



Kahramanmaraş Sütçü İmam University

Journal of Engineering Sciences



Geliş Tarihi : 23.07.2023
Kabul Tarihi : 28.09.2023

Received Date : 23.07.2023
Accepted Date : 28.09.2023

YÜKSEK ÇÖZÜNÜRLÜKLÜ SPEKTROGRAM GÖRÜNTÜLERİNDEN AKAZE YÖNTEMİ İLE SES SAHTECİLİĞİ TESPİTİ

AUDIO FORGERY DETECTION FROM HIGH-RESOLUTION SPECTROGRAM WITH AKAZE METHOD

Beste ÜSTÜBİOĞLU^{1*} (ORCID: 0000-0001-7451-0634)
Gül TAHAOĞLU¹ (ORCID: 0000-0000-0002-8828-5674)

¹ Karadeniz Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü, Trabzon, Türkiye

*Sorumlu Yazar / Corresponding Author: Beste ÜSTÜBİOĞLU

ÖZET

Ses sahteciliği alanında yaygın olarak kullanılan Kopyala-yapıştır sahteciliği, ses içerisindeki bir sesli kısmın kopyalanıp yine aynı ses içerisinde farklı bir konuma yapıştırılmasıyla oluşturulmaktadır. Gelişmiş ses yazılımları sayesinde bu tür bir sahteciliğin uygulanması oldukça kolay olmakla birlikte, saldırganlar tarafından sahtecilik izlerini gizlemek için sahte sese uygulanan son işlem operasyonları bu sahtecilik tespitini oldukça zor hale getirmektedir. Bu amaçla, sestten elde edilen yüksek çözünürlüklü spektrogram görüntüsü üzerinde anahtar nokta tabanlı bir yaklaşım kullanarak, ses kopyala-yapıştır sahteciliğini tespit eden son işlem operasyonlarına dayanıklı yeni bir yöntem önerilmiştir. Önerilen yöntemde öncelikle ses dosyasından yüksek çözünürlüklü spektrogram görüntüsü elde edilir. Ardından, Akaze yöntemi ile spektrogram görüntüsünden anahtar noktalar ve özellik tanımlayıcıları çıkarılmaktadır. Çıkarılan özellikler g2NN algoritması ile eşleştirilmektedir. Spektrogram üzerindeki noktaların ses üzerine iz düşürülmesiyle de ses kopyala-yapıştır sahteciliği tespit edilmektedir. Elde edilen sonuçlar önerilen yöntemin son işlem operasyonları uygulansa dahi literatürdeki çalışmalarla kıyaslandığında çok yüksek doğrulukla ses kopyala-yapıştır sahteciliği tespitini yaptığını göstermektedir.

Anahtar Kelimeler: Akaze, Spektrogram, Ses sahteciliği.

ABSTRACT

Copy-paste forgery, which is widely used in the field of audio forgery, is created by copying an audio part in the audio and pasting it in a different location in the same audio. While this type of forgery is quite easy to implement thanks to advanced audio software, post-processing operations applied to forged audio by attackers to hide traces of forgery make this forgery detection extremely difficult. For this purpose, a new post-processing-robust method for detecting audio copy-paste forgery using a key point-based approach on the high-resolution spectrogram image obtained from the audio is proposed. In the proposed method, firstly, a high-resolution spectrogram image is obtained from the audio file. Then, with the Akaze method, key points, and feature descriptors are extracted from the spectrogram image. Extracted features are matched with the g2NN algorithm. Audio copy-paste forgery is detected by tracing the key points on the spectrogram onto the audio. The results obtained show that the proposed method detects audio copy-paste forgery with very high accuracy when compared to the studies in the literature, even if post-processing operations are applied.

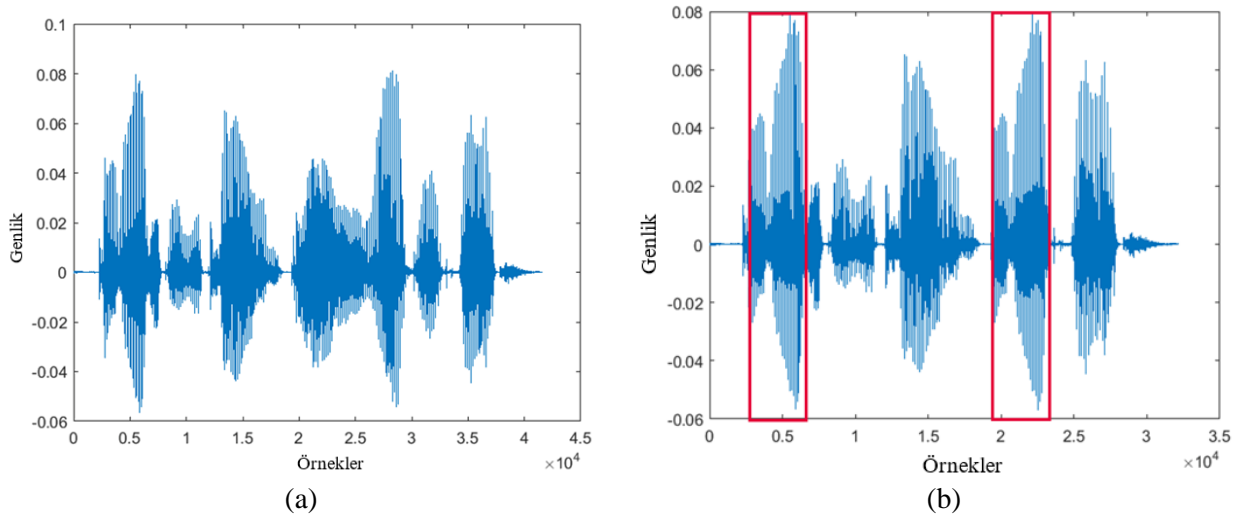
Keywords: Akaze, Spectrogram, Audio forgery.

ToCite: ÜSTÜBİOĞLU, B., & TAHAOĞLU, G., (2023). YÜKSEK ÇÖZÜNÜRLÜKLÜ SPEKTROGRAM GÖRÜNTÜLERİNDEN AKAZE YÖNTEMİ İLE SES SAHTECİLİĞİ TESPİTİ. *Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi*, 26(4), 962-971.

