

MFKK Özniteliklerine Eklenen Logaritmik Enerji ve Delta Parametrelerinin Yaş ve Cinsiyet Sınıflandırma Üzerindeki Etkileri

Ergün YÜCESOY^{1*}

ÖZET: Konuşmacıların yaş ve cinsiyet gruplarının otomatik olarak belirlenmesi önemli bir araştırma konusudur ve başta çağrı merkezleri olmak üzere birçok alanda farklı amaçlarla kullanılmaktadır. Bu çalışmada Mel Frekanslı Kepstrum Katsayılarına (MFKK) eklenen logaritmik enerji ve delta parametrelerinin otomatik yaş ve cinsiyet tanıma üzerindeki etkileri araştırılmıştır. Konuşma sinyallerinden çıkarılan MFKK öznitelikleri, Gauss Karışım Modeli (GKM) süpervektörlerine dönüştürüldükten sonra Destek Vektör Makinesine (DVM) uygulanmış ve gerçekleştirilen optimizasyon süreci sonunda konuşmacıların yaş ve cinsiyet gruplarına karar verilmiştir. Çalışmada MFKK'ya eklenen parametrelerin yanı sıra MFKK sayısının ve GKM bileşen sayısının başarı üzerindeki etkileri de araştırılmıştır. MFKK sayısı 8 ile 20, GKM bileşen sayısı ise 32 ile 256 arasında değiştirilerek sistem üzerinde testler yapılmıştır. aGender veritabanının geliştirme bölümündeki 299 konuşmacının 1388 konuşması ile yapılan testlerde en yüksek sınıflandırma oranı, 12 keprstral katsayıya logaritmik enerji, delta ve delta-delta parametrelerinin eklenmesi sonucunda %60.23 olarak hesaplanmıştır. Çalışmada optimum GKM bileşen sayısı 128 olarak belirlenirken, logaritmik enerji, delta ve delta-delta parametrelerinin başarı üzerindeki etkileri sırasıyla %1.17, %3.24 ve %4.61 olarak saptanmıştır.

Anahtar Kelimeler: Yaş ve cinsiyet sınıflandırma, Konuşma işleme, Destek vektör makineleri, Gauss karışım modeli

Effect of Inclusion of Delta Derivatives and Log Energy to MFCC Features on Age and Gender Classification

ABSTRACT: Automatic recognition of the age and gender groups of the speakers is an important research topic and is used for different purposes in many fields, especially in call centers. In this study, the effects of logarithmic energy and delta parameters added to Mel Frequency Cepstral Coefficients (MFCC) on automatic age and gender recognition were investigated. After transforming the MFCC features extracted from speech signals into Gaussian Mixture Model (GMM) supervectors, they were applied to the Support Vector Machine (DVM) and the age and gender groups of the speakers were decided at the end of the optimization process. In the study, besides the parameters added to MFCC, the effects of MFCC number and GMM component number on success were also investigated. MFCC number was changed between 8 and 20 and GMM component number was changed between 32 and 256 and tests were performed on the system. In tests performed with 1388 speeches of 299 speakers in the development section of aGender database, the highest classification rate was calculated as 60.23% by adding logarithmic energy, delta and delta-delta parameters to 12 cepstral coefficients. In the study, the optimum GMM component number was determined as 128, while the effects of logarithmic energy, delta and delta-delta parameters on success were 1.17%, 3.24% and 4.61%, respectively.

Keywords: Age and gender classification, Speech processing, Support Vector Machine, Gauss Mixture Model

¹Ergün YÜCESOY (Orcid ID: 0000-0003-1707-384X), Ordu Üniversitesi, Teknik Bilimler Meslek Yüksekokulu, Elektronik ve Otomasyon Bölümü, Ordu, Türkiye

*Sorumlu Yazar/Corresponding Author: Ergün YÜCESOY, e-mail: yucesoy@hotmail.com

GİRİŞ

Konuşma insanların birbirleriyle olan iletişimde kullanılan fizyolojik bir sinyaldir. Gelişen teknoloji ile birlikte bu sinyal insanların bilgisayarlarla olan iletişimde de kullanılmaya başlanmıştır. Konuşma sinyali yalnızca seslendirilen ifadenin dilsel (linguistic) bilgilerini içermez. O aynı zamanda konuşmacının kimliği, yaşı, cinsiyeti, duygusal durumu ve etnik kökeni gibi paralinguistik bilgilerini de içerir. Paralinguistik bilgilerin konuşma sinyalinden otomatik olarak çıkarılması oldukça zor bir iştir ve başta ticari, adli ve tıbbi olmak üzere birçok uygulama alanına sahiptir. Bir konuşmadaki en belirgin paralinguistik bilgi olan konuşmacının yaşı ve cinsiyeti genellikle insan-bilgisayar iletişiminin kişiye özel olarak uyarlanması için kullanılır. Örneğin otomatik diyalog sistemlerinde konuşma sentezleyicinin hızı, kullanıcının tespit edilen yaşına göre değiştirilebilir (van Heerden ve ark., 2010). Benzer şekilde bir e-ticaret platformunda kullanıcının yaşı ve cinsiyeti tespit edilerek kişiye özel ürün veya hizmet önerisi yapılabilir. Konuşmacının yaş ve cinsiyet bilgisinin kullanıldığı diğer bir alan da adli vakalardır. Ses kayıtlarının delil olarak kullanıldığı adli vakalarda suçlunun belirlenmesinde otomatik yaş tespitinden yararlanılabilir. Güvenlik alanında ise akıllı izleme ve kullanıcının yaşına göre yetkilendirilmiş farklı erişim seviyelerinin tanımlandığı sistemler düşünülebilir (Choukri ve Wu, 2019). Otomatik yaş ve cinsiyet tanıma, aynı zamanda ruhsal durum tanıma, sigara kullanan konuşmacıların belirlenmesi, sarhoşluk seviyesinin belirlenmesi ve hatta otomatik konuşma tanıma gibi diğer ses teknolojilerinin performanslarının artırılması amacıyla da kullanılabilir (Bahari ve ark., 2014).

