

Ardışıl ileri yönlü öznitelik seçim algoritmasında etkin özniteliklerin belirlenmesi

Önder AYDEMİR^{*1}

¹ Karadeniz Teknik Üniversitesi, Elektrik Elektronik Mühendisliği Bölümü, Trabzon

Makale Gönderme Tarihi: 24.10.2016

Makale Kabul Tarihi: 15.02.2017

Öz

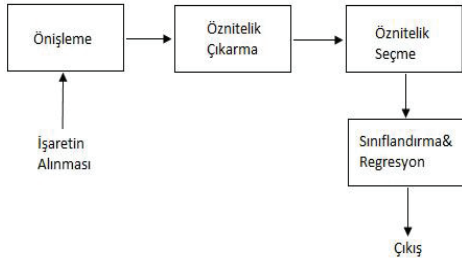
Bu çalışmada, örüntü tanıma ve makine öğrenmesi uygulamalarında öznitelik çıkarma işleminden sonra etkin özniteliklerin belirlenmesi için kullanılan yöntemlerden; ardışıl ileri yönlü öznitelik seçme (AİYÖS) ve ardışıl geri yönlü öznitelik seçme (AGYÖS) algoritmaları sınıflandırma doğruluğu ve hız bakımından karşılaştırılmıştır. Bu yöntemler, eğitim kümesinden çapraz doğrulama yöntemi ile en yüksek doğrulama başarısını veren öznitelikleri belirlerken, alt eğitim kümeleri rastgele seçilir. Bundan ötürü bu yöntemlerin her koşulmasında farklı öznitelikler sonuç olarak seçilebilmektedir. Dolayısıyla farklı özniteliklerin seçimi ise önerilecek modelin test performansını olumlu/olumsuz etkilemektedir. Bu çalışmada bu rastgele seçimin dezavantajını ortadan kaldırmak için bir yöntem önerilmiştir. Önerilen yöntemin kararlılığını göstermek amacıyla eğitim aşamasında AİYÖS ve AGYÖS algoritmaları 1000 defa koşurulmakta ve belirlenen eşik değerden fazla sayıda seçilen öznitelikler etkin öznitelikler olarak belirlenmektedir. Elde edilen sonuçlara göre; AİYÖS algoritmasının AGYÖS'e göre yaklaşık 40 kat daha hızlı olduğu ve %22 daha fazla sınıflandırma doğruluğu sağladığı görülmüştür.

Anahtar Kelimeler: öznitelik seçme; ardışıl ileri yönlü öznitelik seçme; ardışıl geri yönlü öznitelik seçme; sınıflandırma doğruluğu; hesaplama hızı.

^{*}Yazışmaların yapılacağı yazar: Önder AYDEMİR. onderaydemir@ktu.edu.tr; Tel: (462) 377 40 15

Giriş

Örüntü tanıma ve makine öğrenmesi son yıllarda üzerinde çokça araştırma yapılan konuların başında gelmektedir (de la Fraga ve Coello, 2011; Jain vd., 2000; Xian ve Xian, 2013). Bu uygulamalarda genel işlem bileşenleri Şekil 1’de gösterildiği gibi işaretin alınması, önışleme, öznelik çıkarma, öznelik seçme ve sınıflandırma/regresyondur. Bu bileşenlerden öznelik çıkarma; işaretlerin önemli özelliklerinin çıkarılıp işaretleri daha düşük boyutlu vektörle temsil etmek ve farklı kategorideki işaretleri birbirinden ayırt edebilmek için yapılmaktadır. Örneğin beyin bilgisayar arayüzü teknoloji bir tekerlekli sandalyenin kullanıcısı, durum ileri götürmek istediği durumda diğer durumlardan (geri, sağa, sola vb.) farklı olarak elektroensefalogram işaretleri nasıl bir örüntü oluşturuyor sorusunun cevabını bulmak bir öznelik çıkarma işlemidir. Diğer taraftan, öznelik çıkarma aşaması örüntü tanıma ve makine öğrenmesi sisteminde önemli ve kritik bir aşamadır. Çünkü çıkarılan özneliklerin kullanılabilirliği direkt olarak sınıflandırıcının performansını, dolayısıyla da uygulamanın çalışma performansını etkiler (Kayıkcioglu ve Aydemir, 2010).



