

Grafik Tablet Kullanılarak Makine Öğrenmesi Yardımı ile El Yazısından Cinsiyet Tespiti

Berna Arı^{1**}, Ali Arı², İlknur Ucuz³, Filiz Çiledağ Özdemir⁴, Abdulkadir Şengür¹

¹ Elektrik ve Elektronik Mühendisliği Bölümü, Teknoloji Fakültesi, Fırat Üniversitesi, Elazığ, Türkiye

² Bilgisayar Mühendisliği, Mühendislik Fakültesi, İnönü Üniversitesi, Malatya, Türkiye

³ Çocuk Psikiyatrisi, Tıp Fakültesi, İnönü Üniversitesi, Malatya, Türkiye

⁴ Fizyoterapi ve Rehabilitasyon, Sağlık Bilimleri Fakültesi, İnönü Üniversitesi, Malatya, Türkiye

*¹bernagurler.a@gmail.com, ²ali.ari@inonu.edu.tr, ³ilknur_27@yahoo.com, ⁴fciledag@gmail.com, ¹ksengur@gmail.com

(Geliş/Received: 15/12/2019;

Kabul/Accepted: 13/02/2020)

Öz: Günlük hayatın bir rutini olan el yazısı; ruh hali, kişilik özellikleri ve var olan bazı hastalıklar hakkında ipuçları vermektedir. Bunun yanında adli tıp, tıp ve arkeoloji gibi birçok disiplin el yazısını kendi alanlarına yönelik çalışmalarda sıkça kullanmaktadır. Örneğin adli tıp; bazı vakaların aydınlatılmasında el yazısından yaş aralığı ve hangi elin kullanıldığı gibi bilgilere ulaşabilmektedir. Bu çalışmada, el yazısından cinsiyet tespiti yapan bir sistem önerilmiştir. Önerilen sistem el yazısından, el yazısını karakterize edecek bir dizi öznitelik çıkarıp bu öznitelikleri makine öğrenmesi teknikleri kullanarak cinsiyete göre sınıflandırmıştır. El yazılarının kaydedildiği tabletle hem kalemin tablete temas halindeki durumunda hem de harfler ve kelimeler arası geçişte kalemin havada izlediği eğri hareketlerinden öznitelik çıkarılmıştır. Bu öznitelikler sırası ile kalem hızı, ivmesi, yazarken oluşan sarsıntı hareketleri, eğim açısı, yazıdaki kavislenmeler, kalemin havada kalma oranı, kalemin yaptığı basınç değeri ve kalemin yükseklik açıdır. Sınıflandırıcı olarak da Karar Ağaçları (KA), Naive Bayes (NB), Destek Vektör Makineler (DVM) ve k-en Yakın Komşu (k-EYK) yaklaşımları kullanılmıştır. Deneysel çalışmalarda kullanılan veri setinde toplam 410 örnek mevcut olup, deneysel çalışmaların başarımları doğruluk kriteri ile değerlendirilmiştir. Elde edilen sonuçlara göre en iyi başarımın DVM ile elde edildiği ve doğruluk değerinin de %85,1 olduğu görülmüştür.

Anahtar kelimeler: cinsiyet sınıflandırma, el yazısı tanıma, makine öğrenmesi, kelime bölütleme, grafik tabletle el yazısı alımı

Gender Detection from Handwriting with the Help of Machine Learning Using a Graphic Tablet

Abstract: As a routine of daily life handwriting gives clues about mood, personality traits and some diseases. In addition, many disciplines such as forensic medicine, medicine and archeology often use handwriting in their own fields. For example, forensic medicine can reach information from the handwriting, such as age range and which hand is used in some cases. In this study, a handwriting system was proposed. The proposed system subtracted a series of features from the handwriting that would characterize the handwriting, classifying them according to gender that used in machine learning techniques. With the tablet in which handwriting is recorded, the feature is extracted from the curve movements that the pen follows in the air both in case the pen is in contact with the tablet and in the transition between letters and words. These features are the speed of the pen, acceleration, jerking movements while writing, the angle of inclination, the curves in the writing, the rate of pen stay in the air, the pressure value of the pen and the height angle of the pen. Decision Trees (DT), Naive Bayes (NB), Support Vector Machines (SVM) and k-Nearest Neighbor (k-NN) approaches were used as the classifier. There are 410 samples in the data set used in the experimental studies and the performances of the experimental studies were evaluated with accuracy criterion. According to the results that have been observed, the best performance was obtained with SVM and the accuracy value was 85.1%.

Keywords: gender classification, handwriting recognition, machine learning, word segmentation, getting the handwriting with graphic tablet

1. Giriş

Yazı, icadından günümüze kadar geçen süreçte insanların kendini ifade etmesini sağlayan en etkili araçlardan biridir. Yazının kişilere spesifik olması ve bireysel özellikler taşıması bilimin farklı alanlarında kullanımını gündeme getirmiştir. Bu duruma paralel, el yazısının karakterize edilmesini ve analizini gerçekleştiren grafoloji alanı gelişmiştir. Grafoloji, bir kişinin el yazısı özelliklerine dayalı meslek seçimi, ruh hali, kişilik özellikleri ve sahip olabileceği hastalıklar hakkında fikir vermeyi amaçlayan deneysel bir bilim alanı olarak karşımıza çıkmaktadır [1].

Birçok disiplin, el yazısı özelliklerini kendi sahasında kullanımına yönelik çalışmalar gerçekleştirmişlerdir. Örneğin tıp alanında Parkinson hastalığı, Otizm Spektrum Bozukluğu (OSB), Dikkat Eksikliği ve Hiperaktivite

* Sorumlu yazar: bernagurler.a@gmail.com. Yazarların ORCID Numarası: ¹0000-0003-1000-2619, ²0000-0002-5071-6790, ³0000-0003-1986-4688, ⁴0000-0001-9421-0233, ⁵0000-0003-1614-2639

