1st International Mediterranean Science and Engineering Congress (IMSEC 2016)
Çukurova University, Congress Center, October 26-28, 2016, Adana / TURKEY
Pages: 1055-1061, Paper ID:309

SAA proteinlerinin yapısal özelliklerinin makina öğrenme metotları ile ortaya konulması

Turgay İbrikçi[1], İrem Ersöz Kaya[2]

[1]Çukurova Üniversite, Mühendislik-Mim. Fakültesi, Elektrik-Elektronik Müh. Böl. Adana, TÜRKİYE; ibrikci@cu.edu.tr
[2]Mersin Üniversitesi, Tarsus Tek. Fakültesi, Yazılım Müh. Böl. Tarsus/Mersin, TÜRKİYE

Özet

Ailevi Akdeniz Ateşi (FMF) tam adıyla tanımlanan genetik bir hastalıktır. Ancak, onun belirtileri diğer bazı hastalıklarla özellikle romatizmal hastalıklar ile karıştırılabilir. Yaklaşık ailevi Akdeniz ateşi olan 1000 kişi başına 1 ila 3 kişilik arasında değişen Türkiye'de, günlük yaşam üzerindeki oldukça etkili genetik bir hastalıktır. Bu hastalığın ana belirtilerinden biri, belirli organlarda serum amiloid A (SAA) proteinlerinin birikmesidir. Bu çalışmada, araştırma biyoinformatik destekli makine öğrenme yöntemleri kullanılarak insan vücudunda biriken dört SAA proteinlerin ortak biyolojik yapılarını sergilemek amacıyla yapılmıştır. Her şeyden önce, birincil ve ikincil yapıların ortak yanları, sonra parametreleri ortaya çalışılmıştır. Bu nedenle, SAA proteinlerin ortak yapıları tanımlanması için sonuçlar, biyoinformatik destekli makine öğrenme yöntemleri sayesinde kolay ve kullanılabilir bir şekilde elde edilmiştir.

Anahtar Kelimeler: Ailevi Akdeniz Ateşi-AAA, SAA proteinleri, Makine Öğrenmesi, Biyoinformatik

Revealing the structural features of SAA proteins with machine learning methods

Abstract

Familial Mediterranean Fever (FMF) is a genetic disease that is fully defined with the name. However, it can be confused with some other diseases especially rheumatic diseases because of its symptoms. In Turkey, approximately ranging from 1 to 3 persons per 1000 people with familial Mediterranean fever is a genetic disease that affects on the daily life of quite. One of the basic symptoms of this disease is accumulation of Serum Amyloid A (SAA) proteins on certain organs. In this study, a research was performed to exhibit the common biological structures of four SAA proteins deposited in human body by using bioinformatics-assisted machine learning methods. First of all, the common sides of primary and secondary structures, then the parameters were tried to reveal. Therefore, the results on defining the common structures of SAA proteins were obtained in an easy and robust way owing to the bioinformatics-assisted machine learning methods.

Keywords: Familial Mediterranean Fever-FMF, Proteins of SAA proteins, Machine Learning, Bioinformatics

1. GİRİŞ

Ailevi Akdeniz Ateşi (AAA) hastanın günlük yaşamını etkileyerek, yaşam kalitesini düşürecek kadar etkili olan bu genetik hastalık, ülkemizde de önemini korumaktadır. Bölgesel olarak Akdeniz ülkelerinde-Türkiye, İsrail- ve Ermenistan'da sıkça görülmektedir. Ülkemizde de 1000 kişide 1 – 3 hasta gibi yüksek oranda, özellikle doğu ve güneydoğu Anadolu'da bölgelerinde görülmektedir [1-3].

AAA'nın Türkiye'deki ilk teşhisi, 1946 yılında "Garip Bir Karın Ağrısı Sendromu" isimli makale ile Samuel Abrevaya Marmaralı aynı zaman da siyasetçi olan bir tıp doktoru tarafından ortaya konulmuştur [4]. Bu hastalığın genel belirtileri diğer hastalıklar ile çok sık karışabilmektedir. Belirgin belirtilerinden biri de, böbrek ve karaciğer gibi bazı organlarda erimez amyloid proteinlerin birikip, vücut dışına atılamamasıdır. Atılamayan SAA proteinlerinin yapısal ortak özelliklerinin makina öğrenme ile ortaya konulması çalışmanın ana noktasıdır.

