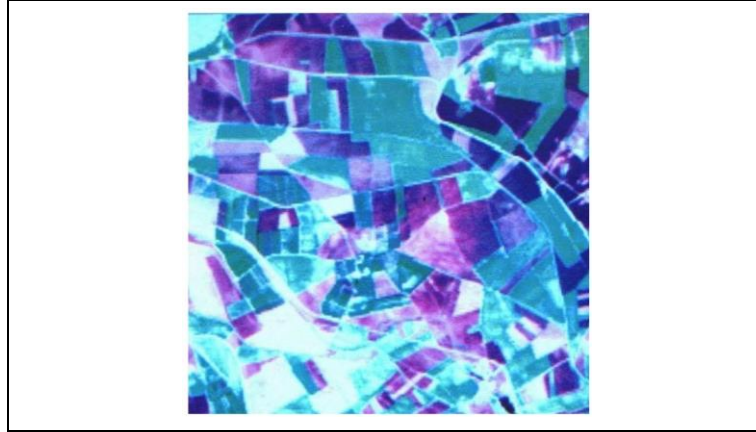
Briem ve ark. (2002). değişik uydu görüntü verilerine çoklu sınıflandırma algoritmaları uygulamışlar ve performanslarını incelemişlerdir. Vitti ve Bezzi (2004). geniş alan arazide ekili alanlardaki bitki örtüsünü sınıflandırmak amacıyla birleştirilmiş veri madenciliği yöntemlerini kullanmışlardır. Çalışmalarında farklı veri madenciliği sınıflandırma algoritmalarının performansını karşılaştırmışlardır. Yang ve ark. (2010). çok ölçekli ve çok özellikli işleme teknikleriyle ilgili uzaktan algılanmış görüntülerden bilgi madenciliği çalışmışlar ve uygulama yapmışlardır. Vieira ve ark. (2012), geniş arazilerde şeker kamışı alanlarının belirlenmesi otomasyonu için yöntem geliştirmek amacıyla zaman serilerini kullanarak uzaktan algılanmış görüntülerin analizini çalışmışlardır. Gao ve ar. (2012), geniş arazilerin termal uydu görüntülerinde alanları netleştirmek için bir veri madenciliği yaklaşımı geliştirmişlerdir. Arun (2013), uydu görüntülerinin sınıflandırma doğruluk tahmini için bir veri madenciliği yöntemi geliştirmiştir. Gañçarski (2014), okyanus büyük veri madenciliğinde uzaktan algılanmış uydu görüntülerinin analizini çalışmıştır. Kumar ve Chatterjee (2016), UCI verisetinde uydu görüntü verilerinin veri madenciliği yöntemleriyle analizinde WEKA yazılımını kullanmışlardır. Gögebakan ve Erol (2016), değişken veri parçalanmalarını kullanılarak model tabanlı kümelemeye dayalı yeni bir sınıflandırma yöntemi geliştirerek uzaktan algılanmış çok bantlı uydu görüntülerinin sınıflandırılması için veri madenciliği yaklaşımında bulunmuşlardır.

## 2. YÖNTEM

Bu bölümde, çalışmada algoritma seçimi için önerilen stratejinin çalışma prensibi, WEKA veri madenciliği yazılımının: veriyi açıklama - explorer ortamı ve algoritma seçimi - experimenter ortamı kullanılarak uzaktan algılanmış çok bantlı uydu görüntü verisinin sınıflandırılması uygulaması üzerinde açıklanacaktır.

## 2.1.Uzaktan Algılanmış Çok Banthlı Uydu Görüntü Verisinin WEKA Veri Madenciliği Yazılımı İle Analizi İçin Hazırlanması

Bu çalışmada Seyhan Ovası Adana - Türkiye ( $\approx 37^{\circ}\text{N}$ ,  $36^{\circ}\text{E}$ ) bölgesinde yer alan tarımsal bölgenin 27 Mart 1992 tarihli (Path 175 - Row 34) 198 satır ve 200 sütundan oluşan Landsat Thematic Mapper çok banthlı uydu görüntü verisinin 3.bant, 4.bant ve 5.bant değerleri kullanılmıştır. Çalışılan tarımsal bölgenin uydu görüntüsü Şekil 1 de gösterilmiştir.