Konuşmacının kısa süreli konuşmasından kişinin yaşının ve cinsiyetinin tanınması konusunda birçok çalışma yapılmıştır. Bu çalışmaların bir kısmında konuşmacıların yaş ve cinsiyet özellikleri ayrı ayrı ele alınırken bir kısmında ise bu bilgiler birlikte değerlendirilmiştir. Cinsiyet tanıma konusunda yapılan çalışmalarda genellikle yetişkin konuşmacılar kullanılmış ve konuşmacıların erkek ve kadın olarak iki sınıfa ayrılması amaçlanmıştır. Diğer taraftan yetişkin ve çocuk konuşmacıların birlikte ele alındığı çalışmalar da yapılmıştır. Bu çalışmalarda genellikle çocuk konuşmacıların cinsiyet ayrımı yapılmamış ve konuşmacıların çocuk, erkek ve kadın olarak sınıflandırılması yapılmıştır. Yaş tanıma konusunda yapılan çalışmalarda ise yaş regresyonu ve yaş grubu sınıflandırma olmak üzere iki farklı yaklaşım söz konusudur. Yaş regresyonunda konuşmacının yıl olarak yaşının tahmin edilmesi amaçlanırken, yaş grubu sınıflandırmada ise konuşmacının yaş aralığına göre çocuk, genç, yetişkin ve yaşlı gibi sınıflara ayrılması amaçlanmaktadır. Özellikle son yıllarda yapılan çalışmalarda yaş grubu sınıflandırma ile cinsiyet sınıflandırma birlikte ele alınmaya başlanmıştır. Bu çalışmalarda genellikle konuşmacılar yaş grubuna göre çocuk, genç, yetişkin ve yaşlı olarak dört sınıfa ayrıldıktan sonra çocuklar hariç diğer konuşmacıların cinsiyetlerine göre erkek ve kadın olarak sınıflandırıldığı 7 sınıflı bir sınıflandırma yaklaşımı kullanılmaktadır.

Başlangıçta konuşma ve konuşmacı tanıma gibi diğer konuşma işleme uygulamaları için geliştirilen yöntemler, küçük değişikliklerle yaş ve cinsiyet tanımaya uyarlanmaktadır. Literatürde yaş ve cinsiyet tanıma konusunda farklı öznitelik çıkarma ve sınıflandırma yöntemlerinin kullanıldığı birçok çalışma yapılmıştır. Metze ve arkadaşları tarafından yapılan çalışmada dört farklı yaklaşım insan performansı ile karşılaştırılmıştır (Metze ve ark., 2007). Bu yaklaşımların ilkinde otomatik dil tanıma sisteminden türetilmiş paralel bir fonem tanıyıcı kullanılmıştır. Birkaç prosodik özneliğin dinamik Bayes ağı ile birleştirildiği sistem ikinci, doğrusal öngörü analizine dayanan sistem üçüncü, MFKK özniteliklerine dayalı GKM sistemi ise dördüncü yaklaşım olarak incelenmiştir. Çalışmada paralel fonem tanıyıcı sistemin performansının insan performansına yakın olduğu ancak kısa konuşmalarda performansının düştüğü rapor edilmiştir. Her bir yaklaşımın SpeechDat-II veritabanı üzerindeki doğrulukları sırasıyla %54, %40 %27 ve %42 olarak verilmiştir. Bocklet ve arkadaşları tarafından

yapılan çalışmada, GKM-Genel Arkaplan Modeli (GAM) ve GKM süpervektör DVM yaklaşımlarının 7 sınıflı yaş ve cinsiyet sınıflandırma performansları karşılaştırılmıştır (Bocklet ve ark., 2008). Yapılan testlerde DVM yaklaşımı %77 kesinlik (precision) ve %74 duyarlılık (recall) oranı ile en başarılı yöntem olarak belirlenmiştir. Meinedo ve arkadaşları tarafından yapılan çalışmada yaş ve cinsiyet sınıflandırma için iki ayrı sistem tanıtılmıştır (Meinedo ve Trancoso, 2010). Bu sistemlerden yaş sınıflandırma sistemi dört, cinsiyet sınıflandırma sistemi ise altı alt sistemin birleşimi ile oluşturulmuştur. Her bir alt sistemin eğitiminde uzun ve kısa süreli akustik ve prosodik öznitelikler, farklı sınıflandırıcılar (GKM-GAM, Çok Katmanlı Algılayıcılar (ÇKA) ve DVM) ve farklı veri kümeleri kullanılmıştır. Geliştirilen yaş ve cinsiyet sınıflandırma sistemleri aGender veri kümesinin geliştirme bölümü ile test edilmiş ve bu testlerde cinsiyet sınıflandırma sisteminin doğruluğu %83.1, yaş sınıflandırma sisteminin doğruluğu ise %51.2 olarak belirlenmiştir. Li ve arkadaşları tarafından yapılan çalışmada yedi alt sistemin akustik ve prosodik seviyede birleşimine dayanan yeni bir yaklaşım sunulmuştur (Li ve ark., 2013). Çalışmada GKM-GAM, GKM/DVM ve uzun dönemli 450 özniteliğin DVM ile sınıflandırıldığı üç temel alt sisteme ilave olarak dört yeni alt sistem daha önerilmiştir. Bu alt sistemlerin farklı kombinasyonları incelenmiş ve en yüksek başarı, yaş ve cinsiyet sınıflandırma için %51.1, yaş sınıflandırma için %52.2, cinsiyet sınıflandırma için ise %88.2 olarak tüm alt sistemlerin skor seviyeli birleşimi sonucunda elde edilmiştir. Grzybowska ve arkadaşları tarafından yapılan çalışmada i-vektörlerin kullanımına dayalı yeni bir yaş sınıflandırma yaklaşımı sunulmuştur (Grzybowska ve Kacprzak, 2016). Test i-vektörleriyle yaş sınıflarının i-vektörleri arasındaki kosinüs uzaklığına dayalı olarak yapılan sınıflandırma sonucunda %62.9 doğruluğa ulaşılarak Interspeech 2010 yarışmasındaki en iyi sonuç %16.7 oranında geliştirilmiştir. Mallouh ve arkadaşları tarafından yapılan çalışmada yaş ve cinsiyet sınıflandırma için iki Derin Sinir Ağının (DSA) (yaş ve cinsiyet) birleşimine dayalı yeni bir DSA sistemi önerilmiştir (Mallouh ve ark., 2017). Önerilen sistemin sonuçları ile i-vektör ve GKM-GAM sistemlerinin sonuçları aGender veritabanı kullanılarak karşılaştırılmıştır. Yapılan testlerde üç sistemin genel doğruluk oranları sırasıyla %55.16, %47.89 ve %43.8 olarak hesaplanırken önerilen sistemin hem genel doğruluk açısından hem de bireysel sınıflar açısından daha başarılı olduğu görülmüştür. Safavi ve arkadaşları tarafından yapılan çalışmada çocuk konuşmalarından konuşmacı, cinsiyet ve yaş grubu tanıma konusu üzerinde durulmuştur (Safavi ve ark., 2018). Çalışmada GKM-GAM, GKM-DVM ve i-vektör tabanlı yaklaşımları içeren birkaç sınıflandırma yönteminin performans karşılaştırması yapılmıştır. Yapılan testlerde yaş arttıkça konuşmacı tanıma oranının arttığı ancak yaşın, cinsiyet ve yaş grubu tanımadaki etkisinin daha karmaşık olduğu görülmüştür. Çalışmada yaştan bağımsız GKM-DVM sisteminin cinsiyet sınıflandırma başarısı AG1, AG2 ve AG3 sınıfları için sırasıyla %76.80, %84.14 ve %75.91 olarak, i-vektör tabanlı cinsiyetten bağımsız yaş sınıflandırma sisteminin başarısı ise %82.62 olarak verilmiştir. Ertam tarafından yapılan çalışmada ise bir ses veri kümesinden konuşmacının cinsiyet öngörüsü için Uzun Kısa Süreli Bellek (Long short-term memory-LSTM) ağlarının kullanımı önerilmiştir (Ertam, 2019). Önerilen yaklaşımın ilk aşamasında en etkin 10 öznitelik seçilmiş ve daha sonra çift katmanlı LSTM yapısı ile derin öğrenme tabanlı bir ağ oluşturulmuştur. Önerilen yöntem ağaç, doğrusal ayırıcı, lojistik regresyon, DVM (farklı aktivasyon fonksiyonları ile) ve K-en yakın komşuluk gibi geleneksel makine öğrenme yaklaşımlarının sonuçları ile karşılaştırılmış ve % 98.4 başarı oranı ile önerilen yöntemin en başarılı yöntem olduğu gösterilmiştir. Choukri ve Wu tarafından yapılan çalışmada mobil cihazlara erişim yetkisinin kontrolü için yeni bir yaş ve cinsiyet sınıflandırma yöntemi önerilmiştir (Choukri ve Wu, 2019). Konuşma sinyallerinden çıkarılan 40 MFKK'nın öznitelik olarak kullanıldığı çalışmada altı DVM çekirdek modeli incelenmiş ve kübik, kuadratik ve orta Gauss çekirdek modellerinin tanıma oranını sırasıyla %93.75, %91.25 ve %93.75'e çıkarabileceği sonucuna varılmıştır. Son ve