Günümüzde çok yaygın olarak kullanılan makina öğrenmesi, bilgisayarların açıkça kodlanmadığı halde mevcut bilgileri öğrenerek yeni bilgiler üretmek yoluyla kendi başlarına hareket etmelerini sağlamak üzerine çalışmalar yapan bilim dalıdır. Öğrenme yeteneğini kazandırmak amacıyla geliştirilen algoritmalar aracılığıyla bilgisayarlar gerçek dünya problemlerinin çözümünde kullanılabilmektedir. Yaşamın her alanında bulunan, insanın iş yükünü azaltmak ve insan hatasını en aza indirmek için kullanılan makinelerin, daha akıllı hareket edebilmesini sağlayabilmek için bilimsel çalışmalarda hızla yaygınlaşmaktadır. Makine öğrenmesi yöntemleri, sağlık ve biyoinformatik alanlarında da yoğun olarak kullanılmaktadır [5]. Bu alanlarda, kanser teşhisi [6], tümör bölgelerinin tespiti [7], ilaç yapımı [8], protein yapılarının tahmini [9] gibi birçok önemli konuda çalışmalar yapılmaktadır.

Bu çalışmada, SAA proteinlerinin birincil yapıları kullanılarak, ikincil yapılarının ortaya konulması için MLP (Multi Layer Perceptron-Çok Katmanlı Algılayıcı), SMO (Sequential Minimal Optimization – Sıralı Minimal Optimizasyon), RBF (Radyal Tabanlı Fonksiyon), Logistic (Lojistik) ve Naive Bayes algoritmaları tercih edilmiştir. Proteinlerin düzensiz yapıları ve ortak motifin ortaya konulması için ise web sunucularından faydalanılmıştır.

2. VERİLER VE METOTLAR

Seçilen SAA proteinlerin ortak özelliği, insanda (human) bulunmaları ve AAA hastalarında etkin olarak görülmeleridir. Bu proteinler, birçok protein veri tabanlarında farklı isimler altında anılmaktadır. Tablo 1. ile bu proteinlerin farklı isimleri ve tanımlamaları protein veri setlerine göre listelenmiştir.

Tablo 1. Kullanılan SAA proteinlerinin farklı veri tabanlarına göre isimlendirilmeleri

Sembol	İsim	Entraz	HUGO	OMIM	RefSeq	UniProt	GenPept
SAA1	HumanSerum amyloid A1	6288	10513	104750	NM_199161	P0DJI8	NP_001171477
SAA2	HumanSerum amyloid A2	6289	10514	104751	NM_030754	P02735	NP_035444
SAA3	HumanSerum amyloid A3 pseudogene	6290	10515	--	--	P22614	NP_035445
SAA4	HumanSerum amyloid A4	6291	10516	104752	NM_006512	P35542	NP_035446

Bu proteinlerin birincil yapıları Tablo 2'de gösterilmiştir. Bu veriler, 1971 yılından beri faaliyette bulunan Protein Data Bank (PDB)'den "fasta" formatında indirilmiştir [10]. Bu SAA proteinlerinin, farklı algoritmalar uygulanarak ikincil yapıları ve diğer özellikleri ortaya konulmuştur.

Bu çalışmada SAA proteinlerinin ikincil yapılarının tahmini için tercih edilen MLP, SMO RBF, Logistic ve Naive Bayes algoritmaları makina öğrenme çalışmalarında çok sıkça kullanılan algoritmalardır. Algoritmalar, MATLAB[11] ve WEKA[12] ortamları üzerinde uygulanarak SAA'ların ikincil yapılarının belirlenmesi sağlanmıştır.

Multi Layer Perceptron (MLP): Eğitim aşamasında, girdileri kullanarak çıktıyı oluşturmaya çalışan, çok katmanlı eğitmenli bir makine öğrenmesi yöntemidir. Çıktıyı oluşturabilmek için oluşturulan yapı

giriş, ara (gizli) ve çıkış katmanlarından oluşur. Katmanlar boğum (düğüm) denilen işlem noktaları ile birbirlerine bağlanıp, bu düğümler üzerinden de "ağırlık değerleri" hesaplanmaktadır. Bu ağırlık değerlerinin yenilenerek hesaplanma işlemine "öğrenme" denir.