**Şekil 1.** Seyhan Ovası Adana - Türkiye ( $\approx 37^{\circ}\text{N}$ ,  $36^{\circ}\text{E}$ ) bölgesinde yer alan tarımsal bölgenin 27 Mart 1992 tarihli (Path 175 - Row 34) 198 satır ve 200 sütundan oluşan Landsat Thematic Mapper uydu görüntüsü (Erol ve Akdeniz 2005).

Şekil 1 deki uzaktan algılanmış çok banthlı uydu görüntü verisi daha önce Çalış ve Erol (2013) tarafından çalışılmış ve karma diskriminant analizi kullanılarak hem alan bazında hem de piksel bazında sınıflandırılmıştır. Tarımsal bölgenin karma diskriminant analizi sonucuna göre beş bitki türü belirlemişlerdir. Bunları buğday (altı çeşit), patates (dört çeşit), bostan (altı çeşit), naranciyeye (üç çeşit) ve toprak (beş çeşit) şeklinde belirtmişlerdir. Tarımsal bölgede 5 ana türünü ve bu türlerin 24 alt çeşitini belirlemişlerdir.

```
UACBSX123k5csv.arff - Not Defteri
Dosya Düzen Biçim Görünüm Yardım
@relation UACBSX123k5csv
@attribute X1 numeric
@attribute X2 numeric
@attribute X3 numeric
@attribute Class5 {1,2,3,4,5}
@data59,84,112,4
54,79,114,4
46,83,101,4
57,84,116,4
51,81,109,4
44,80,98,4
45,77,93,4
```

**Şekil 2.** Uzaktan algılanmış çok banthlı uydu görüntü verisinin sayısallaştırılmış 3.bant, 4.bant ve 5.bant değerlerinin 5 türe göre grup etiketli ARFF dosyasının not defteri görünümü.

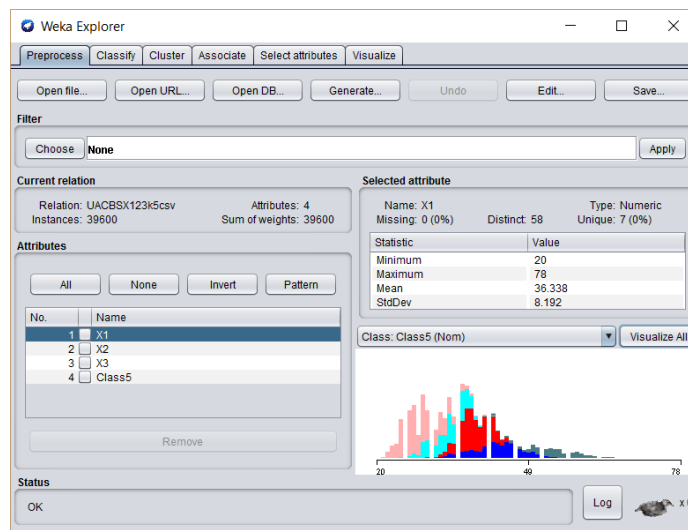
```
UACBSX123k24csv.arff - Not Defteri
Dosya Düzen Biçim Görünüm Yardım
@relation UACBSX123k24csv
@attribute X1 numeric
@attribute X2 numeric
@attribute X3 numeric
@attribute Class24 {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24}
@data59,84,112,16
54,79,114,16
46,83,101,16
57,84,116,16
51,81,109,16
44,80,98,16
45,77,93,20
```

**Şekil 3.** Uzaktan algılanmış çok bantlı uydu görüntü verisinin sayısallaştırılmış 3.bant, 4.bant ve 5.bant değerlerinin 24 alt çeşite göre grup etiketli ARFF dosyasının not defteri görünümü.

Şekil 1 deki uzaktan algılanmış çok bantlı uydu görüntü verisinin sayısallaştırılmış 3.bant, 4.bant ve 5.bant değerlerinin 5 türe ve 24 alt çeşite göre k-ortalamar algoritması kullanılarak grup etiketlenir ve grup etiketli veriler Excel dosyasına aktarılır. Excel dosyasındaki veriler CSV (Comma Separated Value - Virgül ile ayrılmış değer) formatında kaydedilir. CSV formatlı Excel dosyası WEKA veri madenciliği yazılımında yüklenebilir ve ARFF (Attribute Related File Format - Özellik ilişkili dosya formatı) tipinde dosya olarak kaydedilir. Uzaktan algılanmış çok bantlı uydu görüntü verisinin sayısallaştırılmış 3.bant, 4.bant ve 5.bant değerlerinin 5 türe göre grup etiketli ve 24 alt çeşite göre grup etiketli ARFF dosyaları sırasıyla Şekil 2 ve Şekil 3 de gösterilmiştir.