arkadaşları tarafından yapılan çalışmada sözcüksel olmayan üç konuşma özelliğinin (doldurucular-fillers, kesişme-overlapping ve uzatma-lengthening) cinsiyet sınıflandırmaya etkisi araştırılmıştır (Son ve ark., 2019). Çalışmada temel akustik öznitelik olarak MFKK ile sözcüksel olmayan özniteliklerin farklı kombinasyonları ele alınmış ve iki farklı makine öğrenme yöntemi (DVM ve RNN-Recurrent Neural Network) kullanılarak sınıflandırma yapılmıştır. Her iki yöntemde de en yüksek başarı MFKK ile kesişme ve uzatma özniteliklerinin birlikte kullanılması durumunda sağlanırken RNN ile %89.61, DVM ile ise %86.58 sınıflandırma başarısına ulaşılmıştır. Zazo ve arkadaşları tarafından yapılan çalışmada ise LSTM-RNN'ye dayalı yeni bir yaş tahmin sistemi önerilmiştir (Zazo ve ark., 2018). Gerçek zamanlı bir mimaride kolayca uygulanabilmesi için kısa konuşmaların (3-10s) kullanıldığı çalışmada, önerilen yaklaşım (LSTM-RNN) ile i-vektör temelli yaklaşım NIST veri kümesi kullanılarak karşılaştırılmıştır. Yapılan testlerde ortalama mutlak hata açısından önerilen sistemin %28'e varan gelişme sağladığı görülmüştür.

Konuşmaya dayalı tanıma sistemlerinin başarısını etkileyen en önemli faktörlerin başında ses sinyalinden çıkarılan özniteliklerin seçimi gelmektedir. Temel frekans, formantlara karşılık gelen frekans, genlik ve bant genişlikleri, lineer öngörü kepsral katsayıları ve Mel Frekans Kepstral Katsayılar (MFKK) ses işlemede kullanılan özniteliklerden bazılarıdır. Bu özniteliklerden MFKK en yaygın kullanılan öznitelik türü olup başta konuşma ve konuşmacı tanıma sistemleri olmak üzere birçok çalışmada tek başına veya farklı özniteliklerle birlikte kullanılmıştır. Dhonde ve arkadaşlarının çalışmasında MFKK özniteliklerinin sayısı ile delta parametrelerinin (delta ve delta-delta) kişi tanıma etkisi ayrıntılı olarak incelenmiştir (Dhonde ve ark., 2017). Benzer şekilde Fang ve arkadaşları tarafından yapılan çalışmada (Fang ve ark., 2019) MFKK, MFKK+delta ve normalize edilmiş MFKK+delta parametrelerinin patolojik ses tespiti üzerindeki etkisi, Koo ve arkadaşlarının çalışmasında (Koo ve ark., 2020) ise MFKK delta ve ivme parametrelerinin ruhsal durum tanıma etkisi incelenmiştir. Ancak benzer bir çalışma yaş ve cinsiyet tanıma konusunda yapılmamıştır. Bu çalışmada farklı sayıda MFKK özniteliklerine eklenen delta, delta-delta ve logaritmik enerji parametrelerinin yaş ve cinsiyet tanıma üzerindeki etkileri araştırılmıştır. Her konuşmadan çıkarılan öznitelik vektörleri GKM-GAM yapılandırmasına göre GKM-süpervektörlere dönüştürüldükten sonra bir DVM'ye uygulanmış ve gerçekleştirilen optimizasyon sonucunda konuşmacıların yaş ve cinsiyet gruplarına karar verilmiştir.

MATERYAL VE YÖNTEM

Mel-Frekans Kepstral Katsayılar (MFKK)

Mel-Frekans Kepstral Katsayılar sesin spektral özelliklerini temsil etmek için kullanılan bir gösterim şeklidir. Bu gösterimde insanın frekans algılaması dikkate alınır ve bu nedenle konuşma tanıma için en iyi öznitelik olarak kabul edilir (Kerkeni ve ark., 2018). MFKK özniteliklerinin çıkarılma süreci sinyalin yüksek geçiren bir filtreyle filtrelenmesi ile başlar. Filtreleme işlemi insanın ses üretim mekanizması sırasında bastırılan yüksek frekanslı kısımların telafi edilmesini sağlar. Daha sonra sinyal durağan akustik özelliklere sahip olduğu varsayılan kısa zaman aralıklarına (20-30ms) bölünür. Bu amaçla bir pencere fonksiyonu (genellikle Hamming) bir kısmı (yaklaşık 10sn) çakışacak şekilde sinyal boyunca kaydırılır. Elde edilen çerçeveler üzerinde Fourier dönüşümü uygulanarak sinyalin önce genlik spektrumu daha sonra ise mel ölçeğine göre haritalanarak mel-spektrumu tahmin edilir. Son aşamada ise logaritması alınan mel-spektrum üzerinde ayrık kosinüs dönüşümü uygulanır ve elde edilen ilk birkaç katsayı MFKK olarak değerlendirilir. Sıfırıncı katsayı ortalama logaritmik enerjiyi temsil eder ve genellikle öznitelik kümesinin dışında tutulur (Rao ve ark., 2017).