GİRİŞ

Dijital ses kayıtları telefon kayıtları, sesli mesajlar, müzik dosyaları gibi günlük hayatımızın her yerinde karşımıza çıkmaktadır. Dijital ses kayıtlarının hukuk davalarında delil olarak kullanıldığı ve hatta bazı davalarda bu delillerin kilit rol oynadığı düşünülürse, ses kayıtlarının sahte olup olmadığının tespiti büyük önem arz etmektedir.

Gelişmiş ses düzenleme yazılımlarının kolay kullanılabilirliği sayesinde, bir saldırgan uzman sahtecilik yöntemleri ile sahte ses kayıtları oluşturabilmektedir. Bu sahtecilik yöntemlerinden bir tanesi de Kopyala-yapıştır sahteciliğidir. Kopyala-yapıştır sahteciliğinde, ses kaydı içerisindeki bir kelime kopyalanıp yine aynı konuşma kaydı içerisinde başka bir kelime üzerine yapıştırılmasıyla gerçekleştirilmektedir. Şekil 1’de kopyala-yapıştır sahteciliğine örnek bir ses kaydı verilmiştir. Sahte ses oluşturmak için Şekil 1(a)’da TIMIT veri tabanından alınan orijinal konuşma kaydındaki (“sa2.wav”) birinci sözcük kopyalanmış ve aynı konuşma kaydında dördüncü sözcüğün yerine yapıştırılarak Şekil 1(b)’deki sahte ses elde edilmiştir.



Şekil 1. a. Orijinal Ses b. Kopyala-yapıştır Sahteciliği ile Oluşturulmuş Sahte Ses.

Şekilden de görüleceği üzere diğer ses sahteciliği türlerinden farklı olarak kopyala yapıştır sahteciliğinde kopyalanan kısım farklı bir ses kaydına değil, aynı ses kaydı içerisindeki başka bir kısma yapıştırılmaktadır. Bu durumda, sesle ilgili temel özellikler (genlik, frekans, uzunluk, gürültü, ton) aynı ses kaydındaki kopyala-yapıştır yapılmış kelimeler için değişmeyecektir. Ses kaydının bir uzman tarafından dilbilimsel veya ses bilimsel olarak manuel bir şekilde analizinin yapılmadığı durumda sesin otomatik olarak orijinal veya sahte olarak değerlendirilmesi zor hale gelecektir. Kopyala yapıştır sahteciliğinin tespitini zorlaştıran bir başka durum ise saldırgan tarafından sahtecilik izlerini gizlemek için uygulanan son işlem operasyonlarıdır (sıkıştırma, gürültü ekleme, medyan filtreleme vb.). Çünkü son işlem operasyonları sesin dalga biçimini değiştirmekte bu da sahtecilik tespitini zorlaştırmaktadır.

Bir ses kaydının bir manipülasyona maruz kalıp kalmadığının değerlendirilmesinde literatürde fen bilimleri ve sosyal bilimleri olmak üzere farklı disiplinlerin bakış açılarıyla da ele alınması gereken önemli bir problem olarak görülmüştür (Burucu, 2023). Literatürde ses sinyaline ilişkin özelliklerin algoritmik olarak değerlendirilmesi ile kopyala-yapıştır sahteciliği tespiti alanında yapılan çalışmalar VAD (ses aktivite tespiti) ve pencere tabanlı yöntemler olarak iki kategoriye ayrılmaktadır. VAD tabanlı yöntemlerde; önerilen VAD algoritması ile ses dosyası sesli ve sessiz kısımlara ayrılmakta sonrasında bu sesli kısımlardan öznelikler çıkarılmaktadır. Çıkarılan özelliklerin benzerliklerine göre de sesteki tekrarlanan kelimeler tespit edilmektedir. Pencere tabanlı yöntemlerde ise sesteki sesli/sessiz kısımlar çıkartılmak yerine, ses dosyası eşit uzunluktaki örtüşen ve örtüşmeyen pencerelere bölünmekte ve pencerelerden özellik çıkarılmaktadır. Benzer özelliklere sahip pencereler, tekrarlanan pencereler olarak işaretlenmektedir.

Pencere tabanlı yöntemler: (Xiao vd., 2014), seste tekrarlanan kelimeleri tespit etmek için ses dosyasını T süreli pencerelere bölmüştür. Önerdikleri benzerlik fonksiyonu ile bu çerçeveler arasındaki benzerliğe göre tekrarlanan kelimeleri işaretlemiştir.

VAD tabanlı yöntemler: Huang vd. (Huang vd., 2020), öncelikli olarak belirledikleri VAD algoritmasıyla ses dosyasının sesli kısımlarını belirlemiştir. Ayrık Fourier dönüşümü (DFT) ve Mel-Frekans Cepstral Katsayıları

(MFCC) yöntemlerinden faydalanarak bu sesli kısımlardan özellik çıkarmışlardır. Benzer özelliklere göre kopyala-yapıştır yapılmış kelimeyi tespit etmişlerdir. Wang vd. (Wang vd., 2010) Tekil Değer Ayırıştırma (SVD) dönüşümüne dayalı bir yöntem önermiştir. Sesli kısımlardan özellik olarak Ayrık Kosinüs Dönüşümü (DCT) katsayıları elde edilmiştir. Sonrasında, belirli bir özvektörü elde etmek için DCT katsayılarının kare matrisine tekil değer ayırıştırma (SVD) dönüşümünü uygulamışlardır.

(Xie vd., 2018) birden çok özellik kullanarak karar ağacı yaklaşımı ile ses kopyala-yapıştır sahteciliğini tespit eden bir yöntem önermişlerdir. Yöntemde, sesli kısımlardan Gamaton, Mel-frekans cepstral katsayıları (MFCC), perde (pitch) ve ayrık Fourier dönüşümü (DFT) katsayıları olmak üzere 4 ayrı özellik çıkarılmış ve bu özelliklerin benzerliklerine göre dört ayrı tespit sonucu elde edilmiştir. Tespit sonuçlarını birleştirmek için ise C4.5 karar ağacını kullanılmıştır.