Şekil 1. Örüntü tanıma ve makine öğrenmesi genel işlem bileşenleri

Öznelik çıkarma işleminde çok farklı yöntemler kullanılabilirliği için işaretleri temsil eden farklı öznelikler elde edilebilmektedir. Bazı durumlarda bir işareti temsil eden özneliklerin sayısı yüzlerce hatta binlerce olabilmektedir. Öznelik sayısının yüksek olması hem zaman kaybına neden olmakta, hem de bazı durumlarda uygulamanın doğruluk

performansını düşürmektedir. Bu dezavantajı ortadan kaldırmak ve öznelik sayısını düşürmek için çeşitli yöntemler önerilmiştir. Bu yöntemler çıkarılan öznelikler arasında etkin olanları belirleyerek hem örüntü tanıma uygulamasının sınıflandırma doğruluğunu arttırmakta hem de karar verme süresini hızlandırmaktadır. Ardışıl ileri yönlü öznelik seçme (AİYÖS) ve ardışıl geri yönlü öznelik seçme (AGYÖS) yöntemleri literatürde bu alanda kullanılan yöntemlerin başında gelmektedir (Ververidis ve Kotropoulos, 2008; Li vd., 2014; Mao, 2004). Bu yöntemler eğitim kümesi içerisinde genelde çapraz doğrulama yöntemi ile en yüksek doğrulama (validasyon) başarısını veren öznelikleri belirler. AİYÖS bu işlemi öznelikleri teker teker seçim kümesine ekleyerek belirlerken, AGYÖS ise var olan bütün öznelikler içerisinde etkin olmayan öznelikleri çıkartmak suretiyle gerçekleştirir. Ancak, doğrulama aşamasındaki rastgele seçimden ötürü her koşulda aynı etkin öznelikler belirlenmez. Dolayısıyla test aşamasında elde edilen sınıflandırma doğruluğu (SD) da farklılık gösterir. Reif ve Shafait yaptıkları çalışmada sınıflandırıcının hesaplama süresini azaltmak için AİYÖS yaklaşımını kullanmışlardır. Çalışmalarında bu yaklaşımın sınıflandırma doğruluğunu arttırdığını göstermişlerdir (Reif ve Shafait, 2014). Benzer biçimde Gan ve arkadaşları da AİYÖS yöntemini doğrusal ayırma ayırıcı (DAA), destek vektör makineleri ve k- en yakın komşuluk yöntemleri ile test ederek AİYÖS yönteminin sınıflandırıcı doğruluğunu arttırdığını 5 ayrı öznelik kümesi üzerinde test etmişlerdir (Gan vd., 2014). Başka bir çalışmada ise Mao, AİYÖS ve AGYÖS algoritmalarını birlikte kullanarak etkin özneliklerin seçileceğini hibrit bir model ile göstermiştir (Mao, 2004). Literatürdeki bu ve benzeri çalışmalara karşın, AİYÖS ve AGYÖS yöntemlerinin eğitim aşamasındaki alt-eğitim kümelerinin rastgele seçiminden kaynaklanan problem üzerinde durulmamıştır.

Bu çalışmada *BCI Competition 2005* yarışmasında sunulan Data Set I isimli veri kümesinden Aydemir ve Kayıkcioglu (2011)

çalışmasında önerilen yöntemle öznelikler çıkartılmış; AİYÖS ve AGYÖS algoritmaları kıyaslanarak kararlı, sabit ve etkin özneliklerin belirlenebilmesi için bir yöntem önerilmiştir. Ayrıca, AİYÖS ve AGYÖS yöntemleri etkin özneliklerin seçilmesi kabiliyeti ve hesaplama hızı bakımından da karşılaştırılmıştır. Bulunan sonuçlar; AİYÖS yöntemi ile daha kararlı ve etkin özneliklerin seçilebileceğini ve bu yöntemin AGYÖS'e göre çok daha hızlı olduğunu göstermiştir.