Bozukluğu (DEHB), nörolojik ve fiziksel birçok hastalıkta tanı koyma aşamasında el yazısı özelliklerinden faydalanılmıştır [2]. El yazısının klinik pratikte sık kullanıldığı alanlardan biri de adli tıptır. Adli tıpta birçok araştırmanın incelenmesi ve aydınlanması aşamasında el yazısı üzerinden kişinin karakteri, yaş aralığı, hangi elini daha fazla kullandığı, ırkı ve daha birçok bilgiye ulaşılabilmektedir. Bir başka alan olan tarihi araştırmalarda da geçmişin değerlendirilmesi ve tarihi belgelerin çözümlenmesi amacıyla cinsiyet, ırk, yaş gibi geçmiş dönemde yaşayan insanlara ait özellikler el yazısı aracılığıyla tespit edilmiştir [3]. El yazısı ile cinsiyet tahmini üzerine yapılan araştırmalarda erkek bireylerin el yazılarının daha “sert”, “aceleci”, “düzensiz”, “dağınık” ve “eğimli” olduğu, kadınların el yazılarının ise daha “belirleyici”, “homojen”, “narin”, “düzenli”, “tutarlı”, “muntazam” ve “büyük” olduğu görülmüştür [4,5]. El yazısına bağlı cinsiyete dayalı farklılıklar erkek ve kadınlardaki motor kontrol farklılıklarıyla ilişkili bulunmuştur [6]. Bu çalışmalar ışığında el yazısı özelliklerinin birçok farklı bilim alanında farklı bireysel özelliklerin saptanabilmesinde kullanılabileceği görülmektedir.

Kişilerin el yazısı ile sahip olduğu kişisel özellikler arasındaki benzerlikler ve farklılıklar gruplandırılarak belirli sonuçlara ulaşılmaya çalışılmış, ancak manuel olarak yapılan değerlendirmeler zaman kaybına ve verilerin objektif olarak değerlendirilememesine neden olmuştur. Günümüzde bilgisayarlı görme sistemleri ile el yazısında daha objektif bir değerlendirme yapılarak el yazısına ait birçok öznelik tespit edilebilmektedir. Tasarlanan sistemlerde harflerin boyutu, harflerin ve kelimelerin eğikliği, taban çizgisi, kalemin basıncı, harflerin ve kelimelerin arasındaki boşluklar gibi el yazısına ait pek çok öznelik saptanabilmektedir [7]. Bu öznelikler sayesinde el yazısı görüntülerinin amaca bağlı olarak sınıflandırılabilmesi sağlanabilmektedir.

Araştırmalarda el yazıları üzerinde; farklı sınıflandırma yöntemleri, farklı veri setleri ve farklı öznelikler kullanılarak cinsiyet sınıflandırma işlemi yapılmıştır. Bu alandaki literatür çalışmalarına baktığımızda; Moetesum ve arkadaşları [8] tarafından, Qatar University Writer Identification (QUWI) veri setinden alınan 1000 örnek ile Evrişimsel Sinir Ağları (ESA) üzerinden öznelik çıkarımı yapılmıştır. Sonrasında NB, KA, Lineer Ayırma Analizi (LAA) ve DVM kullanılmış ve el yazısından cinsiyet tespit edilmiştir. Yapılan çalışmada en iyi sınıflandırıcının %70,1 ile LAA olduğu bildirilmiştir. Gattal ve arkadaşları [3], el yazısından cinsiyet tespit etmede eğitim, düzgün ve düzensiz yazma, sert ve yumuşak yazma gibi ayırt edici öznelikleri dokusal bilgiler olarak tanımlamış ve çıkarımını, yönlendirilmiş Temel Görüntü Öznelikleri (yTGÖ) ile yapmıştır. Erkeklerden ve kadınlardan alınan el yazısı örneklerinden elde edilen yTGÖ histogramları ile DVM sınıflandırıcısı eğitilmiştir. Emran ve arkadaşları [9], el yazısından cinsiyet, yaş ve el tercihi (sağlak-solak) belirlemek için akıllı bir yöntem önermiştir. Öznelik çıkarımı aşamasında yazının eğrilik derecesi, yönü ve eğimi gibi özelliklerin kullanımı tercih edilmiştir. Öznelik çıkarımı için En Küçük Artıklık En Büyük İlgililik algoritması; boyut indirgeme için ise Temel Bileşenler Analizi (TBA) kullanılmıştır. QUWI veri setindeki Arapça ve İngilizce el yazısı metinlerinden alınan görüntüler, bulanık mantık kenar algılama yöntemiyle bölütlenmiştir. El yazılarının sınıflandırılması ve karakterlerinin tanınması, k-EYK, Rastgele Orman (RO) ve DVM ile yapılmıştır. Son adımda, sınıflandırmalar önce Arapça karakterler, sonra İngilizce karakterler daha sonra ise hem Arapça hem İngilizce karakterler için ayrı ayrı gerçekleştirilmiştir. Hem Arapça hem İngilizcede k-EYK, DVM ve RO için cinsiyet sınıflandırma oranları sırasıyla %74,5, %76,4 ve %75,2 olarak bulunmuştur. Morera ve arkadaşlarının [10] çalışmalarında ESA ile öznelik çıkarımı yapılmış ve kişiler hem cinsiyetine hem de el tercihinin göre sınıflandırılmıştır. Sınıflandırmalar, İngilizce metinleri içeren IAM ve Arapça metinleri içeren KHATT olmak üzere iki veri seti üzerinde ayrı ayrı yapılmıştır. Ayrıca cinsiyet ve el tercihinin birleştirilerek farklı bir sınıflandırma kategorisi oluşturulmuştur. Bu kategoriler; kadın-sağ el, kadın-sol el, erkek-sağ el, erkek-sol el şeklinde adlandırılmıştır. El tercihi sınıflandırmasını IAM veri seti için %90,7, KHATT veri seti için %70,9 olarak elde edilmiştir. Cinsiyet sınıflandırması ise IAM veri seti için %80,7 ve KHATT veri seti için %68,9 için bulunmuştur. Ahmed ve arkadaşları [11], Yerel İkili Örnekleme (YİÖ), Yönlendirilmiş Gradyanların Histogramı (YGH), Gri Seviyeli Eş-oluşum Matrisi (GSEM) ile öznelik çıkarımı yapmıştır. Bu öznelikler eğitim, sözcük aralığı, kalem basıncı, gradyan bilgisi olarak verilmiştir. Sınıflandırma için Yapay Sinir Ağları (YSA), DVM, k-EYK, KA ve RO kullanılmıştır. QUWI veri tabanındaki kadın ve erkeklere ait el yazısı örneklerinden yaklaşık 400 adedi alınmıştır. Arapça yazılar üzerinden cinsiyet sınıflandırması yapılmış ve en iyi sınıflandırma başarımı DVM ile %78 oranında elde edilmiştir. Sonra İngilizce yazılarda YSA ile %84 başarılı cinsiyet sınıflandırma yapılmıştır. Sınıflandırmanın eğitim setinde Arapça örnekler, test setinde İngilizce örnekler kullanılarak cinsiyet sınıflandırılmıştır ve en iyi sınıflandırma başarımı YSA ile %79 bulunmuştur. Bir sonrakinde sınıflandırmanın eğitim setinde İngilizce örnekler, test setinde Arapça örnekler kullanılarak cinsiyet sınıflandırılmış ve YSA ile %74 doğrulukla sınıflandırılmıştır. Topaloğlu ve arkadaşlarının [1] yaptığı çalışmada, el yazısı konusunda iki amaç üzerinde durulmuştur. İlki, veri seti oluşturulurken el yazısı örneği alınan her bir kişinin kendi el yazısını tanıyıp tanıyamadığının belirlenmesi iken ikinci amaç el yazısı üzerinden cinsiyet tespiti yapmak olmuştur. Verileri analiz etmede; kâğıda uygulanan basınç, kâğıt üzerindeki sınırlar, harf boyutları, kenar çizgileri, eğim ve yazı için kullanılan alan gibi öznelikler çıkarılmıştır. En az lise mezunu olan 80 kişiden 5 sayfalık metni yazmaları