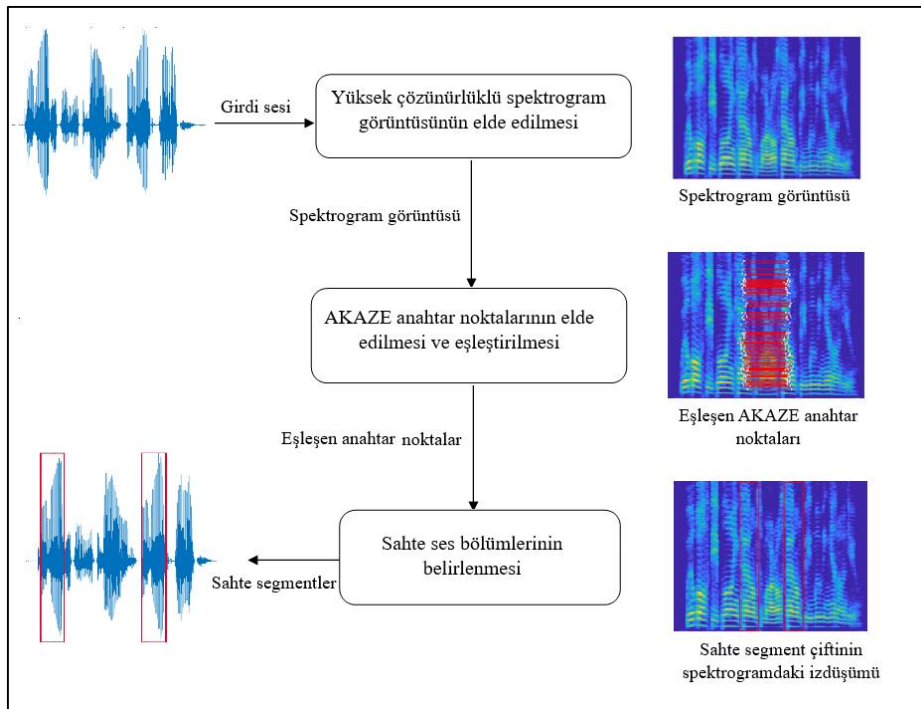
(İmran vd. 2017), ses dosyasındaki sesli kısımları belirlemek için bir VAD yöntemi önermişler ve bu kısımlardan özellik çıkarmak için 1 boyutlu yerel ikili model (LBP) yöntemini uygulamışlardır. Yöntemde tekrarlanan kelimelerin bulunması için sesli kısımlardan çıkarılan LBP histogramları birbirleri ile kıyaslanmıştır.

(Yan vd., 2014), sestem YAAPT yöntemi ile temel frekans dizisini çıkarmıştır. Temel frekans dizilerinin benzerliklerini hesabında ise ortalama mesafe ve faz korelasyonu yöntemlerinden faydalanılmıştır. Yan vd. (Yan vd., 2019) yaptığı bir diğer çalışmada perde dizileri yanında format dizileri de kullanılmıştır. Çıkarılan özelliklerin benzerlik hesabında ise Dinamik zaman atlama (DTW) yönteminden faydalanılarak sesteki kopyala-yapıştır sahteciliği tespit edilmiştir.

Bu çalışmada, son işlem operasyonlarına dayanıklı yeni bir ses kopyala-yapıştır sahteciliği tespit yöntemi önerilmiştir. Önerilen yöntemde, şüpheli sesteki sahte kısımları tespit etmek için giriş sesinin yüksek çözünürlüklü spektrogram görüntüsünden faydalanılmaktadır. Bunun için öncelikle Akaze yöntemi ile spektrogram görüntüsünden anahtar noktalar ve özellik tanımlayıcıları çıkarılmakta ve çıkartılan özellikler g2NN algoritması ile eşleştirilmektedir. Yanlış eşleşmeler elendikten sonra kalan eşleşmeler spektrogram üzerinde tekrarlanan kelimelere karşılık gelmektedir. Spektrogram üzerindeki noktaların ses üzerine iz düşürülmesiyle de ses kopyala-yapıştır sahteciliği tespit edilmektedir. Elde edilen sonuçlar önerilen yöntemin literatürdeki çalışmalarla kıyaslandığında çok yüksek doğrulukla sahtecilik tespiti yaptığını göstermektedir.

MATERYAL VE YÖNTEM

Bu çalışmada ataklara karşı dayanıklı yeni bir kopyala-yapıştır sahteciliği uygulanmış seslerin tespit yöntemi sunulmaktadır. Şekil 2’de görüldüğü gibi önerilen yöntem üç aşamadan oluşmaktadır. İlk olarak, giriş sesinin yüksek çözünürlüklü spektrogram temsili elde edilmektedir. Sesteki tekrarlanan bölümlerin aranması, elde edilen spektrogram görüntüsü aracılığıyla gerçekleştirilmektedir. İkinci aşamada ise spektrogram görüntüsünün tamamından anahtar noktaları elde edilmekte ve eşleştirilmektedir.



Şekil 2. Önerilen Yöntem

Önerilen yöntemin ataklara karşı dayanıklılığı sağlamak için anahtar nokta çıkarımında AKAZE yönteminden faydalanılmıştır. Anahtar nokta çıkarımından sonra anahtar noktalarının özelliklerini barındıran özellik tanımlayıcı vektörler kullanılarak en benzerlerinin eşleştirilmesi sağlanmıştır. Olası hatalı eşleşmelerin elenmesi de yine bu aşamada gerçekleştirilmektedir. Üçüncü aşama olan son aşamada ise, eşleşen kaynak ve hedef bölgelerde yer alan anahtar noktalarının x koordinatlarından faydalanılmıştır. Spektrogram üzerindeki kaynak ve hedefteki noktaların x koordinatlarının orta noktaları, sese iz düşürülerek seste kopyala-yapıştır yapılan segmentler tespit edilmiştir. Bu aşamalarda kullanılan yaklaşımların detayları şu alt başlıklar halinde sunulmuştur;

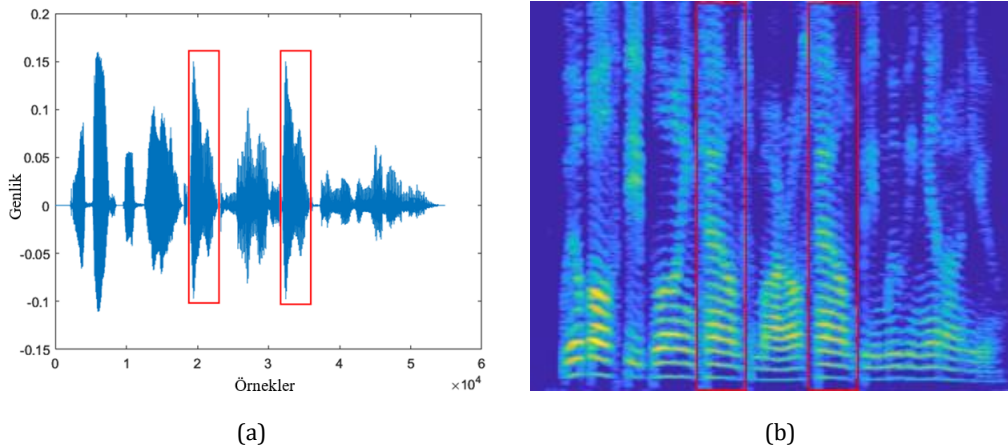
- i. Yüksek çözünürlüklü spektrogram görüntülerinin elde edilmesi,
- ii. AKAZE anahtar noktalarının elde edilmesi ve eşleştirilmesi,
- iii. Sahte ses segmentlerinin belirlenmesi.