Materyal ve Yöntem

Veri Kümesi Tanıtımı

BCI Competition 2005 Data Set 1 veri kümesi bir epilepsi hastasından yaklaşık bir hafta ara ile eğitim ve test verileri olmak üzere iki oturumda kaydedilen elektrokortigogram (ECoG) işaretlerinden oluşmaktadır. Veriler epilepsi hastasından, sol elinin küçük parmağını (*smif 1*) ya da dilini (*smif 2*) hareket ettirdiğini düşünmesi istendiği sırada 64 elektrot ve 1 kHz'lik örnekleme frekansı ile kaydedilmiştir. Her bir denemenin süresi ise 3 saniyedir. Sunulan veri kümesi ilk oturumda 278 (eğitim kümesi) ikinci oturumda 100 (test kümesi) deneme olmak üzere toplam 378 denemeden oluşmaktadır. Eğitim ve test verileri yaklaşık 1 hafta ara ile kaydedilmiştir. Bu kümelerin sınıflara göre dağılımı Tablo 1'de verilmiştir. Veri kümesi hakkında daha detaylı bilgi (Lal vd., 2004; Aydemir ve Kayıkçıoğlu, 2010;) çalışmalarında bulunabilir.

Tablo 1. Deneme sayılarının sınıflara göre dağılımı

Sınıf	Eğitim Kümesi	Test Kümesi
SINIF 1	139	50
SINIF 2	139	50
Toplam	278	100

Öznelik Çıkarma ve Seçme

Öznelikler, *BCI Competition 2005 Data Set 1* veri kümesinden Aydemir ve Kayıkçıoğlu'nun (2011) çalışmasında önerilen sürekli dalgacık dönüşümü (SDD) esaslı yöntem ile çıkarılmıştır. SDD, genel olarak daraltılıp, genişletilebilen bir dalgacık fonksiyonu $\psi(t)$ ile analiz edilecek bir $x(t)$ işaretinin evrişimi ile tanımlanır. Eşitlik 1

dalgacık fonksiyonunu, Eşitlik 2 ise evrişim işlemini ifade etmektedir.

$$\psi_{\tau,s} = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \quad (1)$$

$$SDDK_x^\psi(\tau,s) = \frac{1}{\sqrt{|s|}} \int x(t) \psi^*\left(\frac{t-\tau}{s}\right) dt \quad (2)$$

Bu eşitliklerde t , τ ve s sırasıyla zaman, öteleme ve ölçek parametreleridir. SDDK ise sürekli dalgacık dönüşümü katsayılarını ifade etmektedir. SDD 64 elektrottan kaydedilen işaretlere Morlet dalgacığı kullanılarak uygulanmış, her bir deneme için hesaplanan SDDK'lerin hepsi tek bir vektör yapılarak bu vektörün ortalama ve standart sapma değerleri hesaplanarak öznelikler çıkarılmıştır. Böylece bir ECoG denemesi için $64 \times 2 = 128$ öznelik çıkarılmıştır. Burada 64 elektrot sayısı, 2 ise her bir denemeden çıkarılan öznelik sayısıdır.

Bu çalışmada kolay ve hızlı uygulanabilmesi açısında DAA sınıflandırıcısı kullanılmıştır. Bu yöntem, önceden belirlenmiş iki veya daha fazla sınıfın ortalama niteliklerinin istatistiksel olarak anlamlı bir farklılık gösterip göstermediğini ve grupların farklılıklarını ayırt etmede en fazla katkıyı hangi değişkenlerin yaptığını test etmekte kullanılan bir sınıflandırma tekniğidir. DAA, mevcut verilerin bileşkesi olan yeni bir değişken üretir ve ilk tanımlanan sınıfların yeni değişkene göre aralarındaki farklılıklar maksimumlaştırır. Ayrıca DAA, bir bağımlı değişkenin diğer özelliklerin veya kıstasların doğrusal bileşeni olarak ifade edildiği yöntemler olan varyans analizi ve regresyon analizi ile de yakından ilgilidir. Fakat bu iki yöntemde bağımlı değişken, sayısal bir büyüklük olmasına rağmen, DAA'da bu sınıf etiketi gibi kategorik bir değişkendir (Aydemir, 2013).