Tablo 2. SAA Proteinlerinin Birincil Yapılar

	Birincil Yapıları (Sequence)
SAA1	MKLLTGLVFCSLVLGVSSRSFFSFLGEAFDGARDMWRAYSDMREANYIGSDKYFHARGNYDAAKRGPGGVWAAEAISDA-RENIQRFFGHGAEDSLADQAANEWGRSGKDPNHFRPAGLPEKY
SAA2	MKLLTGLVFCSLVLSVSSRSFFSFLGEAFDGARDMWRAYSDMREANYIGSDKYFHARGNYDAAKRGPGGAWAAEVIS-NARENIQRLTGRGAEDSLADQAANKWGRSGRDPNHFRPAGLPEKY
SAA3	MKPSIAIILCILILGVDSQRWVQFMKEAGQGSRDMWRAYSDMKKANWKNSDKYFHARGNYDAARRGPGGAWAAKVIS-DAREAVQKFTGHGAEDSRADQFANEWGRSGKDPNHFRPAGLPKRY
SAA4	MRLFTGIVFCSLVMGVTSESWRSFFKEALQGVGDMGRAYWDIMISNHQNSNRYLYARGNYDAAQRGPGGVWAAKLISRS-RVYLQGLIDCYLFGNSSTVLEDSKSNEKAEEWGRSGKDPDRFRPDGLPKKY

Sıralı Minimum Optimizasyon (SMO): SMO, kuadratik programlama problemi olarak destek vektör makinasının eğitim anında ortaya çıkan hesaplama problemini çözmek için ortaya atılan bir algoritmadır [13, 14]. Çok terimli kabuk (kernel) fonksiyonları kullanarak destek vektör makinalarında "öğrenme"yi gerçekleştirir.

Radyal Tabanlı Fonksiyon (RBF): RBF, çok boyutlu verilerin üç katmanlı bir sisteme uygulanarak sınıflandırmasını sağlayan çok bilinen bir yapay sinir ağı algoritmasıdır. Bu katmanların ilki verilerin alındığı giriş katmanı, ikinci katman ise giriş katmanından gelen değerleri doğrusal olmayan bir yapılanmaya (kümelemeye) dönüştüren katmandır. Son katman, çıkış katmanı ise doğrusal bir katman olarak verilerin sınıflandırmasını gerçekleştirmektedir.

Lojistik: Lojistik algoritması, bir veya daha fazla tanımlayıcı değişkenler ile bir binom sonucu olarak elde edilen güçlü bir istatistiksel yoldur. Bu kümülatif lojistik dağıtım, lojistik işlevini kullanarak olasılıkları tahmin ederek kategorik bağımlı değişken ve bir veya daha fazla bağımsız değişkenler arasındaki ilişkiyi ölçer.

Naive Bayes: Bayes teoreminin basit bir şekilde uygulamasıdır [15]. Bir verinin üyeliğini istatistiksel kestirim yapan Bayes teoriminden, bir takım başlangıç kabuller yapılarak üretilmiştir. Bunların en önemlilerinden biri, kullanılan parametrelerin birbirlerinden bağımsız olmalarıdır.

3. SONUÇLAR

İkincil Yapılarının Belirlenmesi

İkincil yapılar, 20 farklı aminoasitlerin bir dizim ile oluşan birincil yapısının farklı açılar ve bağlar oluşturarak meydana çıkan, proteinlerin ikinci seviye temel yapılarında biridir. Düzenli bir yapıya sahip olup, alfa, beta ve coin yapılarına sahiptirler. Bu yapılar, proteinlerin bölgesel yapılarının üç boyutlu halleridir. Birincil yapıları üzerinden, ikincil yapıları beş farklı makina öğrenme metotları ile bulunulmaya çalışılmıştır.

Tablo 3. Doğruluk değerlerine göre sonuçlar

	SAA1	**SAA2**	**SAA3**	**SAA4**
MLP	60,33	63,27	64,75	58,46
SMO	50,41	60,65	58,19	51,53
RBF	58,67	58,19	59,83	57,69
Logistic	62,02	55,73	60,65	45,38
Navie Bayes	57,85	62,29	68,85	63,84

Metotların başarımını değerlendirmek için "doğruluk" parametresi aşağıdaki formül kullanılarak bulunmuştur.