istenererek veri seti oluşturulmuştur. Katılımcıların kendi el yazılarını tanımaları konusunda düzenlenen ankette yaklaşık %48'inin kendi el yazısını tanıyabildiği sonucuna ulaşılmıştır. Cinsiyet tespiti için de ID3 ve J48 (C5.0) algoritmalarını kullanılmıştır. Kullanılan yöntemler sonucunda, J48 algoritması cinsiyet tespitinde %70,8 başarılı bulunmuşken ID3 algoritması ile de %93,7 sınıflandırma başarımı elde edilmiştir. Bir diğer el yazısından cinsiyet tespiti çalışmasını Guerbai ve arkadaşları [12] yapmıştır. Öznitelik çıkarımını Curvelet dönüşüm ile yapmış ve 44 bileşenli bir öznitelik vektörü oluşturulmuştur. Sistemleri, Bir Sınıf Destek Vektör Makinesi (BS-DVM) kullanarak sınıflandırılmıştır. Bunun için IAM veri seti içerisinde 200'den fazla örneği kullanılmıştır. Önerilen sistemin başarımını artırmak için, birden fazla BS-DVM sınıflandırıcısı birleştirilmiş ve yaklaşık %77 başarımla elde edilmiştir. Akbari ve arkadaşlarının çalışmasında [13], kişilerin el yazısından cinsiyetlerinin tespiti QUWI veri setindeki 475 kişiden [14] ve Multi-script El Yazısı Veri setindeki (MSEV) [15] 84 kişiden alınan el yazısı örnekleriyle yapılmıştır. Her bir yazı örneğinin görüntüleri dalgacık alt bantlarına ayrıştırılmış ve doku blokları (her bir yazı, doku blokları olarak göz önüne alınmıştır) üretmek için kullanılmıştır. Daha sonra her bir alt bant, Olasılıklı Sonlu Durum Otomati (OSDO) oluşturmak için sembolize edilen bir veri dizisine genişletilmiştir. OSDO sonrasında, cinsiyet ayırımı yapmak için kullanılacak öznitelik vektörünü üretmek için kullanılmıştır. Cinsiyet sınıflandırma, YSA ve DVM ile yapılmış ve iki veri seti üzerinde yaklaşık %80 oranında başarımla sağlanmıştır. Mirza ve arkadaşlarının [16] çalışması bireylerin cinsiyet tespitinin yanında demografik özelliklerinin tespitini de kapsamaktadır. Kadınların ve erkeklerin görsel olarak ayrı bir yapıda yazdığı (eğim, büyük ve küçük, düzenli ve düzensiz gibi) hipoteziyle, Gabor Filtre kullanılmış ve 96 bileşenli bir öznitelik vektörü elde edilmiştir. Filtreleme sonucu ortalama ve standart sapma değerleri bir matriste toplanmış ve bu matrisin Fourier dönüşümü bir öznitelik olarak kullanılmıştır. Sınıflandırma QUWI veri setindeki 400 kişiden alınan örneklerle ileri beslemeli bir sinir ağı kullanılarak gerçekleştirilmiş ve %70 başarımla sağlanmıştır. Marzinotto ve arkadaşlarının [17] yaptığı çalışmada, yaş ve cinsiyetin çevrimiçi (tablet gibi bir dijitalleştiriciden alınan yazı) olarak alınan el yazılarının karakterize edilmiştir. Çalışmada iki katmanlı bir kümeleme yaklaşımı önerilmiş ve katmanlar ayrı ayrı IRONOFF veri setine göre incelenmiştir [18]. Sonrasında k-Ortalamalar algoritması ile kişilerin tüm kelimeleri kümelendirilmiştir. Bulunan kümeleri anlamlandırmak için, her kümede, küme merkezine en yakın kelimeler görselleştirilmiştir. İkinci katman kümelemesinde, her bir kişinin kümelerdeki kelime yüzdeleri hesaplanmış ve el yazısı örneklerindeki kelimeler Prototip Kelimeler Çantası'na dönüştürülmüştür. Bu katmanda Yığınsal Hiyerarşik Kümeleme [19] yapılmıştır. 12 adet küme kullanılarak yaş ve cinsiyet ile el yazısı arasındaki ilişki açısından anlamlı sonuçlar elde edilmiştir. Maadeed ve Hassaine'nin [20] yaptıkları çalışmada, el yazısı ile yaş, cinsiyet ve ırk kategorileri sınıflandırılmıştır. QUWI veri tabanından alınan Arapça ve İngilizce el yazısı örneklerini karakterize etmek için çeşitli geometrik öznitelikler (eğim, büyüklük vb.) önerilmiştir. Öznitelikler RO ve çekirdek ayırma analizi kullanılarak birleştirilmiştir. Aynı metni içeren el yazısı örneklerinden cinsiyet tahmini için %74,1, yaş aralığı tahmini için %55,8 ve ırk tahmini için %53,7 başarımla ulaşılmıştır.