Sonuçlar ve Tartışma

AİYÖS ve AGYÖS yaklaşımlarının kararlılıklarını ve performanslarını test etmek için ilgili öznelik seçme işlemi eğitim kümesine 1000 kez uygulanmıştır. Bu işlem sonucunda özneliklerin seçilme sayısı Şekil 2 ve Şekil

3'te gösterilmektedir. Şekil 2'de görüldüğü gibi AİYÖS yönteminde birçok öznelik 50 defadan az sayıda ve sadece 3 öznelik 300 defadan fazla seçilmesine karşın, AGYÖS yönteminde birçok öznelik 300 defadan fazla sayıda seçilmiştir. Bu sonuçlardan da anlaşıldığı üzere bu öznelik seçme algoritmaları her koşulda farklı öznelikleri seçmektedir. Doğal olarak da bu durum test kümesinde elde edilen sınıflandırma başarımını etkilemektedir. Tablo 2'de 1000 koşul için elde edilen test kümesine ait ortalama SD, standart sapma ve en küçük ve en büyük sınıflandırma doğrulukları verilmiştir. SD'lerin hesaplanması doğru tahmin edilen sınıf sayısının, toplam sınıf sayısına bölünüp, 100 ile çarpılmasıyla % olarak hesaplanmıştır.

Tablo 2. Belirlenen özneliklerle test kümesine ait sınıflandırma doğrulukları

	AİYÖS (%)	AGYÖS (%)
Ortalama SD ve standart sapma	85.27± 4.56	69.06 ± 3.95
En büyük SD	92	79
En küçük SD	72	61

Tablo 2'de verilen sonuçlardan da görüldüğü gibi test kümesi için elde edilen SD, AİYÖS için %72 ile %92, AGYÖS için %61 ile %79 arasında değişmektedir ve sabit bir değerde değildir. Bu dezavantajı ortadan kaldırmak için Eşitlik 3'te verildiği gibi en yüksek sayıda seçilen öznelikle beraber, bunun %70.71'i kadar sayıda seçilen öznelikler etkin öznelikler olarak belirlenmesi önerilmiştir.

$$F \geq \text{round}\left(\max(D)x \frac{70.71}{100}\right) \quad (3)$$

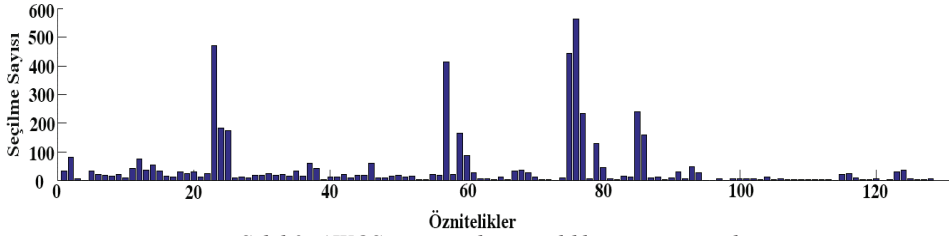
Bu eşitlikte, $\max(D)$; bütün öznelikler içinden en yüksek sayıda seçim sayısını, F ; bu en yüksek sayının en az %70.71'i kadar sayıda seçilen bütün öznelikleri, round ise en yakın tam sayıya yuvarlama fonksiyonunu ifade etmektedir. %70.71 eşiği elektronik filtrelerde, kesim frekansı olarak tepe değerinin $1/\sqrt{2} = 0.7071$ katı olan noktanın belirlenmesinden esinlenerek önerilmiştir (Yavuz ve Aydemir,

2016). Buna göre AİYÖS yönteminde en yüksek sayıda seçilen öznelik 76. öznelik olup 535 defa seçilmiştir. Bunun referanslığında $378 (\text{round}(535x0.7071)=378)$ defadan fazla seçilen öznelikler (23, 57, 75 ve 76. özneliklerdir) etkin öznelikler olarak belirlenmiştir. AGYÖS yönteminde ise en yüksek sayıda seçilen öznelik 59. öznelik olup 995 defa seçilmiştir. Bunun referanslığında $704 (\text{round}(995x0.7071)=704)$ defadan fazla seçilen öznelikler (55 adet) etkin öznelikler olarak belirlenmiştir. Öznelik seçme işlemi olmadan bütün özneliklerin kullanılması durumunda ve AİYÖS ve AGYÖS için belirlenen etkin öznelikler ile test veri kümesi üzerinde elde edilen sınıflandırma doğrulukları Tablo 3'te verilmiştir. Bu sonuçlar AİYÖS yöntemi ile etkin özneliklerin kesin olarak seçilebileceğini ve seçilen özneliklerle elde edilen test sınıflandırma doğruluğunun bütün özneliklerle elde edilene göre %24, Tablo 2'de verilen ortalama sınıflandırma doğruluğuna göre %3.73 daha başarılı sonuç verdiğini göstermiştir. AGYÖS yaklaşımının ise etkin öznelikleri seçmede AİYÖS'e göre %22 daha düşük olduğu görülmektedir.