Yüksek Çözünürlüklü Spektrogram Görüntülerinin Elde Edilmesi

Ses sinyalinin görsel temsilinde, Kısa süreli Fourier dönüşümü (Short Time Fourier Transform, STFT) tabanlı spektrogram çıkarma yaklaşımı en yaygın kullanılan yaklaşımlardan biridir. STFT tabanlı yaklaşımlar, esas olarak zaman ve frekans çözünürlükleri arasındaki değişiklikleri kullanırlar. Bu çalışmada hem zaman hem de frekansta aynı anda yüksek çözünürlük sağlayan spektrogram çıkarma tekniğini kullanılmıştır (Nam vd., 2017). Yöntem, Olasılıksal Gizli Bileşen Analizi' ni (Probabilistic Hidden Component Analysis, PLCA), biri yüksek zaman çözünürlüğüne sahip, diğeri ise Eşleştirilmiş PLCA (Coupled PLCA) adlı yüksek frekans çözünürlüğüne sahip iki spektrogramı birlikte ayırtmak için genişletir. Bu ayırtma adımı sayesinde zaman ve frekans açısından yüksek çözünürlük oluşturulmaktadır. PLCA, bir spektrogramı negatif olmayan spektral ve zamansal bileşenlerin dış çarpımlarının bir toplamına ayırtmak için kullanılır. V_{ft} 'nin temsil edildiğini varsayarsak, f , t 'de (1)'de verildiği gibi elde edilir.

$$V_{ft} = \gamma \sum_z P(z)P(f | z)P(t | z) \quad (1)$$

Burada, $P(f|z)$ spektral bileşenleri, $P(t|z)$ ise karşılık gelen spektral bileşenlerin zaman içinde varlığını belirten zamansal bileşenleri temsil etmektedir. Spektrogram esas olarak çok terimli dağılımların parametreleri tahmin edilerek ayırtılabilir. Bu tahmin, Beklenti Maksimizasyonu (EM) algoritması kullanılarak gerçekleştirilir. PLCA, yalnızca zaman veya frekans alanı için kullanılır, her ikisi için kullanılamaz. Yüksek frekans çözünürlüklü spektrogramı hesaplamak için, Çiftli PLCA kullanılmaktadır (Nam vd., 2017). Örnek olarak Şekil 3'te kopyala-yapıştır sahteciliği uygulanmış sahte ses ve bu sestten elde edilmiş yüksek çözünürlüklü spektrogram görüntüsü verilmiştir. Şekil 3'te görülen kırmızı çerçeveler kopyala-yapıştır yapılmış segmentleri göstermektedir. Şekilden de görüldüğü üzere sahte sesin yüksek çözünürlüklü spektrogram görüntüsünde sahte segmentler oldukça kolay bir şekilde ayırt edilebilmektedir.



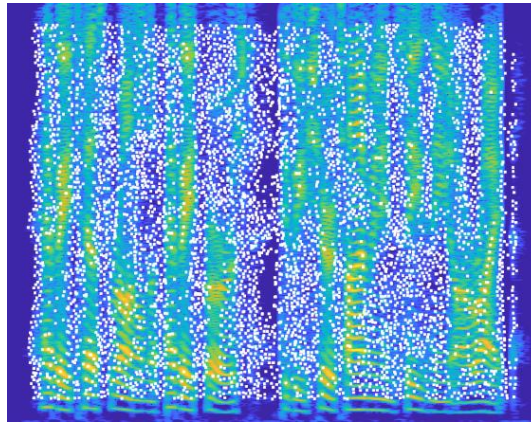
Şekil 3. a. Kopyala-yapıştır Sahteciliği Uygulanmış Sahte Ses. **b.** Sesten Elde Edilen Yüksek Çözünürlüklü Spektrogram.

AKAZE Algoritması ile Anahtar Noktalarının Elde Edilmesi ve Eşleştirilmesi

Spektrum görüntüsünden anahtar nokta çıkarımında AKAZE yönteminden faydalanılmıştır. Kullanılması önerilen bu yöntem doğrusal olmayan ölçek uzayı kullanarak anahtar noktaları elde etmektedir (Alcantarilla vd. 2013). AKAZE yöntemi temel olarak aşağıdaki adımları içermektedir.

- i. FED (Fast Explicit Diffusion) döngüsüne dayalı doğrusal olmayan ölçek uzayının oluşturulması,
- ii. Ölçeklendirilmiş Hessian matrisine dayalı anahtar noktalarının çıkarılması,
- iii. M-LDB(Modified Local Difference Binary) özellik tanımlayıcılarının elde edilmesi.