Kepsral katsayılar analiz edilen çerçevenin spektral özelliklerini iyi temsil eder. Ancak sinyalin zamanla değişen dinamik özelliklerini içermez. Sinyalin zamansal dinamikleri hakkındaki ekstra

bilgiler, kepstal katsayıların birinci ve ikinci türevleri hesaplanarak elde edilir. Kepstral katsayıların birinci türevi delta, ikinci türevi ise delta-delta (ivme) katsayıları olarak adlandırılır. Bu parametrelerin hesaplanmasında genellikle aşağıdaki eşitlik kullanılır (Rabiner ve ark., 2008).

$$\Delta c_m(n) = \frac{\sum_{i=-T}^T k_i c_m(n+i)}{\sum_{i=-T}^T |i|} \quad (1)$$

Burada $c_m(n)$ n 'inci çerçeve için m 'inci özniteliği, k_i i 'inci ağırlığı, T ise ardışık çerçeve sayısını gösterir. Genellikle T 2 olarak alınır. Delta-Delta katsayıları da aynı şekilde hesaplanır, ancak bu hesaplamada sabit kepstal katsayılar yerine delta katsayıları kullanılır.

Gauss Karışım Modeli (GKM)

Gauss Karışım Modeli, bir veri kümesinden bu verileri üreten fonksiyonun olasılık yoğunluk fonksiyonunu tahmin etmek için kullanılan yarı parametrik bir tahmin yöntemidir. Bu olasılık yoğunluk fonksiyonunun M adet Gauss fonksiyonunun doğrusal bir kombinasyonu olduğu kabul edilir. D boyutlu bir öznitelik vektörü için, x , olabilirlik fonksiyonu olarak kullanılan karışım yoğunluğu Eşitlik 2 ile tanımlanır.

$$P(x|\lambda) = \sum_{m=1}^M w_m \cdot p_m(x) \quad (2)$$

Burada $p_m(x)$ i 'inci Gauss yoğunluğunu, w_m ise ağırlığını gösterir. Her bir Gauss yoğunluğu μ_m , $D \times 1$ boyutlu bir ortalama vektörü ve Σ_m , $D \times D$ boyutlu bir kovaryans matrisi olmak üzere Eşitlik 3 ile tanımlanır.

$$p_m(x) = \frac{1}{(2\pi)^{D/2} |\Sigma_m|^{1/2}} e^{-\frac{1}{2}(x-\mu_m)^T \cdot (\Sigma_m)^{-1} \cdot (x-\mu_m)} \quad (3)$$

Gauss karışım modeli tüm bileşenlerin ortalama, kovaryans ve ağırlık parametrelerinin birleşiminden oluşur ve $\lambda = \{w_m, \mu_m, \Sigma_m\} \quad m = 1, \dots, M$ şeklinde gösterilir. Bu parametrelerin tahmininde ise tekrarlamalı bir yöntem olan EM algoritması kullanılır (Dempster ve ark., 1977).

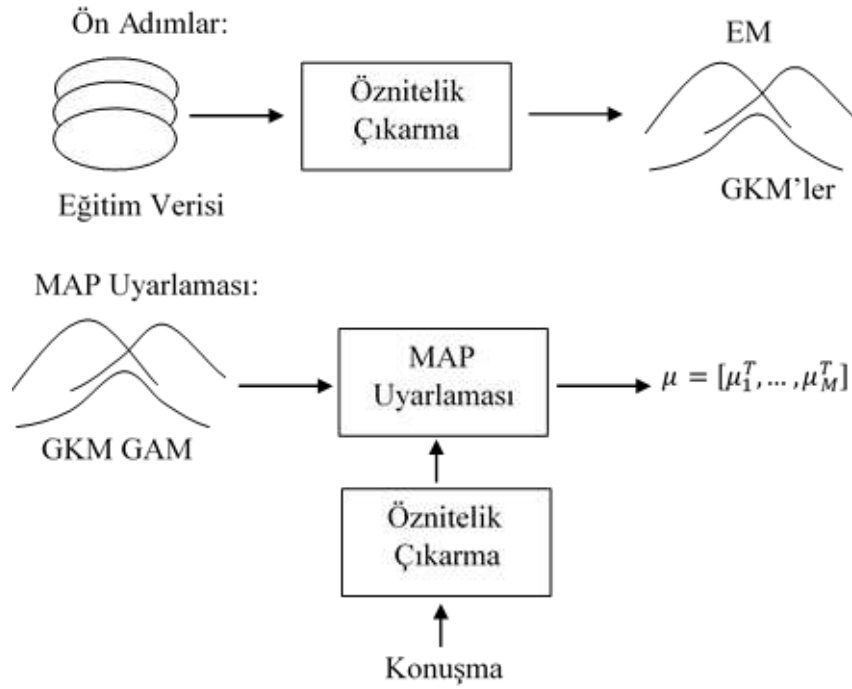
GKM Süpervektörler

Konuşma sinyallerinden çıkarılan öznitelik vektörleri çeşitli sınıflandırıcılara aktararak konuşmacıların kimlik, cinsiyet, yaş, ruhsal durum gibi farklı özellikleri tanınabilir. Ancak konuşma süreleri genellikle farklıdır ve bu durumda konuşma sinyallerinden çıkarılan öznitelik vektörlerinin uzunlukları da farklı olacaktır. Diğer taraftan sınıflandırıcıların neredeyse tamamında, giriş olarak sabit uzunluklu vektörler kullanılır. Bu sebeple konuşma sinyalinden çıkarılan öznitelik vektörlerinin sabit uzunluklu vektörlere dönüştürülmesi gerekir. GKM süpervektörler bu fikirden ortaya çıkmış bir yöntemdir ve aşağıdaki adımlar uygulanarak oluşturulur.

- İlk olarak konuşmacıdan bağımsız özniteliklerin dağılımını temsil eden geniş bir GKM modeli eğitilir. Genel Arkaplan Modeli (GAM) olarak isimlendirilen bu model herhangi bir kişiye, cinsiyete, ruhsal duruma vs. bağımlı değildir ve tüm modellerin temeli olarak düşünülür. GAM'ın eğitiminde kullanılan veritabanının farklı ortamları (örneğin farklı arka plan gürültülerini), farklı duygusal durumları, farklı konuşmacıları (erkek veya kadın) kapsayacak şekilde geniş bir veri kümesinden seçilmesi gerekmektedir.

- Her konuşma için, önceden eğitilmiş GAM, geçerli konuşmanın içerdiği bilgileri temsil edecek şekilde uyarlanır. Uyarlama işlemi için genellikle maksimum-a-posterior (MAP) algoritması kullanılır (Reynolds ve ark., 2000).