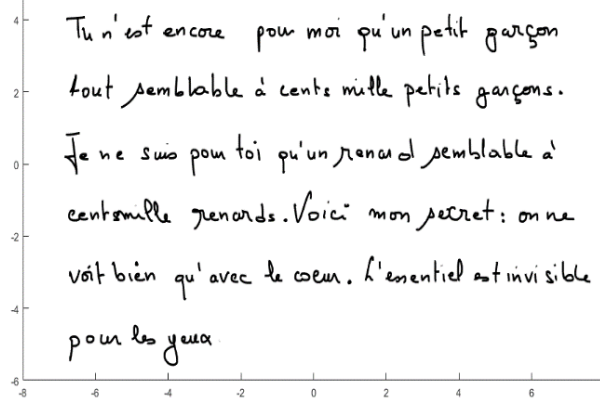
Bu çalışmanın odak noktası bilgisayarlı görme sistemleri yardımı ile el yazısından cinsiyet tespitinin yapılmasıdır. Çalışmada kullanılan el yazıları bir kalem-tablet yardımıyla dijital ortama kaydedilmiştir. Bu kayıtlarda sırası ile kalem ucunun zamansal izlenimi (t), koordinatları (x, y), basıncı (p), genlik açısı (Az) ve yükseklik açısı (Al) gibi özellikler bulunmaktadır. Bu özelliklerden daha sonra cinsiyet tespitinde kullanılacak olan kalem hızı, ivmesi, yazarken oluşan sarsıntı hareketleri, eğim açısı, yazıdaki kavislenmeler, kalemin havada kalma oranı, kalemin yaptığı basınç değeri ve kalemin yükseklik açısı öznitelikleri çıkarılmıştır. Elde edilen özniteliklere dayalı olarak cinsiyetlerin belirlenmesi amacıyla Karar Ağaçları (KA), k-En Yakın Komşu (k-EYK), Destek Vektör Makineleri (DVM), Naive Bayes (NB) sınıflandırıcıları kullanılmıştır. Sınıflandırma başarımları sırası ile %73,4, %79,0, %85,1, %82,4 olarak bulunmuştur. Yapılan çalışmada, ikinci bölümde materyal ve yöntem sunulmuştur. Üçüncü bölümde; deneysel sonuçlar ve tartışmalar paylaşılmıştır. Son bölüm ise çalışmanın sonuçlarını içermektedir.

2. Materyal ve Yöntem

El yazısı tanıma, genel hatlarıyla bakıldığında karakter tanımayı gerektirir. Bu durumda karşımıza etkileşimsiz (çevrimdışı) ve etkileşimli (çevrimiçi) tanıma sistemleri çıkmaktadır. Çevrimdışı tanıma sistemleri, kâğıt üzerine yazılmış bir yazının sayısallaştırma adımlarını içerirken; çevrimiçi tanıma sistemleri, genellikle elektronik tabletler tarafından kalem hareketlerinin koordinatlarının elde edilmesiyle el yazısı ya da çizimlerin otomatik olarak algılanmasını sağlayan sistemlerdir [21]. Verilen bilgiler ışığında bu bölümde çalışmada kullanılan yöntemlere ait bilgiler verilmiştir.

2.1. Çalışmanın amacı ve veri seti

Bu çalışmada önceden verilen belirli bir metnin 10 kişi tarafından tablete yazılması istendiği ve kişilerin özelliklerinin (cinsiyet, yaş, eğitim durumu, hangi elini kullandığı, ana dili vb.) not edildiği Marzinotto ve arkadaşları [17] tarafından hazırlanan bir çevrimiçi el yazısı veri seti kullanılmıştır. Kadın ve erkek bireylerden alınan örnek bir el yazısı görüntüsü Şekil 1'de verilmiştir.



Şekil 1. Veri setinin oluşturulması amacıyla kişilere yazdırılan metin

Tasarlanan sistemde veri setindeki örneklerin boyutu arttırılmak istenmiştir. Bunun için el yazısı örnekleri, tüm metnin incelenmesi yerine veri setindeki 60 satırlık yazı kelime kelime bölütlenmiş ve 410 adet örnek elde edilmiştir. Elde edilen kelime görüntüleri Şekil 2'de verilmiştir.



Şekil 2. Kişiy yazdırılan metnin kelimelere bölütlenmiş hali

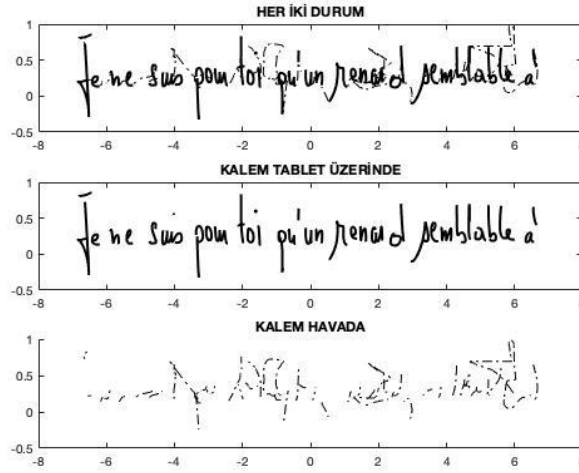
2.2. Öznelik çıkarımı

Kalem ekran yüzeyine değmeden de Şekil 3'de beyaz noktalarla belirli aktif alanın biraz üzerindeki hareketleri algılayabilmektedir. Tablete kalemle yazarken oluşturulan basınç hassasiyeti ilgili kalem ile ölçülebilmektedir. Tabletın basıncının yanında düzlem üzerindeki koordinatları ve yaptığı açı da yine alınabilen önemli değerler arasındadır [22].