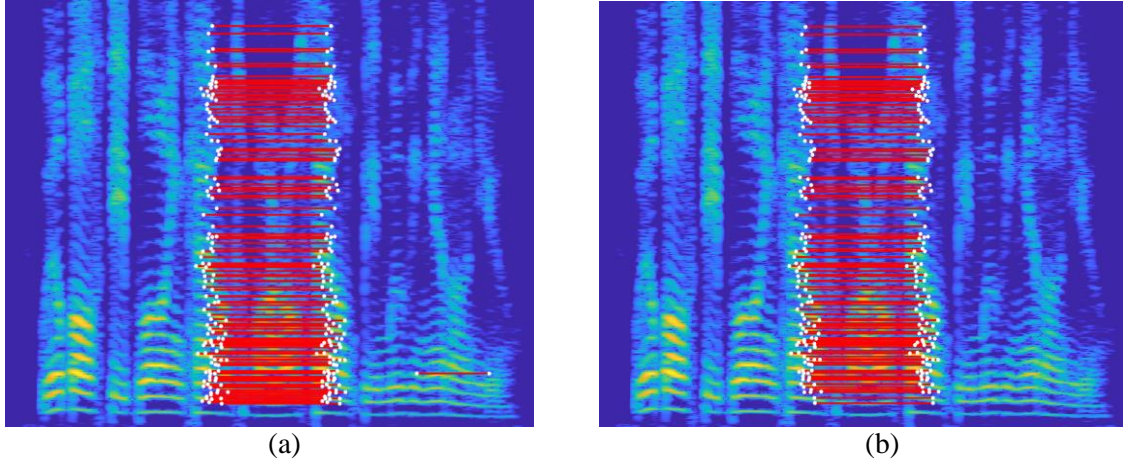
Anahtar nokta elde etme yaklaşımlarından olan SIFT ve SURF algoritmalarının temelinde çoklu Gauss ölçek uzayı kullanılmaktadır. Bu ölçek uzayı kullanımı adımı SIFT algoritmasında piramitsel yapı oluşturma amacı ile gerçekleştirilir. SURF yönteminde de kutu filtreleri aracılığıyla yaklaşık Gauss türevleri kullanılmaktadır. Bu iki yaklaşımda da Gauss bulanıklaştırması ile detaylar ve gürültü aynı seviyede bulanıklaştırılır. Bu problemi çözümlenebilmek için başlangıçta aynı yazarlar tarafından KAZE yöntemi önerilmiştir (Alcantarilla vd., 2012). Yöntem doğrusal olmayan difüzyon denklemlerini çözmek için Additive Operator Splitting (AOS) yaklaşımını kullanmaktadır. Ancak bu yaklaşım yüksek hesaplama karmaşıklığı problemini doğurmuştur. Bu yöntemi hızlandıran Fast Explicit Diffusion (FED) matematiksel yapısı kullanılarak Hızlandırılmış KAZE (AKAZE) olarak adlandırılan AKAZE yöntemi ortaya çıkmıştır (Alcantarilla vd., 2013). Elde edilen anahtar noktalarına ait özellik tanımlayıcıların elde edilmesinde Modified Local Difference Binary (MLDB) algoritması önerilmiştir. Algoritmanın icrası sonrası her bir anahtar noktası için 1x486 boyutunda özellik tanımlayıcı vektör elde edilmiş olur. Şekil 4'te, Şekil 3(b)'de verilen sahte sese ait spektrum görüntüsünden çıkarılan AKAZE anahtar noktaları gösterilmiştir. Bu görüntüde 5852 adet anahtar noktası elde edilmiştir.



Şekil 4. Sahte Sesten Elde Edilen Spektrumdan Çıkarılan Anahtar Noktalar.

Bir sonraki aşamada elde edilen anahtar noktalarına ait 486 boyutlu özellik tanımlayıcı vektörlerinin, $f_1 \dots f_n$, elde edilmesi gerçekleştirilmektedir. Bu özellik vektörlerinde anahtar noktasına ait koordinat bilgisi de yer almaktadır. Elde edilen bu vektörlerin eşleştirme işlemi (Amerini vd., 2011) önerilen hiyerarşik kümeleme yaklaşımı kullanılarak gerçekleştirilmektedir. Bu yaklaşıma göre mevcut f_i özellik vektörünün diğer özellik vektörleri ile arasındaki Öklid uzaklığı hesaplanmakta ve $D = d_1, d_2, \dots, d_{n-1}$ uzaklık vektörü oluşturulmaktadır. f_i vektörüne en yakın özellik vektörü ile ikinci en yakın özellik vektörü arasında bir T eşik değeri ile karşılaştırma yapılır. $d_1 / d_2 > T, T \in (0,1)$ şartını sağlayan özellik vektörleri eşleştirilir. Yapılan çalışmada T eşik değeri 0,6 olarak belirlenmiştir.

Olası hatalı eşleşmelerin giderilmesi için RANSAC (Random Sample Consensus yaklaşımından faydalanılmıştır. Bu yaklaşım ile yüksek oranda yanlış eşleşmelere sahip veri setindeki hatalar minimize edilir (Fischler vd., 1981). Üretilen modele ait parametrelerin tahmini için en az sayıda veri içeren küme kullanılır ve model ile tutarlı veriler ile bu küme genişletilerek aday çözümler özyinelemeli olarak üretilir. İterasyon tamamlanması ile üretilen modele uygun noktalar doğru eşleşme, uygun olmayan eşleşmeler ile hatalı eşleşme olarak değerlendirilir. Şekil 5'te kopyala yapıştır sahteciliği uygulanmış sahte sesin spektrum temsili üzerindeki anahtar nokta eşleşmelerinde hatalı işaretlemenin giderilmesi örneğine ilişkin görsel sunulmuştur. Şekil 5(a)' da siyah oval içine alınmış eşleşme hatalı eşleşmedir ve önerilen yaklaşımın uygulanması ile bu yanlış pozitif işaretlemenin giderildiği görülmektedir. Eleme sonrası eşleşmeler 5(b)'de sunulmuştur.



Şekil 5. Ransac ile Hatalı Eşleşen Noktanın Elenmesi Örneği **a.** Ransac Öncesi **b.** Ransac Sonrası.

Sahte Ses Parçalarının Belirlenmesi

Olası hatalı eşleşmelerin giderilmesi için RANSAC (Random Sample Consensus yaklaşımından faydalanılmıştır. Bu yaklaşım ile yüksek oranda yanlış eşleşmelere sahip veri setindeki hatalar minimize edilir (Fischler vd., 1981). Üretilecek modele ait parametrelerin tahmini için en az sayıda veri içeren Anahtar noktaların eşleştirilmesi ve hatalı eşleşmelerin elenmesinden sonra, spektrogram üzerindeki noktaların iki küme halinde öbeğlendiği görülmektedir. Bu iki öbek sesteki tekrarlanan kısımlara karşılık gelmektedir. Spektrogram üzerindeki eşleşen noktaların ses üzerinde hangi kısma karşılık geldiğini tespit etmek için, eşleşen noktalar ses üzerine iz düşürülmüştür. Bunun için öncelikle kümelerdeki anahtar noktaların x koordinatlarının ortalaması alınmaktadır. Daha sonra iki küme için hesaplanan orta noktalar M1 ve M2' in spektrogram genişliğine bölünüp, sesteki eleman sayısı ile çarpılmasıyla bu noktalar sese iz düşürülmektedir. Son olarak ise bu noktaların seste hangi segmentte yer aldığı araştırılarak kopyala-yapıştır yapılmış segmentler işaretlenmektedir.