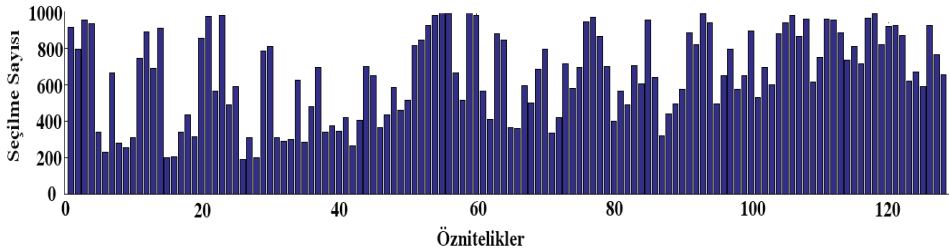
Tablo 3. Etkin öznelikler ile elde edilen test sınıflandırma doğruluğu

	Bütün öznelikler	AİYÖS ile belirlenen öznelikler	AGYÖS ile belirlenen öznelikler
SD	65	89	67

AİYÖS ve AGYÖS yaklaşımları hesaplama karmaşıklığı bakımından da karşılaştırılmıştır. Eğitim kümesinde tek bir koşul için AİYÖS'ün öznelikleri belirleme süresi 2.53 saniye iken AGYÖS için bu süre 102.25 saniyedir. Bütün hesaplamalar Matlab R2015b yazılımı ve Intel Core i7-3770 CPU 3.40 GHz'lik masaüstü bilgisayar ile yapılmıştır. Bu sonuçlara göre de AİYÖS yönteminin AGYÖS'e göre yaklaşık 40 kat daha hızlı olduğu görülmektedir.



Şekil 2. AIYOS yöntemi ile özneliklerin seçim sayıları



Şekil 3. AGYÖS yöntemi ile özneliklerin seçim sayılar

Teşekkür

Bu çalışma 215E155 proje numarası ile TÜBİTAK tarafından desteklenmiştir.

Kaynaklar

- Aydemir, Ö., (2013). İmlecın İki Boyutlu Hareketinin Hayali Sırasında Kaydedilmiş EEG İşaretlerinin Karar Ağaç Yapısı Esaslı Sınıflandırılması, Doktora Tezi, Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü.
- Aydemir, O. ve Kayıkcıoğlu, T., (2011). Wavelet transform based classification of invasive brain computer interface data, *Radioengineering*, 20:1, 31-38.
- Aydemir, O., Kayıkcıoğlu, T., (2010). Classifying ECoG based mental tasks using wavelet transform features”, In 33rd International Conference on Telecommunications and Signal Processing. Baden near Vienna (Austria), p. 103-107.
- de la Fraga, L. G., Coello, C. A. C., (2011). A review of applications of evolutionary algorithms in pattern recognition In *Pattern Recognition, Machine Intelligence and Biometrics* Springer Berlin Heidelberg, 3-28.

- Gan, J. Q., Hasan, B. A. S., ve Tsui, C. S. L., (2014). A filter-dominating hybrid sequential forward floating search method for feature subset selection in high-dimensional space, *International Journal of Machine Learning and Cybernetics*, 5(3), 413-423.
- Jain, A. K., Duin, R. P., Mao, J., (2000). Statistical pattern recognition: A review, *Pattern Analysis and Machine Intelligence*, 22:1, 4-37.
- Kayıkcıoğlu, T., Aydemir, O., (2010). A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data, *Pattern Recognition Letters*, 31:11, 1207-1215.
- Lal, T. N., Hinterberger, T., Widman, G., Schröder, M., Hill, N. J., Rosenstiel, W., Schölkopf, B., (2004). Methods towards invasive human brain computer interfaces, In *Advances in neural information processing systems*, 737-744.
- Li, L., Yu, S., Xiao, W., Li, Y., Li, M., Huang, L., Yang, H., (2014). Prediction of bacterial protein subcellular localization by incorporating various features into Chou's PseAAC and a backward feature selection approach”, *Biochimie*, 104, 100-107.
- Mao, K. Z., (2004). Orthogonal forward selection and backward elimination algorithms for feature subset selection. *Systems, IEEE Transactions on Man, and Cybernetics, Part B: Cybernetics*, 34(1), 629-634.