Şekil 3. Verilerin alındığı dijital kalem ve grafik tablet

Bu durum göz önüne alındığında öznitelikler üç durum için tanımlanmıştır. Kişi tablete yazı yazarken harfler ve kelimeler arasında geçişlerde elini kaldırdığı haller olmaktadır. Tablet 1 cm'ye kadar yapılan tüm yazma işlemini kaydettiğinden bu geçiş hareketleri “kalem havadayken” başlığı altında toplanmıştır. Kalem tablete temastayken gerçekleştirilen hareketler yani yazma işlemi “kalem tablet üzerindeyken” başlığı altında kaydedilmiştir. Son olarak kişinin yazı yazma hali verilen iki durum da kullanılarak yapıldığından bu hareketler “her iki durum” başlığı altında toplanmıştır. Verilen bu üç başlığın görüntüsü Şekil 4’te verilmiştir.



Şekil 4. Grafik tablette verilerin kaydedilmesi esnasında oluşan yazı ve geçiş görüntüsü örneği

Açıklanan kalemin havada olması durumunda; kalem hızı, ivmesi, yazarken oluşan sarsıntı hareketleri, eğim açısı, yazıdaki kavislenmeler, kalemin havada kalma oranı, kalemin yaptığı basınç değeri ve kalemin yükseklik açısı öznitelikleri çıkarılmıştır. “Kalemin tablet üzerinde olması” ve “her iki durumda” ise kalemin hızı, ivmesi, sarsıntısı, eğim açısı, kavislenme öznitelikleri çıkarılmıştır. Çıkarılan 18 öznitelik Tablo 1’de verilmiştir.

Tablo 1. Kalemin tablet üzerindeki konumuna göre öznitelikler

Kalem Havadayken	Kalem Tablet Üzerindeyken	Her İki Durum
Kalemin Hızı	Kalemin Hızı	Kalemin Hızı
Kalemin İvmesi	Kalemin İvmesi	Kalemin İvmesi
Sarsıntı	Sarsıntı	Sarsıntı
Eğim açısı	Eğim açısı	Eğim açısı
Kavislenme	Kavislenme	Kavislenme
Kalemin havada kalma oranı		
Kalemin Basıncı		
Yükseklik açısı		

El yazısıyla yazılan her kelime için dinamik ve uzamsal olmak üzere iki öznelik türü çıkarılabilir. Birincisi, hız, ivme ve sarsıntı gibi yerel dinamik değerleri hesaplarken [23], ikincisi yerel vuruş açıları, eğriler [24] ya da kelimeler arası boşluklar gibi ölçümlerle kalem yörüngelerinin statik şeklini tanımlar [25]. Dinamik parametreler olarak, n noktasında yerel olarak hesaplanan yatay (x) ve dikey (y) hız aşağıdaki denklemlerle hesaplanır:

$$V_x = | \Delta x / \Delta t | \quad (1)$$

$$V_y = | \Delta y / \Delta t | \quad (2)$$

$$\Delta x (n) = x (n + 1) - x (n - 1) \quad (3)$$

$$\Delta y (n) = y (n + 1) - y (n - 1) \quad (4)$$

$$\Delta t (n) = t (n + 1) - t (n - 1) \quad (5)$$

Bu değerler tüm kelimeler için hesaplanır. Benzer işlemler, sırasıyla hız ve ivmenin yatay ve dikey türevleriyle ilişkili yerel ivme ve sarsıntı değerlerini elde etmek için de kullanılır. Ayrıca, kalem basıncı, basıncın türeviyle ve kalemin havada kalma oranı olan (Havada kalma süresi) / (Toplam süre) [26] denklemiyle hesaplanır ve böylece dinamik özellikler elde edilir. Vuruş değeri iki minimum hız arasındaki bir yazma hareketi olarak tanımlanır. Böylelikle vuruş değeri; kalemin havada kalma sayısını, yatay düzlemde havadaki mesafeyi, kalemin vuruş sayısını ve ortalama uzunluklarını hesaplarken göz önünde bulundurulur [17].

2.3. Sınıflandırma

2.3.1. Karar ağaçları

Karar ağaçları; veri sınıflandırmada kullanılan popüler ve verimli bir sınıflandırma yaklaşımıdır [27]. Karar ağaçlarının amacı, girdi verilerinden çıkarılan basit karar kurallarını öğrenmek ve sonrasında, bir test örneğinin sınıf etiketini tahmin eden bir model oluşturmaktır. Bir karar ağacının yapısı, yaprak ve iç düğüm olmak üzere iki tür düğümü kapsar. Bir yaprak, eğitim örnekleri sonucunun çoğunluğuna göre belirlenen bir sınıf etiketi içerir. Ayrıca, her bir iç düğüm, öznelikler üzerine bir sorudur. Cevaplara göre dallanmalar olur [28].

2.3.2. Naive Bayes

Naive Bayes (NB) sade ve etkin olmasından dolayı en iyi veri madenciliği algoritmalarından biridir. Örneklerin her bir niteliğinin bağımsız olduğu ve her bir nitelik değerinin, diğer nitelikler hakkında bilgi içermediği temeline dayanmaktadır [28]. Bayes ağları, genellikle bir öğrencinin sınıf etiketleri verilen belirli bir eğitim grubundan sınıflandırıcı oluşturmaya çalıştığı problemler için kullanılır [30].

(a_1, a_2, \dots, a_m) gibi bir öznelik vektörü ile temsil edilen x test örneği verildiğinde, Bayes Ağ Sınıflandırıcıları (BAS) sınıflandırma yapmak için aşağıdaki denklemi kullanır:

$$c(x) = \arg \max_{c \in C} P(c) P(a_1, a_2, \dots, a_m | c), \quad (6)$$

Burada C , tüm olası c sınıfı etiketlerinin kümesidir ve $c(x)$, BAS tarafından tahmin edilen x ' in sınıf etiketidir [31].