DENEYSEL SONUÇLAR

Bu bölümde, önerilen yöntemin performans sonuçları hakkında ayrıntılı bilgiler verilmekte ve son teknoloji yöntemler sunulmaktadır. Öncelikle kullanılan veri seti ve değerlendirme metrikleri verilmiş, ardından karşılaştırmalı performans değerlendirme raporları sunulmuştur.

Kullanılan Veri tabanı Ve Metrikler

Bu çalışmada, TIMIT konuşma veri tabanından oluşturulan ses kopyala-yapıştır veri tabanı kullanılmıştır (Ustubioglu vd., 2022). TIMIT veri tabanı, iki saniyeden altı saniyeye kadar İngilizce konuşmaları içermektedir. Ses kayıtlarının örnekleme hızı 8 kHz'dir. Bu veri tabanında yer alan ses kayıtları sessiz ortamda kaydedilmiş olup gürültü içeriğine sahip değildir. Oluşturulan kopyala-yapıştır veri tabanında, ses içerisindeki bir segment rastgele kopyalanmakta ardından aynı sesteki rastgele bir başka konuma yapıştırılarak sahte ses dosyası oluşturulmaktadır. Her sahte segment, yaklaşık 0,2 saniye ila 0,6 saniye uzunluğundadır. Veri tabanında bu sahtecilik türü ile oluşturulmuş ek atak içermeyen 368 adet sahte ses dosyası yer almaktadır. Ayrıca sahte seslere yaygın olarak kullanılan bazı son işleme işlemlerini uygulanmıştır bu ataklara ilişkin bilgiler şu şekildedir: 30 dB ve 20 dB beyaz Gauss gürültüsü, medyan filtre, 32 kbps ve 64 kbps sıkıştırma. Ataklı sahte sesler dâhil olmak üzere toplam 2208 adet sahte ses dosyası mevcuttur.

$$\text{Kesinlik (Precision)} = \frac{DP}{DP + YP}$$

$$\text{Duyarlılık, DPO(Recall, TRP)} = \frac{DP}{DP + YN} \quad (2)$$

$$F - \text{ölçütü} = 2x \frac{\text{Kesinlik} \cdot \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}}$$

Yöntemlerin performans değerlendirmesinde Precision, Recall ve F-measure metrikleri kullanılmıştır. Dikkate alınan metrikler Denklem (2)'de verilmiştir. Kullanılan parametrelerin anlamları Tablo 1'de verilmiştir.

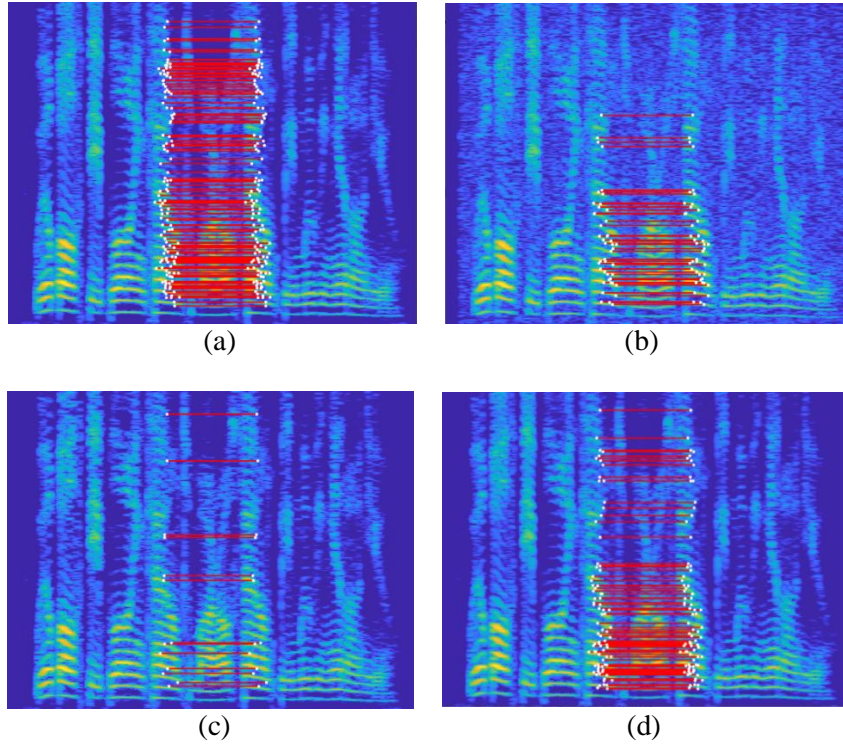
Tablo 1. Değerlendirme Metriklerindeki Kısaltmaların Anlamı.

Kısaltma	Anlamı
DP	Gerçekte sahte olup sahte olarak etiketlenen ses sayısı
YP	Orjinal olup sahte olarak etiketlenen ses sayısı
YN	Sahte olup orjinal olarak etiketlenen ses sayısı

Keskinlik metriği, bütün sesler arasından sahte olarak tespit edilen seslerin oranını temsil ederken, Duyarlılık metriği sahte sesler arasından sahte olarak tespit edilen seslerin oranını temsil etmektedir. F-ölçütü metriği ise Keskinlik ve Duyarlılık metriklerinin harmonik ortalamasıdır hem yanlış pozitifleri hem de yanlış negatifleri tek bir değerle hesaba katar. Bu üç metrik de 0 ile 1 arasında değerlerdir. Bu metriklerin 1'e yakın olması daha yüksek performansı temsil etmektedir.

Önerilen Yöntemin Performansının Literatürdeki Çalışmalarla Karşılaştırılması

Bu bölümde önerilen yöntemin literatürdeki popüler çalışmalardan olan (Ustubioglu vd. 2023), (Imran vd., 2017), (Huang vd., 2020), (Yang vd. 2019) ve (Wang vd., 2010) ile karşılaştırması yapılmıştır. Bu amaçla yapılan tüm ölçümler, Matlab R2019b çalıştıran, 3,4 GHz Intel Core i7 CPU ve 8 GB RAM'e sahip bir masaüstü bilgisayarda gerçekleştirilmiştir. Önerilen yöntemin performansı hem ataksız sahte sesler hem de medyan filtreleme, 20 dB ve 30 dB gürültü ekleme, 32 kbps ve 64 kbps sıkıştırma saldırıları durumunda değerlendirilmiştir. Bu değerlendirmede her bir saldırı tipi için 368 adet ses dosyası kullanılmıştır. İlk olarak önerilen yönteme ait spektrogram görüntüleri üzerindeki anahtar nokta eşleşmelerine ilişkin sonuçlar Şekil 6'da sunulmuştur.