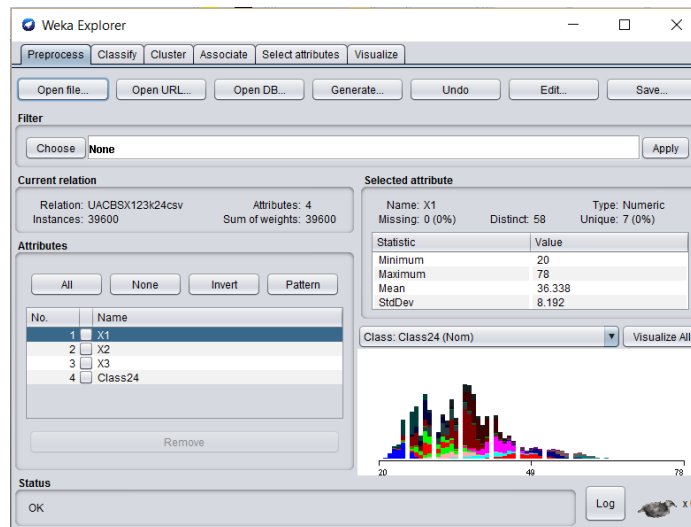
## 2.2.Uzaktan Algılanmış Çok Bantlı Uydu Görüntü Verisinin WEKA Veri Madenciliği Yazılımına Yüklenmesi

Uzaktan algılanmış çok bantlı uydu görüntü verisinin sayısallaştırılmış 3.bant, 4.bant ve 5.bant değerlerinin 5 türe göre grup etiketli ARFF dosyasının WEKA veri açıklama - explorer ortamına yüklenmiş hali Şekil 4 de gösterilmiştir.



**Şekil 4.** Uzaktan algılanmış çok bantlı uydu görüntü verisinin sayısallaştırılmış 3.bant, 4.bant ve 5.bant değerlerinin 5 türe göre grup etiketli ARFF dosyasının WEKA veri açıklama - explorer ortamına yüklenmiş hali.

Uzaktan algılanmış çok bantlı uydu görüntü verisinin sayısallaştırılmış 3.bant, 4.bant ve 5.bant değerlerinin 24 alt çeşite göre grup etiketli etiketli ARFF dosyasının WEKA veri açıklama - explorer ortamına yüklenmiş hali Şekil 5 te gösterilmiştir.

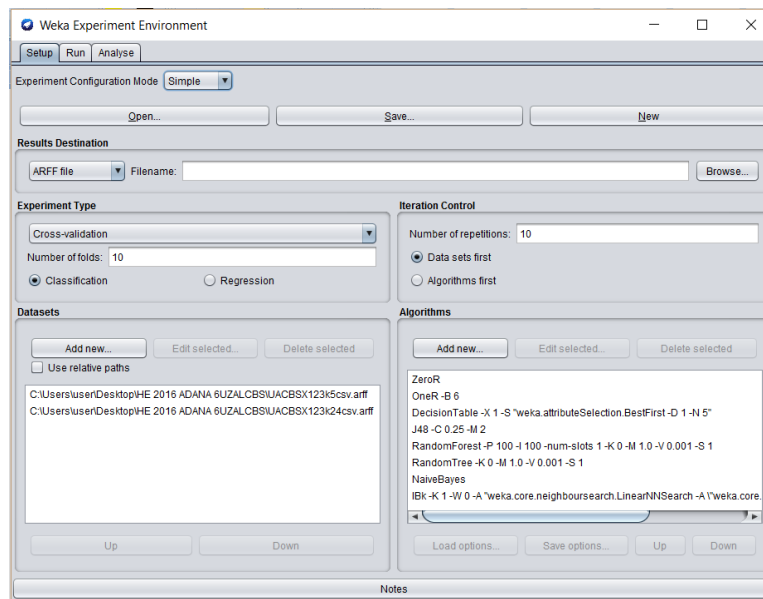


**Şekil 5.** Uzaktan algılanmış çok bantlı uydu görüntü verisinin sayısallaştırılmış 3.bant, 4.bant ve 5.bant değerlerinin 24 alt çeşite göre grup etiketli ARFF dosyasının WEKA veri açıklama - explorer ortamına yüklenmiş hali.