2.3.3. Destek vektör makineleri

DVM, önemli ve verimli bir denetimli sınıflandırma algoritmasıdır [32]. İkili sınıflandırma probleminde doğrusal olarak ayrılabilen bir verisetinin olduğu düşünülürse, bu veri setini ayırabilen sonsuz sayıda hiper-düzlem vardır. DVM karar düzeyini oluştururken iki sınıfa olan uzaklığı maksimum yapmaya çalışır. Bu düzlemler arasında maksimum sınıfa ait sadece bir hiper-düzlem bulunmaktadır. Sınır genişliğini sınırlandıran noktalara ise destek vektörleri adı verilir [33]. Bir dizi N eğitim $\{(x_i, y_i)_{n=1}^N\}$ veri noktası verilmiştir. Burada x_i çok boyutlu bir özneliktir ve y_i de buna karşılık gelen etikettir. Bir DVM, ayırıcı hiper düzlem olarak eğitim verileri sınıfları arasında bir karar sınıfını modellemektedir. DVM karar fonksiyonu şöyle tanımlanır:

$$f(x_*) = \text{sign}[\sum_{i=1}^N a_i y_i \varphi(x_*, x_i) + b] \quad (7)$$

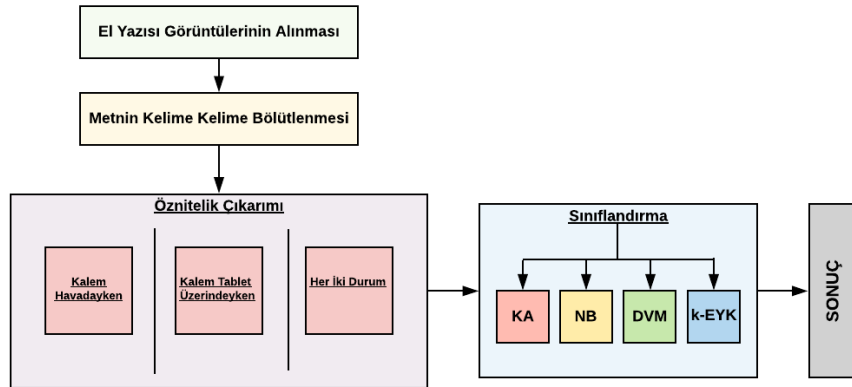
Burada; x_* bir test vektörü, a_i eğitim örneği olan x_i ile ilişkilendirilmiş Lagrangian çarpanı ve b de öğrenilen bias değeridir. DVM'ler; $g(x_i)$ 'nin $\varphi(x_*, x_i) = g(x_*)^t$, $g(x_i)$ noktasal çarpımı üzerinden çekirdek fonksiyonu aracılığıyla uygulanan bir eşleme fonksiyonu olan φ kullanılarak doğrusal olmayan hale getirilir [28].

2.3.4. k-en yakın komşu

k-En Yakın Komşular (k-EYK) sınıflandırıcısının basit işleyişe sahip bir sınıflandırıcı olduğu bilinmektedir. k-EYK sınıflandırma prosedüründe, tüm eğitim örnekleri test örneğini, önceden tanımlanmış bir mesafe fonksiyonuna ve en yakın komşu k sayısına göre sınıflandırmak için kullanılır. Başka bir deyişle, k-EYK yaklaşımı başlangıçta, bir mesafe fonksiyonuna sahip test örneği ve eğitim örnekleri arasındaki benzerlik ölçüsünü hesaplar. Daha sonra, hesaplanan benzerlik ölçüsüne dayanarak, test örneğinin en yakın komşuları belirlenir. Test örneğinin sınıf etiketi, önceden belirlenmiş k en yakın komşuların etiketlerinin çoğunluğunun oyuyla belirlenir [28].

3. Deneysel Sonuçlar ve Tartışmalar

Bu çalışma; ilk olarak veri setinin alınmasını ve veri seti içerisindeki metinlerin kelime kelime bölütlenmesini içermektedir. Sonrasında Tablo 1'de belirtilen durumlara göre öznelik çıkarımı yapılmıştır. Son adımda ise en iyi sınıflandırma türünün belirlenebilmesi için Matlab yazılımı üzerinde; KA, NB, DVM, k-EYK gibi bir dizi sınıflandırıcı ile veri seti test edilmiştir. Önerilen sistemin akış prensibi Şekil 3'te verilmiştir.



Şekil 5. Önerilen çalışmanın akış diyagramı

Yapılan cinsiyet sınıflandırma çalışmasında, kadın ve erkek el yazılarına ait 410 kelimedenden oluşan örnek işleme alınmıştır. Sınıflandırmayı yapmak için veri setinin analiz işlemi 5 katlı çapraz doğrulama ile gerçekleştirilmiştir. Yapılan sınıflandırmalar Tablo 2 üzerinde gösterilmiştir. Sınıflandırma yöntemleri içerisindeki her bir sınıflandırıcı tipinin el yazısı ile cinsiyet sınıflandırma doğruluk yüzdesi karşısında verilmiştir.

Tablo 2. El yazısı ile cinsiyet sınıflandırma yöntem ve doğrulukları

Sınıflandırma yöntemi	Sınıflandırıcı tipi	Doğruluk (%)
Karar ağaçları	Doğruluk Ağacı	73,4
Naive Bayes	Çekirdek Naive Bayes	79,0
DVM	Kuadratik DVM	85,1
K-EYK	Ağırlıklandırılmış k-EYK	82,4

En yüksek cinsiyet sınıflandırma doğruluk oranı Kuadratik DVM ile %85,1 bulunmuştur. Ağırlıklandırılmış k-EYK ile %82,4, KA ile ortalama %73 doğruluk elde edilirken Çekirdek NB ile bu başarı oranı %79 bulunmuştur.

Literatürde yapılan son 5 yıldaki çalışmalara bakıldığında verilerin çevrimiçi ve çevrimdışı alınabildiği el yazısı örnekleri ile oluşturulan veri setleri kullanılmıştır. Araştırmaların QUWI veri seti başta olmak üzere birkaç veri seti üzerinde yoğunlaştığı görülmüştür. Yapılan çalışma el yazısı örneklerinin çevrimiçi olarak alındığı bir veri seti kullanmıştır. Bu yöntemle alınan el yazısı örnekleri, verilerden çıkarılan özniteliklerin daha belirleyici olmasını sağlamıştır. Bu çalışma, el yazısından cinsiyet tespiti için farklı veri setlerini kullanan fakat aynı sınıflandırma yöntemleriyle sonuca ulaşmaya çalışan literatürdeki diğer çalışmalarla karşılaştırıldığında üstünlük göstermiştir. Özellikle, bu çalışmada önerilen sistem üzerinde en iyi sınıflandırmayı veren DVM ile yapılan son çalışmalara baktığımızda, Emran ve arkadaşları [9] öznitelik olarak eğitim, yazı yönü ve eğrilik derecesi gibi değerler kullanmış %76,4 sınıflandırmaya ulaşmıştır. Ahmed ve arkadaşlarının [11] çalışmalarında ise öznitelik çıkarım yönüyle bu çalışmadan ayrılmakta ancak DVM ile %78 başarı göstermektedir. El yazısına dayalı cinsiyet sınıflandırılması ile ilgili son yıllardaki diğer literatür araştırmalarının karşılaştırılması Tablo 3’de özetlenmiştir. Her bir satır farklı bir çalışmaya ait bilgileri içermektedir. El yazısı ile cinsiyet tahmini yapan bu çalışmaların özetlendiği tablodaki sütunlar sırasıyla; yapılan çalışmanın kime ait olduğu, sınıflandırma yönteminin ne olduğu, kullanılan veri setinin adı ve çalışmadaki sınıflandırmanın yüzde cinsinden doğruluk oranını ifade etmektedir.