Düzensiz Yapılarının Ortaya Konulması

Protein yapısının, dört ana hiyerarşik seviyeden oluştuğu bilinmektedir. Sırasıyla amino asitlerin yan yana gelmesiyle birincil yapı, amino asitleri birleştiren bağların bükülüp dönmesiyle ortaya çıkan bölgesel şekillenmelerle ikincil yapı, 3 boyutlu bir katlanmayla üçüncül yapı ve bu yapıların bir araya gel-

mesiyle dördüncül yapı meydana gelir. Proteinin yapısı ile fonksiyonu arasındaki ilişkiyi ifade eden en etkin görüş, 3 boyutlu yapının fonksiyonun gerçekleştirilmesinde ön koşul olarak kabul edilmesidir [16]. Proteinin katlanmamış olduğu durumda fonksiyonel formunda olduğu kabul edilir. Ancak sonrasında yapılan çalışmalar, birçok protein ve protein parçasının (domain), 3 boyutlu yapıya sahip olmadıkları halde çok önemli fonksiyonlarda yer aldıklarını göstermiştir [17, 18]. Böylece Düzensiz Yapılı Proteinler kavramı ortaya çıkmıştır [19].

Düzensiz yapılı proteinler, içerdikleri düzensiz kısımların uzunluklarına göre tipik olarak iki gruba ayrılmaktadırlar. 30'dan fazla rezidü (amino asit) içeren düzensiz bölgeler uzun, ve diğerleri kısa olarak sınıflandırılmaktadır [20].

Tablo 4. SAA Proteinlerinin Düzenli(O)-Düzensiz(D) Yapı Dağılımları

	Öngörülen Düzensiz Kalıntıları
SAA1	000000000000000000000000000000000000000000000000000000000000000000000-00000000000000000000000000000000***DDDDD***000000000000***DDD***
SAA2	000000000000000000000000000000000000000000000000000000000000000000000-000000000000000000***DDDDDDD***000000***DDDDDD***000000000000***DDD***
SAA3	000000000000000000000000000000000000000000000000000000000000000000000-0000000000000000000***DDDD***0***D***0000000***DDDDD***000000000000***DDD***
SAA4	000000000000000000000000000000000000000000000000000000000000000000000-00000000000000000000000000***DDDDD***0***D***0000000***D***0***DDD***000000000000***DDD***

Seçilen SAA proteinlerinin yapılarına bakıldığında düzensiz bölgeler içerdikleri görülmektedir (Tablo 5). Tablo 4'de proteinlerin düzenli (O) ve düzensiz (D) bölgelerinin, sekans üzerinde dağılımı verilmektedir. Düzensiz yapı içeren parçaların, düzensizliğin başladığı amino asit noktasına göre hizalandığındaki (alignment) durumu ise Tablo 5'da verilmektedir. Tablo 6'daki durum incelendiğinde SAA proteinlerinin düzensiz bölgelerinin çoğunlukla aynı sekans bölgelerinde bulunduğu gözlenmektedir. Özellikle son üç düzensiz bölge tüm proteinlerde mevcuttur. Tüm sekans için başlangıç pozisyonları farklı noktalara denk gelse de ilgili bölgelere göre hizalandıklarında sondan 16 ila 18. pozisyonlar arasının ve 20. pozisyonun tüm proteinlerde ortak olarak düzensiz yapıda olduğu görülmektedir. Benzer şekilde, SAA1 protein hariç diğer proteinlerde, sondan 28 ve 30-34 arası amino asitler düzensiz yapıda bulunmaktadırlar.

Düzensiz bölgelerin uzunluğu 3 ile 8 amino asit olarak değişmektedir. Buna göre kısa düzensiz bölgelere sahip oldukları söylenebilir.

Tablo 5. Hizalanmış Protein Parçalarında Düzenli(O)-Düzensiz(D) Yapı Dağılımları

	Düzensiz Bölgeler	Başlama Pozisyonu
SAA1	00000000000000**DDDDD**000000000000**DDD**	103
SAA2	**DDDDDDD**000000**DDDDDD**000000000000**DDD**	89
SAA3	0**DDDD**0**D**0000000**DDDDD**000000000000**DDD**	90
SAA4	**DDDDD**0**D**0000000**D**0**DDD**000000000000**DDD**	97

Ortak Motif Yapılarının Ortaya Konulması

Motif, protein yapısı içinde bilinen en küçük birim olarak tanımlanan ve bir protein yapısı içindeki bir desendir. Bilinen motifler ile karşılaştırarak, proteinin nasıl bir fonksiyona sahip olduğu ortaya konulmaktadır. Bunun için kullanılan veri tabanlarından bazıları PROSITES, PRINTS, SMART ve Pfam'dır [21].