WEKA veri açıklama - explorer ortamında ARFF dosyasındaki tüm değişkenlerle ilgili sayısal ve grafiksel bilgiler görüntülenebilir.

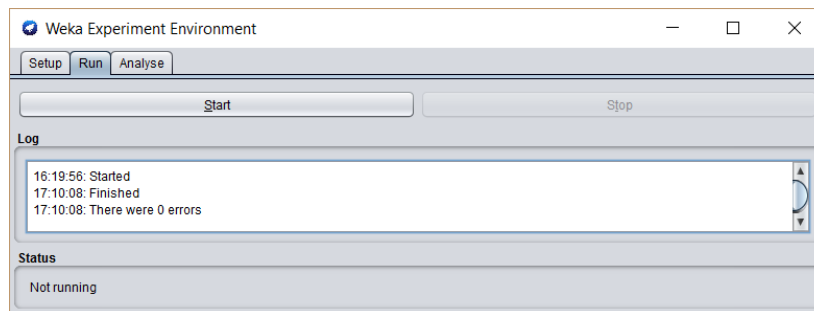
### 2.3.WEKA Veri Madenciliği Yazılımının Algoritma Seçimi - Experimenter Ortamında Sınıflandırma İçin Farklı Algoritmaların Performanslarının Belirlenmesi

WEKA veri madenciliği yazılımının algoritma seçimi - experimenter ortamında sınıflandırma için farklı algoritmaların performanslarının belirlenmesi amacıyla uzaktan algılanmış çok bantlı uydu görüntü verisinin sayısallaştırılmış 3.bant, 4.bant ve 5.bant değerlerinin 5 türe göre grup etiketli arff dosyası ve 24 alt çeşite göre grup etiketli ARFF dosyası WEKA veri madenciliği yazılımının algoritma seçimi - experimenter ortamı Setup sekmesinden yüklenir. Algoritma bölümünden WEKA veri madenciliği yazılımında yer alan sınıflandırma algoritma gruplarından farklı algoritmalar seçilir. Bu analiz için Rules grubundan: Decision Table, ZeroR, OneR; Trees grubundan: J48, RandomForest, RandomTree; Bayes grubundan: NaiveBayes; Lazy grubundan: IBK algoritmaları olmak üzere toplam sekiz algoritma seçilmiştir. Bu durum Şekil 6 da gösterilmiştir.



**Şekil 6.** WEKA veri madenciliği yazılımının algoritma seçimi - experimenter ortamında uzaktan algılanmış çok bantlı uydu görüntü verisinin sayısallaştırılmış 3.bant, 4.bant ve 5.bant değerlerinin 5 türe göre grup etiketli arff dosyası ve 24 alt çeşite göre grup etiketli ARFF dosyasının yüklenmesi; Algoritma bölümünden Rules grubundan: Decision Table, ZeroR, OneR; Trees grubundan: J48, RandomForest, RandomTree; Bayes grubundan: NaiveBayes; Lazy grubundan: IBK algoritmalarının seçilmesi.

WEKA veri madenciliği yazılımının algoritma seçimi - experimenter ortamı Run sekmesinde sınıflandırma için farklı algoritmaların performanslarının belirlenmesi amacıyla Start düğmesine basılır ve algoritmaların performanslarının belirlenmesi işlemine başlanır. İşlem bitene kadar beklenir. İşlem bitince Run sekmesinde bilgi mesajı gelir. Bu durum da Şekil 7 de gösterilmiştir.