Tablo 3. El yazısı ile cinsiyet sınıflandırma konusunda son yıllarda yapılan bazı çalışmaların sınıflandırma yöntemleri, veri setleri ve sınıflandırma doğrulukları

Yapılan çalışma	Sınıflandırma yöntemi	Veri seti	Doğruluk (%)
Moetesum ve ark. [8]	NB, KA, LAA, DVM	QUWI	70,1
Emran ve ark. [9]	k-EYK, RO, DVM	QUWI	76,4
Morera ve ark.[10]	ESA	IAM, KHATT	80,7
Ahmed ve ark.[11]	DVM+YSA+k-EYK	QUWI	85,0
Topaloğlu ve ark. [1]	KA	Topaloğlu ve ark [1]	93,8
Guerbai ve ark. [12]	BS-DVM	IAM	77,0
Akbari ve ark. [13]	YSA, DVM	QUWI, MSHD	80,0
Mirza ve ark. [16]	ESA	QUWI	70,0
Maadeed ve Hassaine [20]	RA	QUWI	74,1
Önerilen yöntem	KA, NB, DVM, k-EYK	Marzinotto[17]	85,1

3.1. Performans değerlendirme kriterleri

Makine öğrenmesi sınıflandırma modellerinin değerlendirilmesi amacıyla hata matrisi yorumlanırken Gerçek Pozitif (GP), Gerçek Negatif (GN), Yanlış Pozitif (YP) ve Yanlış Negatif (YN) metrikleri işleme alınır [34]. İlgili sistemde, bahsedilen değerler cinsiyeti belirleme üzerinden açıklanacak olursa: GP: Sistem erkeğe erkek demiştir. YP : Sistem kadına erkek demiştir. GN : Sistem kadına kadın demiştir. YN : Sistem erkeğe kadın demiştir.

		TAHMİN EDİLEN SINIF	
		Gerçek Negatif (73)	Yanlış Pozitif (55)
GERÇEK SINIF	Gerçek Negatif (73)		
	Yanlış Negatif (42)		
		Doğru Pozitif (240)	

Şekil 6. Hata Matrisi Metrikleri

Tasarlanan sistemde Hata Matrisine ait metriklerden Doğruluk, Duyarlılık ve Özgüllük değerleri hesaplanarak performans ölçülmek istenmiştir [34].

Doğruluk (Accuracy): Doğru sınıflandırma değerinin toplam değere bölünmesi ile elde edilir [34].

$$\frac{GP+GN}{GP+GN+YP+YN} = \frac{240+73}{410} = 0,76 \quad (8)$$

Duyarlılık (Sensitivity): Sistemin doğruları bilme konusundaki etkinliğidir. GP ‘nin YN ve GP değerleri toplamına bölümüyle bulunur ve bu değer, gerçekte bulmak istenilen cinsiyetin ne kadarının saptanabildiğidir [34].

$$\frac{GP}{GP+YN} = \frac{240}{240+42} = 0,85 \quad (9)$$

Özgüllük (Specificity): Sistemde GN olan durumlara GN ve YP değerleri toplamının oranlanmasıyla bulunur ve gerçek durumu diğer cinsiyet olan bireylerin ne derece doğru tespit edildiğidir [34].

$$\frac{GN}{GN+YP} = \frac{73}{73+55} = 0,57 \quad (10)$$

4. Sonuçlar

Bu çalışma, KA, NB, DVM, K-EYK sınıflandırıcılarıyla bireylere ait el yazısı örneklerinden cinsiyet tahmini yapmak amacıyla kullanan bir sistem önermektedir. Kullandığımız bu sınıflandırıcılar kalemde havada olduğu, tablet üzerinde olduğu ve her iki durumdaki özneliklerini (kalem hızı, kalem ivmesi, sarsıntı, eğim açısı, kavislenme, kalemde havada kalma oranı, kalem basıncı yükseklik açısı) girdi olarak alır, sonrasında ise çıktı olarak cinsiyeti verir.

Sistem üzerinde, verilen el yazısı örneklerini cinsiyete göre tespit işlevinde %85,1 sınıflandırma elde edilmiştir. Önerilen yöntem, mevcut teknikler ve benzer sınıflandırıcı türleri ile karşılaştırıldığında yüksek bir başarı oranı göstermektedir. El yazısı özelliklerinin değerlendirildiği bu sistem ile farklı bireysel özellikler saptanabilecektir. Ayrıca multidisipliner gözlem ve izlem çalışmalarına katkı sunacaktır.