- Uyarlanmış modelin bileşenlerinin ortalama vektörleri birleştirilerek $\mu = [\mu_1^T, \dots, \mu_M^T]$ ile gösterilen süpervektörler oluşturulur. Bu süreç Şekil 1'de gösterilmiştir.



Şekil 1. GKM süpervektörlerin hesaplanması

Destek Vektör Makineleri (DVM)

Destek Vektör Makinesi, ayırıcı bir hiper düzlem ile sınıflandırma yapan iki sınıflı bir sınıflandırıcıdır. Hiper düzlemin konumu, eğitim vektörleri ile hiper düzlem arasındaki mesafe en az olacak şekilde belirlenir. Hiper düzlemdeki veri noktalarına destek vektörleri denir. Eğitim veri kümesi doğrusal olarak ayrılamaz ise, destek vektörleri doğrusal olmayan bir dönüşümle yüksek boyutlu bir alana dönüştürülebilir. Doğrusal olmayan bu dönüşüm çekirdek işlevi ile gerçekleştirilir. Çekirdek işlevi, öznelik uzayında iki nokta arasındaki uzaklığı temsil eder ve Eşitlik 4 ile gösterilir.

$$K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j) \quad (4)$$

Burada $\varphi(x_i)$ giriş uzayından öznelik uzayına bir haritalama fonksiyonudur. Mercer koşulunu sağlayan çekirdek işlevi ile doğrusal olmayan sınırlara genişletilen bir DVM sınıflandırıcısı,

$$f(x) = \sum_{i=1}^L \lambda_i y_i K(x, x_i) + d \quad (5)$$

ifadesi ile tanımlanır. Burada y_i , $y_i \in \{-1, 1\}$ olmak üzere eğitim vektörlerinin sınıf etiketini, L destek vektörlerinin sayısını, d ise öğrenilen bir sabiti gösterir. λ_i 'ler negatif olmayan Lagrange çarpanlarıdır ve değerleri kuadratik programlama probleminin doğrusal kısıtlamalar altında çözülmesiyle hesaplanır.

DVM'nin doğrusal olmayan sınırlara genişletilmesinde kullanılan birçok çekirdek işlevi mevcuttur. Bu çekirdeklerden doğrusal, radyal tabanlı, polinomial ve GKM KL çekirdeklerinin yaş ve cinsiyet sınıflandırma üzerindeki etkilerinin ele alındığı bir çalışma yapılmıştır (Yücesoy., 2020). Bu çalışmada GKM süpervektörleri ile temsil edilen konuşmaların yaş ve cinsiyet gruplarına göre sınıflandırılmasında en başarılı çekirdek işlevinin GKM-KL çekirdeği olduğu gösterilmiştir. Bu nedenle bu çalışmada da çekirdek işlevi olarak GKM-KL çekirdeği kullanılmıştır. μ^a ve μ^b ile verilen iki GKM süpervektörü için GKM KL çekirdeği aşağıdaki eşitlikle tanımlanır (Campbell ve ark., 2006).

$$\begin{aligned} K(utt_a, utt_b) &= \sum_{i=1}^M w_i \mu_i^a \Sigma_i^{-1} \mu_i^b \\ &= \sum_{i=1}^M \left(\sqrt{w_i} \Sigma_i^{-\frac{1}{2}} \mu_i^a \right)^t \left(\sqrt{w_i} \Sigma_i^{-\frac{1}{2}} \mu_i^b \right) \end{aligned} \quad (6)$$

Burada μ_i^a ve μ_i^b iki Gauss karışım modelinin i 'inci bileşenlerinin ortalamalarını gösterir. w_i ve Σ_i sırasıyla i 'inci Gauss bileşeninin ağırlık ve kovaryans matrisidir ve genellikle GAM'dan doğrudan alınır. Çalışmada DVM sınıflandırıcısının eğitimi sırasında ızgara tarama yöntemine göre parametre optimizasyonu yapılmıştır. GKM KL çekirdeğinin kullanıldığı durumda belirlenmesi gereken tek parametre ceza parametresidir. Bu nedenle çalışmada yalnızca ceza parametresinin optimizasyonu yapılmıştır. Ceza parametresi -7 ile 15 arasında taranmış ve elde edilen sonuçlara göre en uygun ceza parametresi belirlenerek DVM eğitimi tamamlanmıştır.

BULGULAR VE TARTIŞMA

Çalışmada GKM süpervektör DVM yaklaşımına dayalı olarak bir yaş ve cinsiyet sınıflandırma sistemi geliştirilmiş ve bu sistem üzerinde çeşitli testler yapılmıştır. aGender veri kümesi kullanılarak gerçekleştirilen bu testlerde MFKK özniteliklerine eklenen zamansal türev (delta ve delta-delta) ve logaritmik enerji parametrelerinin yaş ve cinsiyet tanıma üzerindeki etkileri araştırılmıştır. Çalışmada MFKK katsayıları ile delta türev ve logaritmik enerji parametrelerinin farklı birleşimlerinden oluşan 8 farklı öznitelik vektörü kullanılmıştır (Çizelge 1). MFKK sayısı 8 ile 20 (4'er aralıklarla), GKM bileşen sayısı ise 32 ile 256 arasında (ikinin üstel katlarında) değiştirilerek toplam 128 (8x4x4) durum test edilmiştir.

Çizelge 1. Çalışmada kullanılan öznitelik vektörleri

Öznitelik birleşimleri	Sembolik gösterimi
MFKK	M
MFKK+log Enerji	Me
MFKK+delta	Md
MFKK+delta-delta	MD
MFKK+log Enerji+delta	Med
MFKK+log Enerji+delta-delta	MeD
MFKK+delta+delta-delta	MdD
MFKK+log Enerji+delta+delta-delta	MedD

Çalışmada kullanılan aGender veritabanı 965 Alman konuşmacının farklı oturumlarda (en fazla 6) seslendirdiği toplam 65241 ses kaydından oluşmaktadır. Bu kayıtlar ana dili Almanca olan konuşmacıların bir ses portalını arayarak (sabit ve cep telefonu ile) seslendirdiği metin okumalarını ve bazı açık uçlu sorulara verdikleri cevapları içermektedir. Normal konuşma süresi yaklaşık 2 saniyedir. Ancak uzunluğu 3 ile 6 saniye arasında olan konuşmalar da vardır. aGender veritabanı eğitim, geliştirme ve test olmak üzere üç bölümden oluşmaktadır. Her bölümdeki konuşmacıların yaş ve cinsiyet gruplarına göre dağılımı yaklaşık eşittir. Eğitim bölümü 471 konuşmacının 32527 konuşmasından, geliştirme bölümü 299 konuşmacının 20549 konuşmasından, test bölümü ise 175 konuşmacının 12165 konuşmasından oluşmaktadır. Eğitim ve geliştirme bölümlerindeki konuşmaların hangi yaş ve cinsiyet sınıflarına ait oldukları veritabanında kayıtlıdır. Ancak test bölümündeki konuşmaların sınıf etiketleri veritabanında mevcut değildir. Eğitim ve geliştirme bölümlerindeki konuşmacıların yaş ve cinsiyet gruplarına göre dağılımları Çizelge 2'de gösterilmiştir.