Çalışmada, dört SAA proteininin yapılarında ortak bir motif taşıdıkları belirlenmiştir. Bu SAA motifinin, farklı veri tabanlarındaki isimleri ve kodları aşağıda listelenmiştir;

SMART - SM00197 (SAA)

PRINTS - PR00306 (SERUMAMYLOID)

Pfam - PF00277 (SAA)

PROSITE - PS00992 (SAA)

PIRSF - PIRSF002472 (Serum_amyloid_A) [22]

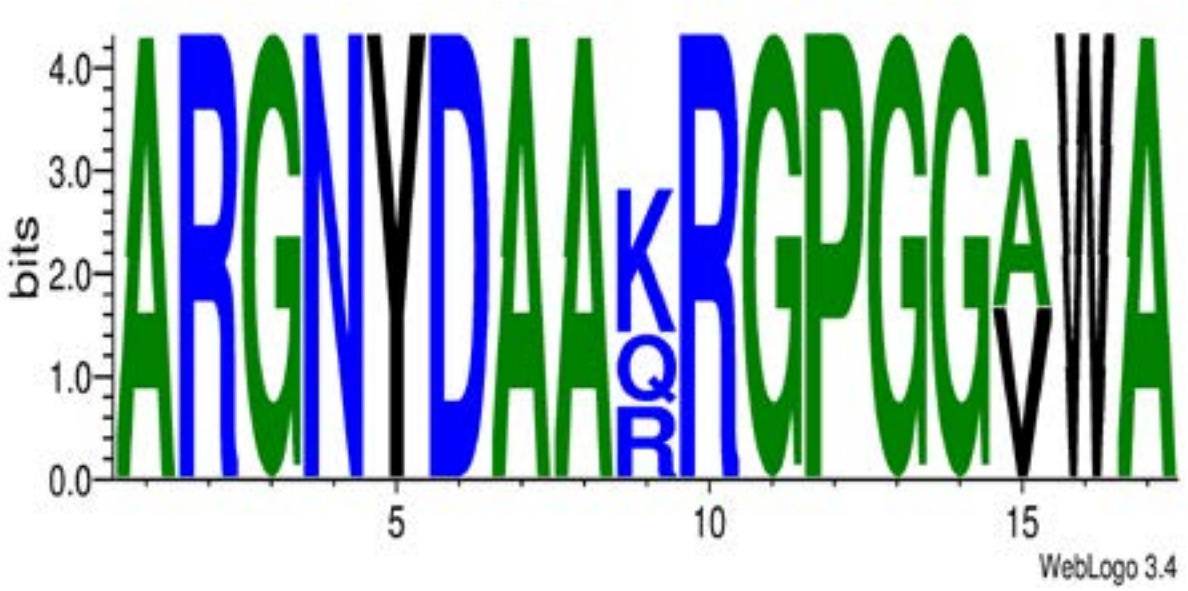
Bulunan SAA motifinin proteinlerdeki pozisyonları, Tablo 6'da verilen amino asit dizilerinde koyu renkle işaretlenerek gösterilmiştir.

Tablo 6. Mevcut motifin protein yapısı içindeki pozisyon

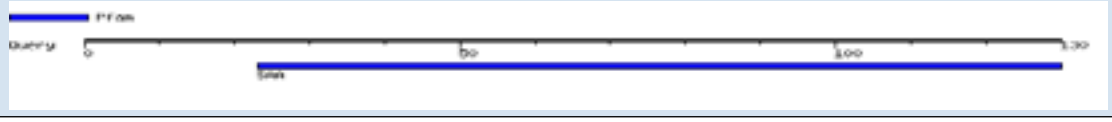
ID	
SAA1	MKLLTGLVFCSLVLGVSSRSFFSFLGEAFDGARDMWRAYSDMREANYIGSDKYFH**ARGNYDAAKRGPGGVWA**AE-AISDARENIQRFFGHGAEDSLADQAANEWGRSGKDPNHFRPAGLPEKY
SAA2	MKLLTGLVFCSLVLSVSSRSFFSFLGEAFDGARDMWRAYSDMREANYIGSDKYFH**ARGNYDAAKRGPGGAWA**AE-VISNARENIQRLTGRGAEDSLADQAANKWGRSGRDPNHFRPAGLPEKY
SAA3	MKPSIAIILCILILGVDSQRWVQFMKEAGQGSRDMWRAYSDMKKANWKNSDKYFH**ARGNYDAARRGPGGAWA**-AKVISDAREAVQKFTGHGAEDSRADQFANEWGRSGKDPNHFRPAGLPKRY
SAA4	MRLFTGIVFCSLVMGVTSESWRSFFKEALQGVGDMGRAYWDIMISNHQNSNRYLY**ARGNYDAAQRGPGGVWA**AK-LISRSRVYLQGLIDCYLFGNSSTVLEDSKSNEKAEEWGRSGKDPDRFRPDGLPKKY