- Reif, M., ve Shafait, F., (2014). Efficient feature size reduction via predictive forward selection, *Pattern Recognition*, 47(4), 1664-1673.
- Xian, G., Xian, G., (2013). Speaker-independent Recognition by Using Mel Frequency Cepstrum Coefficient and Multi-dimensional Space Bionic Pattern Recognition”, *Advances in Information Sciences and Service Sciences*, 5:1, 221.
- Ververidis, D., Kotropoulos, C., (2008). Fast and accurate sequential floating forward feature selection with the Bayes classifier applied to speech emotion recognition, *Signal Processing*, 88(12), 2956-2970.
- Yavuz E., Aydemir, Ö., (2016). EEG Tabanlı Beyin Bilgisayar Arayüzü İşaretlerinin Etkin Zaman Dilimlerinden Çıkarılmış Özniteliklerle Sınıflandırma Doğruluklarının Artırılması, 20. Biyomedikal Mühendisliği Ulusal Toplantısı, 3-5 Kasım 2016.

Determining the effective features in sequential forward feature selection algorithm

Extended abstract

In recent years, pattern recognition and machine learning has become a very active research area due to the applicability in various kind of subjects, including brain computer interface, commercial and financial approaches. Furthermore, it is worthwhile mentioning that there is no definite technique to solve any classification problem. Thus, it can be helpful to get the most discriminative features using feature selection algorithm. Such approaches are generally performed in four steps: 1- preprocessing, 2- feature extraction, 3- feature selection and 4- classification/regression. Among these steps, feature extraction and feature selection are very vital for representing input signals in a reduced feature space and for identifying discriminative information in order to propose fast and high performance application.

Because of there are many kind of feature extraction method, in some cases hundreds of features might be calculated for identifying input signals. However, enormous number of features might reduce the decision performance and speed which are very crucial two parameters in pattern recognition and machine learning approaches. In order to eliminate those disadvantages and reduce the number of features some techniques have been proposed. Those techniques are used by machine learning community for selecting the most suitable feature subset among all the extracted feature sets in order to increase the performance of the proposed model.

The sequential forward feature selection (SFFS) and the sequential backward feature selection (SBFS) algorithms are very widely used feature selection techniques in literature. In this study, we extracted Continuous Wavelet Transform based features from BCI Competition 2005 Data Set I. Afterwards, among the extracted features, the most stable and effective features were selected by SFFS and SBFS techniques. BCI Competition 2005 Data Set I includes electrocorticogram (ECoG) based brain computer interface signals which was taken from an epilepsy subject on two different days with about one week of delay. In the both sessions the ECoG signals were recorded while subject was asked to imagine of either the left small finger or the tongue movement.

The signals were acquired with an 8x8 ECoG platinum electrode grid (totally from 64 points) which was placed on the contralateral (right) motor cortex. All recordings were performed with a sampling rate of 1 kHz (acquired 3000 samples per channel for every trial). Additionally, BCI Competition 2005 Data Set I consist of 278 training trials (139 trials for finger movements, 139 trials for tongue movements) and 100 test trials which were recorded in the first session and the second session, respectively. Each trial's duration was 3 seconds.

In this paper, SFFS and SBFS algorithms tested for determining of effective features after feature extraction procedure. Afterwards, they compared in terms of classification accuracy and speed. While those methods determine the effective features from training data set using cross validation method, the sub-training data sets are selected randomly. So that, different features might be selected for every running of those methods. Thus, selecting different features are influenced the test performance of the model positively/negatively, as well. Moreover, in this paper a method is proposed to overcome this randomly selection disadvantage. In order to show the robustness of the proposed method the SFFS and SBFS algorithms were run 1000 times in the training stage. Afterwards the features, which were selected more than the determining threshold level, were selected as effective features. Moreover, SFFS and SBFS algorithms were compared in terms of the speed and classification accuracy. The obtained results showed that, SFFS is approximately 40 times faster than SBFS and SFFS provides more than 22% classification accuracy.

Keywords: feature selection, sequential forward feature selection, sequential backward feature selection, classification accuracy, calculation speed.