Önerilen sistemin ilk aşmasında yaştan ve cinsiyetten bağımsız Genel Arkaplan Modelinin (GMA) eğitimi gerçekleştirilmiştir. Bu amaçla aGender veritabanının eğitim bölümünden rastgele seçilen 70 konuşmacının (her yaş grubundan 10'ar kişi) 4678 konuşması ile bir GKM (EM algoritması ile) eğitilmiş ve bu model GAM olarak kullanılmıştır. Daha sonra her konuşmaya karşılık gelen GKM'ler GAM'dan uyarlanarak oluşturulmuştur. MAP ile gerçekleştirilen uyarlama aşamasında GAM'ın yalnızca ortalama

bileşenleri uyarlanmış diğer parametreler (kovaryans ve ağırlık) GAM'dan doğrudan alınmıştır. Uyarlanan GKM'lerin ortalama bileşenleri uç uca eklenerek süpervektörlere dönüştürüldükten sonra DVM'ye uygulanmış ve gerçekleştirilen eğitim ve test aşamaları sonunda konuşmacıların yaş ve cinsiyet sınıflarına karar verilmiştir. DVM sınıflandırıcısının eğitiminde aGender veritabanının eğitim bölümündeki 331 konuşmacı tarafından 1571 oturumda seslendirilen konuşmalar kullanılırken, test aşamasında ise geliştirme bölümündeki 299 konuşmacı tarafından 1388 oturumda seslendirilen konuşmalar kullanılmıştır. Her oturumda gerçekleştirilen ses kayıtlarının tamamı tek bir konuşma olarak kabul edilmiş ve her oturum için birer GKM modeli üretilmiştir.

Çizelge 2. Yaş ve cinsiyet grupları ile birlikte eğitim ve geliştirme bölümündeki konuşmacı/konuşma sayıları.

Sınıf	Yaş ve Cinsiyet Grubu	Yaş	#Eğitim	#Geliştirme
1	Çocuk (Ç)	7-14	68/4406	38/2396
2	Genç Kadın (GK)	15-24	63/4638	36/2722
3	Genç Erkek (GE)	15-24	55/4019	33/2170
4	Yetişkin Kadın (YeK)	25-54	69/4573	44/3361
5	Yetişkin Erkek (YeE)	25-54	66/4417	41/2512
6	Yaşlı Kadın (YaK)	55-80	72/4924	51/3561
7	Yaşlı Erkek (YaE)	55-80	78/5549	56/3826

Farklı GKM bileşen sayıları (32, 64, 128 ve 256) kullanılarak geliştirilen yaş ve cinsiyet sınıflandırma sistemleri için elde edilen doğruluk oranları ayrı çizelgeler halinde aşağıda verilmiştir (Çizelge 3 – Çizelge 6). Bu çizelgeler GKM bileşen sayısı sabit olmak üzere 4 farklı MFKK sayısı (8, 12, 16 ve 20) ve 8 farklı öznitelik vektörü ile yapılan test sonuçlarını içermektedir. Öznitelik sütunundaki “M” MFKK katsayılarını, “e” logaritmik enerji parametresini, “d” delta parametresini, “D” ise delta-delta parametresini temsil etmektedir. Bu durumda yalnızca MFKK katsayıları ile oluşturulan öznitelik vektörü “M” ile MFKK, log enerji, delta ve delta-delta parametreleri ile oluşturulan öznitelik vektörü ise “MedD” ile gösterilmiştir.

Çizelge 3. 32 bileşenli GKM ile geliştirilen yaş ve cinsiyet sınıflandırma sisteminin doğruluğu

Öznitelik vektörü	MFKK-8	MFKK-12	MFKK-16	MFKK-20
M	46.61%	51.08%	50.22%	50.50%
Me	48.27%	51.22%	52.38%	52.95%
Md	52.09%	54.03%	54.18%	54.47%
MD	51.87%	53.03%	53.24%	53.89%
Med	49.71%	54.54%	53.31%	55.12%
MeD	50.79%	54.18%	55.12%	54.83%
MdD	53.03%	55.91%	56.84%	56.70%
MedD	52.74%	55.91%	54.76%	55.48%

Çizelge 4. 64 bileşenli GKM ile geliştirilen yaş ve cinsiyet sınıflandırma sisteminin doğruluğu

Öznitelik vektörü	MFKK-8	MFKK-12	MFKK-16	MFKK-20
M	47.05%	51.15%	51.51%	52.38%
Me	47.77%	52.09%	52.88%	53.60%
Md	51.66%	54.61%	54.83%	54.90%
MD	51.95%	55.33%	53.60%	54.25%
Med	53.96%	58.07%	56.56%	56.12%
MeD	52.31%	54.83%	55.76%	54.83%
MdD	54.97%	56.63%	56.63%	56.05%
MedD	55.19%	58.72%	57.64%	57.49%

Çizelge 5. 128 bileşenli GKM ile geliştirilen yaş ve cinsiyet sınıflandırma sisteminin doğruluğu

Öznitelik vektörü	MFKK-8	MFKK-12	MFKK-16	MFKK-20
M	46.11%	51.66%	52.02%	52.74%
Me	47.26%	52.31%	54.03%	53.60%
Md	53.31%	55.69%	56.34%	55.62%
MD	51.08%	54.76%	54.61%	55.04%
Med	54.32%	58.14%	57.56%	57.35%
MeD	52.88%	55.48%	55.76%	56.77%
MdD	54.61%	57.85%	56.63%	57.20%
MedD	56.63%	60.23%	58.21%	58.65%

Çizelge 6. 256 bileşenli GKM ile geliştirilen yaş ve cinsiyet sınıflandırma sisteminin doğruluğu

Öznitelik vektörü	MFKK-8	MFKK-12	MFKK-16	MFKK-20
M	45.75%	50.50%	52.09%	52.38%
Me	48.49%	52.52%	54.68%	54.68%
Md	53.10%	55.91%	56.56%	56.56%
MD	51.44%	53.82%	54.54%	54.68%
Med	54.39%	58.07%	58.29%	58.07%
MeD	53.53%	55.91%	56.05%	57.20%
MdD	54.68%	57.64%	56.92%	58.00%
MedD	56.84%	60.16%	58.93%	59.08%

Elde edilen sonuçlardan GKM bileşen sayısı arttıkça öznitelik türünden ve boyutundan bağımsız olarak sistemin başarısının arttığı görülmektedir. Ancak bu artış öznitelik türüne ve boyutuna göre değişmektedir ve belli bir değerden sonra başarı artışı olmamaktadır. Örneğin öznitelik olarak 8 MFKK'nın kullanıldığı durumda en yüksek başarı 64 bileşenli GKM'lerle elde edilirken, 20 MFKK ve diğer parametrelerin birlikte kullanıldığı durumda ("MedD") en yüksek başarı 256 bileşenli GKM'lerle elde edilmiştir. Yani öznitelik boyutu arttıkça bu vektörleri modellemek için gereken GKM bileşen sayısı da artmıştır.