Tablo 7. Proteinlerde bulunan ortak motif ve pozisyonu [23]

ID	Bulunan Motif	Pozisyonu	Motif Adı
SAA1	ARGNYDAaKRGpGGvWA	56-72	PS000992|SAA
SAA2	ARGNYDAaKRGpGGaWA	56-72	PS000992|SAA
SAA3	ARGNYDAaRRGpGGaWA	56-72	PS000992|SAA
SAA4	ARGNYDAaQRGpGGvWA	56-72	PS000992|SAA



Şekil 1. Dört proteinde oluşan SAA motifi [24]

Pfam	Motifler
SAA1	
SAA2	
SAA3	
SAA4	

Şekil 2. Proteinlerde var olan SAA motifin yerleşimi

WebLogo 3.4 web sunucusu aracılığı ile oluşturulan motif, dört proteininin ortak motifi olarak belirgin bir SAA motifini göstermektedir. Bu motifi bulmak için daha önceden ortaya konulmuş motif veri setlerinde araştırma yapılmıştır ve şekil 1.de verilen motif bulunmuştur. Şekil 2 ise Pfam veri setine göre motiflerin protein içindeki bulunabilecekleri pozisyonları göstermektedir. Tablo 7'de SAA proteinlerindeki motiflerin, başlangıç değerleri verilmiştir.

5. TARTIŞMA

Bu çalışmada, AAA genetik hastalığında belirgin belirtilerden biri olan SAA amoloid proteinlerinden SAA1, SAA2, SAA ve SAA4 üzerinde çalışılmış, bunların ortak biyolojik yapıları ortaya konulmaya çalışılmıştır. Aynı zamanda ortak düzensiz bölgeler tespit edilmiş ve ortak bir motif ortaya konulmuştur. Bu motifin proteinlerin hangi bölgesinde aktif olduğu belirlenmiştir. Seçilen bu proteinlerin öncelikli ortak yönleri ise "insan" proteinleri olmalarıdır.

Sonuçlar göstermiştir ki, bu proteinler arasında yapısal olarak yüksek oranda benzerlikler bulunmaktadır. Ortaya koyulan benzerliklerin, bu tip proteinlerin, AAA hastalarında, vücutta belli organlar üzerinde tutularak o organın özelliklerini kaybetmesine neden olması ile ilişkilendirilmeleri muhtemeldir. Böylece, genetik bir hastalık olarak AAA'nın ortadan kaldırılması olanak dışı olmasına rağmen, belirtilerinin baskılanmasını sağlayabilen uygulamalar ve aynı zamanda SAA proteinlerin hasta üzerinde birikmesini geciktirebilecek veya vücut dışına atılmasına yardımcı olabilecek çalışmalar için, yapılan bu araştırmanın katkısı olabileceği düşünülmektedir.

Bundan sonraki çalışmalarda, diğer SAA proteinleri de kullanarak genel SAA yapılarının ortaya konulmaya çalışılacaktır.

6. TEŞEKKÜR

Bu çalışma, Çukurova Üniversitesi Bireysel Araştırma Projeleri (BAP) birimi tarafından Proje No: MMF2013BAP19 ile desteklenmiştir.

REFERANSLAR

[1] Tunca M. (2006), "Ailevi Akdeniz Ateşinin Tarihçesi ve Türkiye'dek Ailevi Akdeniz Ateşi", *Türkiye Klinikleri Dergisi*, 2(8):4-8.