Şekil 6. Önerilen Yöntem ile Elde Edilen Eşleşmelere İlişkin Örnek Görsel Sonuçlar **a.** Ataksız **b.** Gürültü Ekleme **c.** Sıkıştırma **d.** Medyan Filtreleme Atığı Uygulanmış Seslerde.

Bunun için veri tabanında yer alan “si1798_5-3” isimli sahte ses dosyasına ait ataklı ve ataksız sonuçlar elde edilmiştir. Şekil 6(a)’da ataksız sestten elde edilen eşleşmelerin yer aldığı görsel verilmiştir. Ataksız sestten elde edilen spektrogramda çözünürlük daha yüksek olduğu için diğer durumlara göre daha yüksek sayıda eşleşme elde edilmiştir. (b)’de 20 dB gürültü ekleme atağı uygulanan sestten elde edilen eşleşmeler yer almaktadır. Atak olmayan duruma göre eşleşme sayısının azaldığı görülse de sahte bölgelerin belirlenmesine yetecek sayıda eşleşme elde edilebilmiştir. (c)’de ise 32 bit sıkıştırma atağı olması durumundaki eşleşmeler verilmiştir. Bu senaryo ile en az sayıda eşleşmenin elde edilmesine rağmen yine de yöntemin sahte bölgeleri doğru bir şekilde işaretlediği görülmektedir. (d)’de yer alan örnek görsel sonuç ise Median filtreleme atağı durumunda elde edilmiştir. Bu durumda da yöntem sesin sahte olduğunu işaretleyebilecek olmasının yanı sıra sahte ses bölümlerini rahatlıkla belirleyebilme performansı göstermiştir.

Önerilen yöntemle ilk karşılaştırma tarafımızca önerilen bir önceki çalışmamız ile yapılmıştır (Ustubioglu vd. 2023). Bu iki yöntem ile elde edilen ortalama Kesinlik, Duyarlılık ve F-ölçütü sonuçları Tablo 2’de verilmiştir.

Tablo 2. Bir Önceki Çalışmamız ile Performans Karşılaştırması.

Atak türü	Önerilen Yöntem			(Ustubioglu vd. 2023)		
	Kesinlik	Duyarlılık	F ölçütü	Kesinlik	Duyarlılık	F ölçütü
<i>Ataksız</i>	0,99	0,90	0,95	0,89	0,82	0,84
<i>Median filtreleme</i>	0,99	0,87	0,93	0,83	0,80	0,91
<i>20db gürültü</i>	0,99	0,78	0,87	0,82	0,74	0,84
<i>30db gürültü</i>	0,99	0,85	0,91	0,83	0,78	0,86
<i>32 bit sıkıştırma</i>	0,99	0,8	0,89	0,83	0,73	0,86
<i>64 bit sıkıştırma</i>	0,99	0,86	0,92	0,89	0,75	0,89

Her metrik açısından elde edilen en yüksek değerler kalın olarak gösterilmiştir. Tablo 2’den de görüldüğü üzere önerilen yöntem, ataksız durumda en iyi durum senaryosu için daha yüksek performans sonuçlarına sahiptir. Ataksız seslerde önceki çalışmamızda ortalama %89, %82 ve %84 sırasıyla Kesinlik, Duyarlılık ve F-ölçütü elde edilirken, önerilen çalışmada ise %99 Kesinlik, %90 Duyarlılık ve %95 F-ölçütü metrik sonuçları elde edilmiştir. Yöntemin sahte sese atak uygulanması durumunda sonuçlarına bakıldığında ise kesinlik, duyarlılık ve F-ölçütleri açısından tüm ataklarda daha yüksek performans elde edildiği görülmektedir. Önerilen yöntemden elde edilen performans sonuçlarının tarafımızdan önerilen bir önceki çalışmamıza göre daha yüksek olmasının sebebi bu çalışmada önerilen sahtecilik lokalizasyon algoritmasıdır. Bir önceki çalışmamızda eşleşen anahtar noktalardan x eksenine göre min ve max değere sahip anahtar noktalar alınarak sese izdüşüm gerçekleştiriliyordu. Ancak sese atak uygulanma durumunda anahtar nokta sayısı azalmaktadır. Yeterli anahtar noktası çıkarılamaması sonucunda da min ve max anahtar noktalar ile sahte segmentlerin kaçırılabilme olasılığı oluşmaktadır. Fakat önerilen sahtecilik lokalizasyon algoritmasıyla yukarı da anlatıldığı gibi eşleşen anahtar noktala öbeklerinin orta noktaları alındığı için yeterli sayıda anahtar nokta çıkmasa dahi sahte segmentler işaretlenebilmektedir.

Tablo 3. Yöntemlerin Ortalama Duyarlılık, DPO Değerleri.

Atak türü	Önerilen Yöntem	(Ustubioglu vd., 2023)	(Imran vd., 2017)	(Huang vd., 2020)	(Yan vd., 2019)	(Wang vd., 2010)
<i>Ataksız</i>	0,90	0,82	0,18	0,19	0,23	0,22
<i>Median filtreleme</i>	0,87	0,80	0,10	0,21	0,28	0,22
<i>20db ile gürültü ekleme</i>	0,78	0,74	0,17	0,1	0,37	0,25
<i>30db ile gürültü ekleme</i>	0,85	0,78	0,19	0,16	0,31	0,29
<i>32 bit sıkıştırma</i>	0,80	0,73	0,20	0,20	0,31	0,29
<i>64 bit sıkıştırma</i>	0,86	0,75	0,20	0,19	0,30	0,30