Geliştirilen sistemin başarısını etkileyen bir diğer faktör de öznitelik olarak kullanılan MFKK sayısıdır. GKM bileşen sayısı gibi MFKK sayısının artırılması da sistemin başarısını arttırmış, ancak belli bir değerden sonra başarı artışı durmuş ve düşüşler görülmüştür. Bu sınır kullanılan öznitelik vektörüne göre değişiklik göstermektedir. Öznitelik vektörlerinin yalnızca MFKK katsayıları ile oluşturulduğu durumlarda ("M") en yüksek başarı 20 MFKK ile elde edilirken, log enerji ve delta parametrelerinin birlikte kullanıldığı durumda ("MedD") ise en yüksek başarı 12 MFKK'lı özniteliklerle sağlanmıştır. Yani MFKK sayısını artırmak yerine, düşük sayıda MFKK'ya log enerji ve delta parametreleri dahil edilerek oluşturulan özniteliklerin başarısı daha yüksek olmuştur.

Öznitelik vektörlerine eklenen log enerji, delta ve delta-delta parametrelerinin sistemin başarısını farklı oranlarda arttırdığı yapılan test sonuçlarından görülmektedir. Bu parametrelerden log enerji, sistemin başarısını en az etkileyen parametre olurken, bu parametreyi sırasıyla delta-delta ve delta parametreleri takip etmiştir. En yüksek başarı oranı 12 MFKK'ya log enerji, delta ve delta-delta parametrelerinin eklenmesiyle oluşturulan özniteliklerin ("MedD"), 128 bileşenli GKM'lerle modellenmesi sonucunda %60.23 olarak elde edilmiştir. En yüksek başarı oranının elde edildiği durum için sistemin karışıklık matrisi Çizelge 7'de verilmiştir. Her sınıfın bireysel sınıflandırma oranını ve sınıflar arası geçişleri ayrıntılı olarak gösteren karışıklık matrisi, sınıflandırma sistemlerinin performansını değerlendirmede yaygın olarak kullanılan bir araçtır.

Çizelge 7. En yüksek başarının elde edildiği durum için (MFKK sayısı:12, Öznitelik türü: “MedD”, GKM bileşen sayısı: 128) sistemin karışıklık matrisi.

	Ç	GK	GE	YeK	YeE	YaK	YaE
Ç	62.96%	13.58%	6.17%	8.02%	2.47%	6.79%	0.00%
GK	11.30%	69.49%	0.00%	13.56%	0.00%	5.65%	0.00%
GE	1.36%	1.36%	59.86%	0.00%	21.77%	2.04%	13.61%
YeK	2.23%	21.43%	0.00%	55.80%	0.45%	20.09%	0.00%
YeE	0.59%	0.00%	25.44%	2.37%	43.20%	0.59%	27.81%
YaK	1.22%	10.61%	0.00%	26.94%	0.00%	60.82%	0.41%
YaE	0.00%	0.00%	7.20%	0.00%	23.86%	2.27%	66.67%

Geliştirilen yaş ve cinsiyet sınıflandırma sisteminde, en yüksek başarı ile sınıflandırılan yaş ve cinsiyet sınıfı %69.49 ile genç kadın (GK) sınıfı olmuştur. Bu sınıfı sırasıyla %66.67 ile yaşlı erkek (YaE), %62.96 ile çocuk (Ç), %60.82 ile yaşlı kadın (YaK), %59.86 ile genç erkek (GE), %55.80 ile yetişkin kadın (YeK) ve %43.20 ile yetişkin erkek (YeE) sınıfları izlemiştir. Test edilen 1388 konuşmadan 836 tanesi doğru sınıflandırılırken, 552 tanesi ise hatalı sınıflandırılmıştır. Hatalı kararların %80.25'i (443 tanesi) cinsiyeti aynı olan konuşmacıların yaş grupları arasında, %3.26'si (18 tanesi) yaş grubu aynı olan konuşmacıların cinsiyet grupları arasında, %16.49'u (91 tanesi) ise çocuk grubu ile diğer gruplar arasında olmuştur. Çocuk grubunun dahil olduğu 91 hatalı kararın 74 tanesi (%81.32) çocuk ile kadın yaş grupları arasında, 17 tanesi (%18.68) ise çocuk ile erkek yaş grupları arasında görülmüştür. En yüksek karışıklık %27.81 ile yetişkin erkek (YeE) ile yaşlı erkek (YaE) grubu arasında görülmüştür. Bireysel sınıflandırılma oranı en düşük grup olan yetişkin erkek grubuna ait 169 test konuşmasının 1 tanesi çocuk (%0.59), 43 tanesi genç erkek (%25.44), 4 tanesi yetişkin kadın (%2.37), 73 tanesi yetişkin erkek (%43.20), 1 tanesi yaşlı kadın (%0.59) ve 47 tanesi de yaşlı erkek (%27.81) olarak sınıflandırılmıştır.

SONUÇ

Bu çalışmada MFKK özniteliklerine eklenen log enerji ve delta parametrelerinin (delta ve delta-delta) otomatik yaş ve cinsiyet tanıma üzerindeki etkileri araştırılmıştır. Bu amaçla GKM süpervektörlerinin DVM ile sınıflandırıldığı bir sistem, MFKK ile log enerji ve delta parametrelerinin olası birleşimlerini içeren 8 farklı öznitelik vektörü kullanılarak gerçekleştirilmiştir. Daha sonra bu sistemler üzerinde testler yapılarak MFKK özniteliklerine eklenen log enerji ve delta parametrelerinin başarı üzerindeki etkileri incelenmiştir. aGender veritabanının geliştirme bölümündeki 299 konuşmacının 1388 oturumda seslendirdiği konuşmalarla yapılan testlerde, MFKK'ye eklenen üç parametrenin de sistemin başarısını arttırdığı görülmüştür. Bu parametrelerden log enerjinin başarı üzerindeki etkisi %1.17 olurken, delta-delta parametresinin etkisi %3.24, delta parametresinin etkisi ise yaklaşık %4.61 olmuştur. Bu sonuçlar dört farklı MFKK sayısı (8, 12, 16 ve 20) ile elde edilen sonuçların ortalaması olup, 12 MFKK'nın kullanıldığı durumda öznitelik kümesine log enerjinin eklenmesi ile %0.65, delta-delta parametresinin eklenmesi ile %3.10 ve delta parametresinin eklenmesiyle de %4.03 başarı artışı sağlanmıştır. Çalışmada ayrıca öznitelik olarak kullanılan kepsral katsayıların sayısı ile yaş ve cinsiyet modeli olarak kullanılan GKM'lerin bileşen sayısının başarı üzerindeki etkileri de araştırılmıştır. Bu amaçla MFKK sayısı 8 ile 20 arasında, GKM bileşen sayısı ise 32 ile 256 arasında