[2] Örün, E., Yalçınkaya, F. (2003), "Türk Tıbbında Ailevi Akdeniz Ateşi Hastalığı ve Amiloidoz", *Türk Nefroloji Diyaliz ve Transplantasyon Dergisi*; 12(1):1-7.

[3] Tunca M. (2005), "Familial Mediterranean Fever (FMF) in Turkey: Results of A Nationwide Multicenter Study", *Medicine*, 84(1):1-11.

[4] Marmaralı A. S. (1946), "Garip Bir Karın Ağrısı Sendromu", Türkiye Tıp Mec. No. 12.

[5] Cohen, J. (2004), "Bioinformatics – An Introduction for Computer Scientists", *ACM Computing Surveys*, 36(2):122-158.

[6] Ibrikci, T., Üstün, D., Ersoz Kaya, I., (2012), "Diagnosis of Several Diseases by Using a Novel Hybrid Kernel Support Vector Machine", *Journal of Medical Systems*, 36(3): 1831-1840.

[7] Ibrikci,T., Şen, O., (2012), "Comparison of Nonlinear Classication on Brain MRI's" , Ç.Ü. Journal of Science and Engineering, 28(3): 128-137.

[8] Pehlivanlı, A. Ç., Ibrikci,T., Ersoy, O.K., (2008), "Consensual Classification of Drug and Nondrug Compounds", *International Jour. of Computational Biology and Drug Design,* 1(3): 224-234.

[9] Ersöz Kaya, I, Ibrikci, T., Ersoy, O. K., (2011), "Prediction of Disorder with New Computational Tool: BVDEA", *Expert Systems with Applications*. 38(12): 14451-14459.

[10] World Wide Protein Data Bank, www.wwpdb.org/

[11] www.mathworks.com

[12] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H., (2009), "The WEKA Data Mining Software: An Update", *ACM SIGKDD Explorations*, 11(1).

[13] Platt, J. (1998), Fast Training of Support Vector Machines using Sequential Minimal Optimization", In Advances Kernel Methods-Support Vector Learning, B.Scholkopf, C. Burges, A. Smola, eds., MIT Press.

[14] Shevade, S.K., Keerthi, S.S., Bhattacharyya, C., Murthy, K.K. (2000), "Improvements to SMO Algorithm for SVM Regression", *IEEE Trans Neural Netw.* 11 (5):1188-93.

[15] Orhan, U., I. Ersoz, I., Ibrikci,T., (2007), "Prediction of Disordered Regions of Proteins with N-Pieces Naive Algorithm", *Intelligent Engineering Systems Through Artificial Neural Networks: Smart System Engineering* (ANNIE'07), 17: 43-48.

[16] Romero, P., Obradovic, Z., Dunker, A.K., (2000), "Intelligent Data Analysis for Protein Disorder Prediction", *Artificial Intelligence Review*, 14:447-484.

[17] Klee, C., Draetta, G., Hubbard, M., (1988), "Calcineurin", In Meister, A. (ed.) Advances in Enzymology, 61:149–200.

[18] Dunker, A.K., Obradovic, Z., Romero, P., Kissinger, C., Villafranca, E., (1997), "On The Importance of Being Disordered", *PDB Newsletter*, 81:3-5.

[19] Dunker, A. K., Lawson, J.D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. S., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., Obradovic, Z., (2001), "Intrinsically Disordered Protein", *Journal of Molecular Graphics and Modelling*, 19:26-59.

[20] Atkins, J.D., Boateng, S. Y., Sorensen, T., McGuffin, L., (2015), "Disorder Prediction Methods, Their Applicability to Different Protein Targets and Their Usefulness for Guiding Experimental Studies", Int. J. Mol. Sci, 16:19040-19054.

[21] PFAM, pfam.xfam.org/

[22] Wu, C.H., Nikolskaya, A., Huang, H., Lai-Su L. Yeh, Natale, D.A., Vinayaka, C. R., Z Hu, Z., Mazumder, R., Kumar, S., Kourtesis,P., Ledley, R.S., Suzek, B. E., Arminski, L., Chen, Y., Zhang, J., Cardenas, J.L., Chung S., Castro-Alvear, J., Dinkov, G., Barker, W.C, (2004) "PIRSF: Family Classification System at The Protein Information Resource", Nucleic Acids Res. 2004 Jan 1; 32

[23] PROSITE, prosite.expasy.org/scanprosite/

[24] WEBLOGO, weblogo.berkeley.edu/logo.cgi