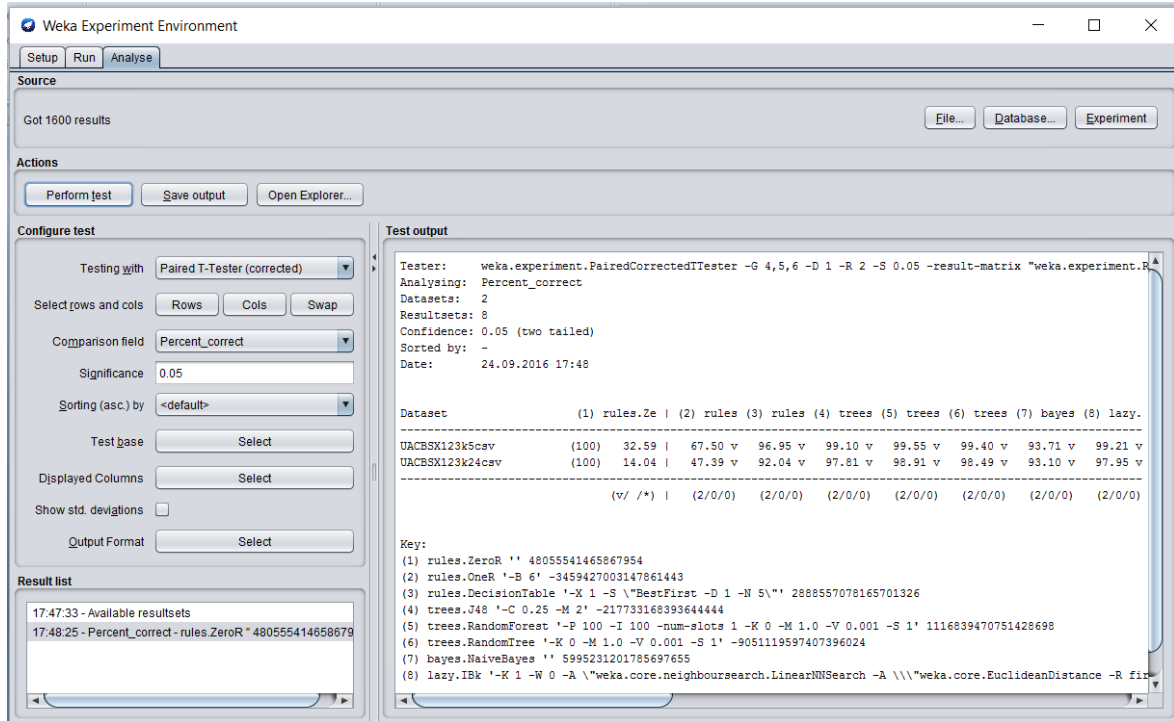


**Şekil 7.** WEKA veri madenciliği yazılımının algoritma seçimi - experimenter ortamı Run sekmesinde sınıflandırma için farklı algoritmaların performanslarının belirlenmesi işleminin tamamlanması sonucunda bilgi mesajı.

WEKA veri madenciliği yazılımının algoritma seçimi - experimenter ortamı Run sekmesinde sınıflandırma için farklı algoritmaların performanslarının belirlenmesi işleminin tamamlanması sonucunda bilgi mesajından işlem

süresi belirlenir.

WEKA veri madenciliği yazılımının algoritma seçimi - experimenter ortamı Analyse sekmesinde sınıflandırma için farklı algoritmaların performanslarının belirlenmesi işleminin sonucu önce Experiment sonra Perform Test düğmelerine basılarak çıktı bölümünde görüntülenir. WEKA veri madenciliği yazılımının algoritma seçimi - experimenter ortamı Analyse sekmesinde sınıflandırma için farklı algoritmaların performanslarının çıktı bölümünde görüntülenmesi Şekil 8 de gösterilmiştir.



Şekil 8. WEKA veri madenciliği yazılımının algoritma seçimi - experimenter ortamı Analyse sekmesinde sınıflandırma için farklı algoritmaların performanslarının çıktı bölümünde görüntülenmesi.

### 3.SONUÇLAR

WEKA veri madenciliği yazılımının algoritma seçimi - experimenter ortamında Analyse sekmesinde sınıflandırma için farklı algoritmaların performanslarının çıktılarına göre uzaktan algılanmış çok bantlı uydu görüntü verisinin sayısallaştırılmış 3.bant, 4.bant ve 5.bant değerlerinin 5 türe göre grup etiketli arff dosyası ve 24 alt çeşite göre grup etiketli ARFF dosyası için Rules grubundan: Decision Table, ZeroR, OneR; Trees grubundan: J48, RandomForest, RandomTree; Bayes grubundan: NaiveBayes; Lazy grubundan: IBk seçilen algoritmalarından en iyi performans gösteren sekiz algoritmadan 5. algoritma olan Trees grubundan RandomForest algoritması en yüksek sınıflandırma performanslı algoritma olarak seçilir. Bu sonuçtan sonra belirlenen en iyi algoritma kullanılarak uzaktan algılanmış çok bantlı uydu görüntü verisinin sınıflandırma işlemine geçilir.