değiştirilerek testler yapılmıştır. Bu testlerde, en yüksek sınıflandırma oranı 12 MFKK'li "MedD" özniteliklerinin, 128 bileşenli GKM'lerle modellenmesi sonucunda %60.23 olarak belirlenmiştir. En yüksek başarı ile sınıflandırılan yaş ve cinsiyet grubu %69.49 ile genç kadın (GK) grubu olurken, en düşük başarı ile sınıflandırılan grup ise %43.20 ile yetişkin erkek (YeE) grubu olmuştur. Bu fark (%26.29) özellikle sınıflandırma oranı düşük olan gruplar için (yetişkin) farklı özniteliklere ihtiyaç olduğunu göstermektedir. Bu bağlamda perde, enerji, formant, konuşma hızı, süre gibi prosodik ve jitter, shimmer, harmonik gürültü oranı (HNR) gibi ses kalitesine dayalı özniteliklerin MFKK ile birlikte kullanımını düşünülebilir. Ayrıca LTSM, RNN ve CNN (Convolutional Neural Network) gibi derin öğrenme temelli yaklaşımların karmaşık ve yüksek boyutlu verilerdeki karmaşık yapıları tanımlamadaki üstün performansı göz önünde bulundurulduğunda bu yaklaşımların kullanımı ile konuşmaya dayalı yaş ve cinsiyet tanıma sistemlerinin performansında ciddi artışlar sağlanabileceği değerlendirilmiştir.

KAYNAKLAR

- Bahari MH, McLaren M, van Leeuwen DA, 2014. Speaker age estimation using i-vectors. *Engineering Applications of Artificial Intelligence*, 34: 99-108.
- Bocklet T, Maier A, Bauer JG, Burkhardt F, Noth E, 2008. Age and gender recognition for telephone applications based on gmm supervectors and support vector machines. In 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, 31 March-4 April, 2008, pp: 1605-1608.
- Campbell, WM, Sturim DE, Reynolds DA, 2006. Support vector machines using GMM supervectors for speaker verification. *IEEE signal processing letters*, 13(5): 308-311.
- Choukri M, Wu S, 2019. Age and Gender Classification for Permission Control of Mobile Devices in Tracking Systems. In International Conference on Artificial Intelligence for Communications and Networks, Harbin, May 25-26, 2019, pp: 318-324.
- Dempster A, Laird N, Rubin D, 1977. Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Stat. Soc.* 39:1-38.
- Dhonde SB, Chaudhari A, Jagade SM, 2017. Integration of mel-frequency cepstral coefficients with log energy and temporal derivatives for text-independent speaker identification. In Proceedings of the International Conference on Data Engineering and Communication Technology, 2017: pp: 791-797
- Ertam F, 2019. An effective gender recognition approach using voice data via deeper LSTM networks. *Applied Acoustics*, 156: 351-358.
- Fang SH, Tsao Y, Hsiao MJ, Chen JY, Lai YH, Lin FC, Wang CT, 2019. Detection of pathological voice using cepstrum vectors: A deep learning approach. *Journal of Voice*, 33(5): 634-641.
- Grzybowska J, Kacprzak S, 2016. Speaker Age Classification and Regression Using i-Vectors. In INTERSPEECH 2016, San Francisco, September 8-12, 2016, pp: 1402-1406.
- Kerkeni L, Serrestou Y, Mbarki M, Raoof K, Mahjoub, MA, 2018. Speech Emotion Recognition: Methods and Cases Study. In ICAART, January 16-18, 2018, pp: 175-182.
- Koo H, Jeong S, Yoon S, Kim W, 2020. Development of Speech Emotion Recognition Algorithm using MFCC and Prosody. In 2020 International Conference on Electronics, Information, and Communication (ICEIC), January 19-22, 2020, pp: 1-4.
- Li M, Han KJ, Narayanan S, 2013. Automatic speaker age and gender recognition using acoustic and prosodic level information fusion. *Computer Speech & Language*, 27(1): 151-167.

- Mallouh AA, Qawaqneh Z, Barkana BD, 2017. Combining two different DNN architectures for classifying speaker's age and gender. In International Conference on Bio-inspired Systems and Signal Processing, Porto, February 21-23, 2017, pp: 112-117.
- Meinedo H, Trancoso I, 2010. Age and gender classification using fusion of acoustic and prosodic features. In Eleventh Annual Conference of the International Speech Communication Association, Makuhari, September 26-30, 2010, pp: 2818-2821.
- Metze F, Ajmera J, Englert R, Bub U, Burkhardt F, Stegmann J, Müller C, Huber R, Andrassy B, Bauer JG, Littel B, 2007. Comparison of four approaches to age and gender recognition for telephone applications. In 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, Honolulu, April 15-20, 2007, pp: IV-1089-IV-1092I.
- Rabiner L, Juang BH, Yegnanarayana B, 2008. Fundamentals of Speech Recognition, Pearson Education, London.
- Rao KS, Manjunath KE, 2017. Speech recognition using articulatory and excitation source features. Springer. (Appendix A MFCC Features)
- Reynolds DA, Quatieri TF, Dunn RB, 2000. Speaker verification using adapted Gaussian mixture models. Digital Signal Processing, 10(1-3), 19-41.
- Safavi S, Russell M, Jančovič P, 2018. Automatic speaker, age-group and gender identification from children's speech. Computer Speech & Language, 50: 141-156.
- Son G, Kwon S, Park N, 2019. Gender classification based on the non-lexical cues of emergency calls with recurrent neural networks (RNN). Symmetry, 11(4): 525.
- van Heerden C, Barnard E, Davel M, van der Walt C, van Dyk E, Feld M, Müller C, 2010. Combining regression and classification methods for improving automatic speaker age recognition. In 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, March 14-19, 2010, pp: 5174-5177.
- Yücesoy E, 2020. Konuşmacının Yaş ve Cinsiyetine Göre Sınıflandırılmasında DVM Çekirdeğinin Etkisi. El-Cezeri Journal of Science and Engineering, 7(3):970-982.
- Zazo R, Nidadavolu PS, Chen N, Gonzalez-Rodriguez J, Dehak N, 2018. Age estimation in short speech utterances based on LSTM recurrent neural networks. IEEE Access, 6: 22524-22530.