Tablo 3’te önerilen yöntemin ve literatürdeki popüler çalışmaların ortalama Duyarlılık, DPO sonuçlarına yer verilmiştir. Önerilen yöntem ile atak olmaması durumunda %90 oranında başarımla yakalanırken ataklı sahte seslerin tespit edilmesinde bu değer %78-87 arasındadır. Referans çalışmalardan, (Ustubioglu vd. 2023) ile nispeten daha yakın performans sonuçları elde edilse de diğer yöntemlerin bu veri seti üzerindeki performans düşüklüğü göze çarpmaktadır. Bunun sebebi ise diğer yöntemlerin önerdikleri algoritmaların genelleme yeteneklerinin olmaması olarak görülmektedir. Çünkü kendi veri setlerinde rapor edilen değerler oldukça yüksektir. Farklı bir veri setinden ses ile test edildiğinde bu algoritmaların performansı görüldüğü gibi oldukça düşmektedir. Elde edilen sonuçlar

gözden geçirildiğinde önerilen yöntemin ses kopyala-yapıştır sahteciliği tespitinde literatürdeki diğer çalışmalara göre üstün bir performans gösterdiği görülmektedir. Yine önerilen yöntemin saldırganlar tarafından sahtecilik izlerini gizlemek adına yapılan son işlem operasyonları da oldukça dayanıklı olduğu görülmektedir.

SONUÇLAR

Çalışmada, ses kopyala-yapıştır sahteciliği tespiti için yeni bir yöntem önerilmiştir. Giriş sesi, yüksek frekans özelliklerine sahip süper çözünürlüklü bir spektrogramla temsil edilmiştir. Spektrogram görüntüsündeki tekrarlanan bölgeleri tespit etmek için, tüm spektrogram görüntüsünden AKAZE anahtar noktaları çıkarılmıştır. AKAZE anahtar noktalarına ait özellik tanımlayıcıların g2NN yaklaşımı ile eşleştirilmesi sonrası sesteki tekrarlanan segmentler görüntü üzerinde bulunmuştur. Eşleşen anahtar noktaların konumlarının ortalamalarından faydalanılarak ses dosyasındaki sahte segmentlerin tespiti gerçekleştirilmiştir. Deneysel analizler, halka açık veri seti üzerinde gerçekleştirilmiştir. Elde edilen sonuçlar, önerilen yöntemin sahte sese uygulanan atak senaryolarında dahi en başarılı tespit yaklaşımı olduğunu göstermektedir.

Özellikle sahte seslerdeki sahtecilik izlerini gizlemek adına uygulanan ataklar göz önüne alındığında, ses kopyala-yapıştır sahteciliği daha yüksek doğrulukla tespit etmeye yönelik iyileştirmeler, gelecekteki çalışmalarımızın temelini oluşturmaktadır. Bu hedefe ulaşmak için atak bağımsız özellikler incelenecektir. Aynı zamanda probleme özel derin öğrenme tabanlı ağ yaklaşımları da değerlendirilecektir. Bunların sonucunda ses dosyalarına uygulanabilecek diğer yeni saldırı türleri de değerlendirilerek, geliştirilecek yöntemin bu saldırı türlerine karşı dayanıklı olunması sağlanacaktır.

KAYNAKLAR

Alcantarilla, P. F., Bartoli, A., & Davison, A. J. (2012). KAZE features. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VI 12* (pp. 214-227). Springer Berlin Heidelberg.

Alcantarilla, P. F., Nuevo, J., & Bartoli, A. (2013). Fast explicit diffusion for accelerated features in nonlinear scale spaces british machine vision conference (BMVC).

Amerini, I., Ballan, L., Caldelli, R., Del Bimbo, A., & Serra, G. (2011). A sift-based forensic method for copy-move attack detection and transformation recovery. *IEEE transactions on information forensics and security*, 6(3), 1099-1110 DOI:10.1109/TIFS.2011.2129512

BURUCU, E. (2023). Adli Bilimlerde Ses Kayıtları Üzerinde Manipülasyon İncelemesi. *Hacettepe Üniversitesi Edebiyat Fakültesi Dergisi*, 40(1) DOI:10.32600/huefd.1106795

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381-395 DOI:10.1145/358669.358692

Huang, X., Liu, Z., Lu, W., Liu, H., & Xiang, S. (2020). Fast and effective copy-move detection of digital audio based on auto segment. In *Digital forensics and forensic Investigations: Breakthroughs in Research and Practice* (pp. 127-142). IGI Global DOI:10.4018/IJDCF.2019040104

Imran, M., Ali, Z., Bakhsh, S. T., & Akram, S. (2017). Blind detection of copy-move forgery in digital audio forensics. *IEEE Access*, 5, 12843-12855 DOI:10.1109/ACCESS.2017.2717842

Nam, J., Mysore, G. J., Ganseman, J., Lee, K., & Abel, J. S. (2010). A super-resolution spectrogram using coupled PLCA. In *Eleventh Annual Conference of the International Speech Communication Association*.

Ustubioglu, B., Küçükuğurlu, B., & Ulutas, G. (2022). Robust copy-move detection in digital audio forensics based on pitch and modified discrete cosine transform. *Multimedia Tools and Applications*, 81(19), 27149-27185 DOI:10.1007/s11042-022-13035-3.

Ustubioglu, B., Tahaoglu, G., & Ulutas, G. (2023). Detection of audio copy-move-forgery with novel feature matching on Mel spectrogram. *Expert Systems with Applications*, 213, 118963 DOI:10.1016/j.eswa.2022.118963.

Wang, F., Li, C., & Tian, L. (2017, October). An algorithm of detecting audio copy-move forgery based on DCT and SVD. In *2017 IEEE 17th International Conference on Communication Technology (ICCT)* (pp. 1652-1657). IEEE.

Xiao, J. N., Jia, Y. Z., Fu, E. D., Huang, Z., Li, Y., & Shi, S. P. (2014). Audio authenticity: Duplicated audio segment detection in waveform audio file. *Journal of Shanghai Jiaotong University (Science)*, 19, 392-397 DOI:10.1007/s12204-014-1515-5.

Xie, Z., Lu, W., Liu, X., Xue, Y., & Yeung, Y. (2018). Copy-move detection of digital audio based on multi-feature decision. *Journal of information security and applications*, 43, 37-46 DOI:10.1016/j.jisa.2018.10.003

Yan, Q., Yang, R., & Huang, J. (2015, April). Copy-move detection of audio recording with pitch similarity. In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1782-1786). IEEE.

Yan, Q., Yang, R., & Huang, J. (2019). Robust copy-move detection of speech recording using similarities of pitch and formant. *IEEE Transactions on Information Forensics and Security*, 14(9), 2331-2341 DOI:10.1109/TIFS.2019.2895965