Kaynaklar

- [1] Topaloglu M, Ekmekci S. Gender detection and identifying one's handwriting with handwriting analysis. *Expert Systems with Applications*, 79, 236-243.
- [2] Langmaid, R. A. Papadopoulos, N. Johnson, B. P. Phillips, J. G. & Rinehart, N. J. Handwriting in children with ADHD. *Journal of Attention Disorders* 2014; 18(6), 504-510.
- [3] Gattal, A. Djeddi, C. Siddiqi, I. & Chibani, Y. Gender classification from offline multi-script handwriting images using oriented basic image features (oBIFs). *Expert Systems with Applications* 2018; 99, 155-167.
- [4] Burr, V. Judging gender from samples of adult handwriting: Accuracy and use of cues. *The Journal of Social Psychology* 2002; 142(6), 691-700.
- [5] Hartley, J. Sex differences in handwriting: A comment on Spear. *British Educational Research Journal* 1991; 17(2), 141-145.
- [6] Dorfberger, S. Adi-Japha, E. & Karni, A. Sex differences in motor performance and motor learning in children and adolescents: an increasing male advantage in motor learning and consolidation phase gains. *Behavioural Brain Research* 2009; 198(1), 165-171.
- [7] Prasad, S. Singh, V. K. & Sapre, A. Handwriting analysis based on segmentation method for prediction of human personality using support vector machine. *International Journal of Computer Applications* 2010; 8(12), 25-29.
- [8] Moetesum, M. Siddiqi, I. Djeddi, C. Hannad, Y. & Al-Maadeed, S. Data Driven Feature Extraction for Gender Classification using Multi-script Handwritten Texts. In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)* 2018; IEEE. pp. 564-569.
- [9] Moetesum, M. Siddiqi, I. Djeddi, C. Hannad, Y. & Al-Maadeed, S. Data Driven Feature Extraction for Gender Classification using Multi-script Handwritten Texts. In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)* 2018; IEEE. pp. 564-569.
- [10] Morera, Á. Sánchez, Á. Vélez, J. F. & Moreno, A. B. Gender and handedness prediction from offline handwriting using convolutional neural networks. *Complexity*, 2018.
- [11] Ahmed, M. Rasool, A. G. Afzal, H. & Siddiqi, I. Improving handwriting based gender classification using ensemble classifiers. *Expert Systems with Applications* 2017; 85, 158-168.
- [12] Guerbai, Y. Chibani, Y. & Hadjadji, B. Handwriting gender recognition system based on the one-class support vector machines. In *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)* 2017; 1-5 IEEE.
- [13] Akbari, Y. Nouri, K. Sadri, J. Djeddi, C. & Siddiqi, I. Wavelet-based gender detection on off-line handwritten documents using probabilistic finite state automata. *Image and Vision Computing* 2017; 59, 17-30.

- [14] Al Maadeed, S. Ayoub, W. Hassaine, A. & Aljaam, J. M. Quwi: An arabic and english handwriting dataset for offline writer identification. In *2012 International Conference on Frontiers in Handwriting Recognition 2012*; IEEE. pp.746-751.
- [15] Djeddi, C. Gattal, A. Souici-Meslati, L. Siddiqi, I. Chibani, Y. & El Abed, H. LAMIS-MSHD: a multi-script offline handwriting database. In *2014 14th International Conference on Frontiers in Handwriting Recognition 2014*; IEEE. pp. 93-97.
- [16] Mirza, A. Moetesum, M. Siddiqi, I. & Djeddi, C. Gender classification from offline handwriting images using textural features. In *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR) 2016*; IEEE. pp. 395-398.
- [17] Marzinotto, G. Rosales, J. C. El-Yacoubi, M. A. & Garcia-Salicetti, S. Age and gender characterization through a two layer clustering of online handwriting. In *International Conference on Advanced Concepts for Intelligent Vision Systems 2015*; 428-439 Springer, Cham.
- [18] Viard-Gaudin, C. Lallican, P. M. Knerr, S. & Binter, P. The ireste on/off (ironoff) dual handwriting database. In *Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 1999*; IEEE. pp. 455-458.
- [19] Bharath, A. Deepu, V. & Madhvanath, S. An approach to identify unique styles in online handwriting recognition. In *Eighth International Conference on Document Analysis and Recognition (ICDAR'05) 2005*; IEEE. pp. 775-778.
- [20] Al Maadeed, S. & Hassaine, A. Automatic prediction of age, gender, and nationality in offline handwriting. *EURASIP Journal on Image and Video Processing, 2014* (1), 10.
- [21] Duygulu, P. Ören, T. & Çölkesen, R. El Yazısı Tanıma. *Bilişim Ansiklopedisi*, 2004.
- [22] Wacom Intuos, T. M. User's Manual for Windows, copyright Wacom Company, 2000.
- [23] Bharath, A. Deepu, V. & Madhvanath, S. An approach to identify unique styles in online handwriting recognition. In *Eighth International Conference on Document Analysis and Recognition (ICDAR'05) 2005*; IEEE. pp. 775-778.
- [24] Matic, N. Guyon, I. Denker, J. & Vapnik, V. Writer-adaptation for on-line handwritten character recognition. In *Proceedings of 2nd International Conference on Document Analysis and Recognition (ICDAR'93) 1993*; IEEE. pp. 187-191.
- [25] Crettez, J. P. A set of handwriting families: style recognition. In *Proceedings of 3rd International Conference on Document Analysis and Recognition 1995*; IEEE. pp. 489-494.
- [26] Rosenblum, S. Engel-Yeger, B. & Fogel, Y. Age-related changes in executive control and their relationships with activity performance in handwriting. *Human movement science 2013*; 32(2), 363-376.
- [27] Safavian, S. R. & Landgrebe, D. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics 1991*; 21(3), 660-674.
- [28] Sengur, D. Turhan, M. & Karabatak, S. Prediction of The School Administrators, Who Attended An Action Learning Course, Based On Their Conflict-Handling Styles: A Data Mining Approach. *International Online Journal of Educational Sciences 2018*; 10(4).
- [29] Utku, A. & Doğru, İ. A. Android kötüçül yazılımlar için izin tabanlı tespit sistemi. *Journal of the Faculty of Engineering and Architecture of Gazi University 2017*; 32(4), 1015-1024.
- [30] Geiger, D. Verma, T. & Pearl, J. Identifying independence in Bayesian networks. *Networks 1990*; 20(5), 507-534.
- [31] Jiang, L. Li, C. Wang, S. & Zhang, L. Deep feature weighting for naive Bayes and its application to text classification. *Engineering Applications of Artificial Intelligence 2016*; 52, 26-39.
- [32] Schurz, G. *Hume's Problem Solved: The Optimality of Meta-induction*. Mit Press 2019.
- [33] Kucuk, H. Eminoglu, I. & Balci, K. Classification of neuromuscular diseases with artificial intelligence methods. *Journal of the Faculty of Engineering and Architecture of Gazi University 2019*; 34(4), 1725-1741.
- [34] Forbes, A. D. Classification-algorithm evaluation: Five performance measures based on confusion matrices. *Journal of Clinical Monitoring 1995*; 11(3), 189-206.