### KAYNAKLAR

**Arun, P. V.**, 2013. A visual mining based fame work for classification accuracy estimation. American Journal of Remote Sensing, 1(2), p.47-52, <http://www.sciencepublishinggroup.com/j/ajrs>, doi: 10.11648/j.ajrs.20130102.15.

**Briem, G. J., Benediktsson, J. A. and Sveinsson, J. R.**, 2002. Multiple Classifiers Applied to Multisource Remote Sensing Data. IEEE Transactions On Geoscience And Remote Sensing, Vol. 40, No. 10, p. 2291-2299.

**Çalış, N. and Erol, H.**, 2013. A new per-field classification method using mixture discriminant analysis. Journal of Applied Statistics 39(10):1-12. DOI: 10.1080/02664763.2012.702263.

**Erol, H. ve Akdeniz, F.**, 2005. A per-field classification method based on mixture distribution models and an application to Landsat Thematic Mapper data. International Journal of Remote Sensing 26(6):1229-1244. DOI: 10.1080/01431160512331326800

- Gançarski, P.**, 2014. Introduction to data mining: Example of remote sensing image analysis. Ocean's Big Data Mining, 2014. Data mining in large sets of complex oceanic data: new challenges and solutions. 8-9 Sep 2014 Brest - France. [https://oceandatamining.sciencesconf.org/data/program/OBIDAM14\\_Gancarski.pdf](https://oceandatamining.sciencesconf.org/data/program/OBIDAM14_Gancarski.pdf) p.1-123.
- Gao, F., Kustas, W. P. and Anderson, M. C.**, 2012. A Data Mining Approach for Sharpening Thermal Satellite Imagery over Land. Remote Sensing, 4, 3287-3319; doi:10.3390/rs4113287.
- Gögebakan, M. ve Erol, H.**, 2016. Değişken Veri Parçalanmaları Kullanılarak Model Tabanlı Kümelemeye Dayalı Yeni Bir Sınıflandırma Yöntemi: Uzaktan Algılanmış Çok Bantlı Görüntü Sınıflandırılması İçin Veri Madenciliği Yaklaşımı. 6. Uzaktan Algılama - Coğrafi Bilgi Sistemleri Sempozyumu 5-7 Ekim 2016. Çukurova Üniversitesi - Adana.
- Kumar, A. and Chatterjee, I.**, 2016. Data Mining: An experimental approach with WEKA on UCI Dataset International Journal of Computer Applications (0975 – 8887) Volume 138 – No.13, p. 23-28.
- Vieira, M. A., Formaggio, A. R., Rennó, C. D., Atzberger, C., Aguiar, D. A. and Mello, M., P.**, 2012. Object Based Image Analysis and Data Mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. Remote Sensing of Environment, Volume 123, Pages 553-562.
- Vitti, A. And Bezzi, M.**, 2004. A Grass - Data Mining Integrated Procedure For Land Cover Classification. <Http://Geomatica.Como.Polimi.It/Workbooks/N4/Articoli/Bezzi-Vitti.Pdf>. Geomatics Workbooks Vol. 4. 5th Italian Grass Users Meeting Proceedings.
- WEKA**, 2016. Weka 3.8 is the latest stable version, <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>
- Yang A, X. M., Cui B, W., Gong A, J. M., And Zhang A., T.**, 2010. Information Mining From Remote Sensing Imagery Based On Multi-Scale And Multi-Feature Processing Techniques. Joint International Conference on Theory, Data Handling and Modelling in GeoSpatial Information Science, 26-28 May 2010, Hong Kong. p. 158-161. [http://www.isprs.org/proceedings/XXXVIII/part2/Papers/56\\_Paper.pdf](http://www.isprs.org/proceedings/XXXVIII/part2/Papers/56_Paper.pdf).