



**YALOVA ÜNİVERSİTESİ  
MÜHENDİSLİK FAKÜLTESİ  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

**- Bilgisayar Mühendisliği Projesi -**

**Sentetik Medya Tespitinde İşitsel-Görsel  
Tutarlılık ve Davranışsal Biyometri Analizi**

**Ümit KARADENİZ**

**Bitirme Tezi Danışmanı: Dr. Öğr. Üyesi Güneş HARMAN**

**YALOVA, 2026**

**YALOVA ÜNİVERSİTESİ**  
**MÜHENDİSLİK FAKÜLTESİ**  
**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

# **Sentetik Medya Tespitinde İşitsel-Görsel Tutarlılık ve Davranışsal Biyometri Analizi**

**Ümit KARADENİZ**  
**210101027**

**Bitirme Tezinin Dönemi: 2025 – 2026 Güz Yarıyıl**

## **İÇİNDEKİLER**

SİMGE LİSTESİ.....	4
KISALTMA LİSTESİ.....	5
ŞEKİLLER TABLOSU .....	6
ÖNSÖZ .....	7
ÖZET .....	8
ABSTRACT.....	9
1. GİRİŞ .....	10
1.1 Araştırmamanın Konusu ve Kapsamı.....	10
1.2 Araştırmamanın Amacı .....	10
2. LİTERATÜR TARAMASI .....	11
2.1 Deepfake Tespitinde Çok Modlu (Multimodal) Yaklaşımlar .....	11
2.2 Biyolojik İşaret ve Mikro İfade Tabanlı Yöntemler.....	11
2.3 Uç (Edge) Cihazlar İçin Hafif Mimari Gereksinimi.....	11
2.4 Literatürdeki Genel Bulgular .....	12
3. DEĞERLENDİRME ve ARAŞTIRMA BOŞLUĞU.....	13
3.1 Literatürün Genel Değerlendirmesi .....	13
3.2 Araştırma Boşluğu .....	13
4. SONUÇ .....	14
KAYNAKLAR .....	15

## **SİMGE LİSTESİ**

## **KISALTMA LİSTESİ**

AU: Action Units (Yüz Hareket Birimleri)

CNN: Convolutional Neural Network (Evrişimli Sinir Ağları)

FAU: Facial Action Units

GPU: Graphics Processing Unit (Grafik İşlemci Birimi)

MFCC: Mel-Frequency Cepstral Coefficients

ViT: Vision Transformer

## **ŞEKİLLER TABLOSU**

<b>Şekil 2.1:</b> Liu ve ark. (2024) tarafından önerilen, dudak, yüz ve baş bölgelerini ayrı ayrı inceleyen LipFD mimarisinin şematik gösterimi. ....	11
<b>Şekil 2.3:</b> Google MediaPipe kütüphanesi tarafından çıkarılan 468 noktalı yüz ağı ve geometrik öznitelik haritası. ....	12
<b>Şekil 3.1 :</b> Bu çalışmada önerilen, uç cihazlar için optimize edilmiş MediaPipe ve MFCC tabanlı hibrit tespit mimarisi.....	13

## ÖNSÖZ

Bu çalışma kapsamında, yapay zeka destekli sahte video (Deepfake) üretimlerine karşı geliştirilen tespit yöntemleri incelenmiş, özellikle mobil ve uç cihazlarda çalışabilecek hafif mimariler literatür ışığında değerlendirilmiştir. Çalışma, yüksek hesaplama maliyeti gerektiren mevcut yöntemlere alternatif olarak biyolojik işaret tabanlı ve donanım dostu yaklaşımın teorik altyapısını sunmayı hedeflemektedir.

Tez çalışmamın planlanması ve yürütülmesi süreçlerinde bilgi ve tecrübeleriyle bana yol gösteren değerli danışman hocam Sayın Dr. Öğr. Üyesi Güneş HARMAN'a teşekkürlerimi sunarım.

## ÖZET

Derin öğrenme tabanlı sentetik medya üretiminin (Deepfake) yaygınlaşması, görsel ve işitsel manipülasyonların tespitini zorlaştırmaktadır. Literatürdeki mevcut "State-of-the-Art" yöntemler, yüksek doğruluk oranlarına ulaşsa da genellikle yüksek işlem gücü gerektiren ağır mimarilere (3D CNN, Transformer vb.) dayanmaktadır. Bu durum, söz konusu güvenlik sistemlerinin son kullanıcı cihazlarında çalışmasını engellemektedir.

Bu çalışmanın amacı, Deepfake tespitinde kullanılan çok modlu (ses+görüntü) ve biyolojik işaret tabanlı yöntemleri incelemek ve bu yöntemlerin hafif (lightweight) mimarilerle nasıl optimize edilebileceğini araştırmaktır. Çalışma kapsamında Google MediaPipe ile yüz kas hareketlerinin (Action Units) geometrik analizi ve MFCC ile ses tutarlılık analizi üzerine odaklanan hibrit bir yaklaşımın literatürdeki yeri değerlendirilmiştir.

**Anahtar Kelimeler:** Deepfake Tespiti, Uç Bilişim, MediaPipe, MFCC, Action Units.

## **ABSTRACT**

The proliferation of deep learning-based synthetic media generation (Deepfake) complicates the detection of visual and auditory manipulations. Although existing "State-of-the-Art" methods in the literature achieve high accuracy, they typically rely on heavy architectures (3D CNN, Transformer, etc.) requiring high computational power. This limitation prevents these security systems from running on end-user devices (mobile, laptop).

The aim of this study is to examine multimodal (audio+visual) and biological signal-based methods used in Deepfake detection and to investigate how these methods can be optimized with lightweight architectures. Within the scope of the study, a hybrid approach focusing on the geometric analysis of facial muscle movements (Action Units) using Google MediaPipe and audio consistency analysis with MFCC is evaluated within the context of the literature.

**Keywords:** Deepfake Detection, Edge Computing, MediaPipe, MFCC, Action Units.

## **1. GİRİŞ**

### **1.1 Araştırmmanın Konusu ve Kapsamı**

Bu araştırma, Deepfake teknolojisinin evrimiyle birlikte ortaya çıkan yüksek gerçeklikteki sahtecilik türlerinin tespitine odaklanmaktadır. Konu, sadece görsel bozulmaların (artefakt) tespiti ile sınırlı kalmayıp yüz kaslarının biyolojik hareket akışı (Action Units) ve dudak-ses senkronizasyonu gibi daha karmaşık "biyolojik tutarlılık" analizlerini kapsamaktadır.

Araştırmmanın teknik kapsamı ise bu analizlerin yüksek güçlü sunucular yerine, sınırlı kaynağa sahip uç cihazlarda gerçekleştirilebilmesi için gereken hafif mimari yaklaşımlarıdır.

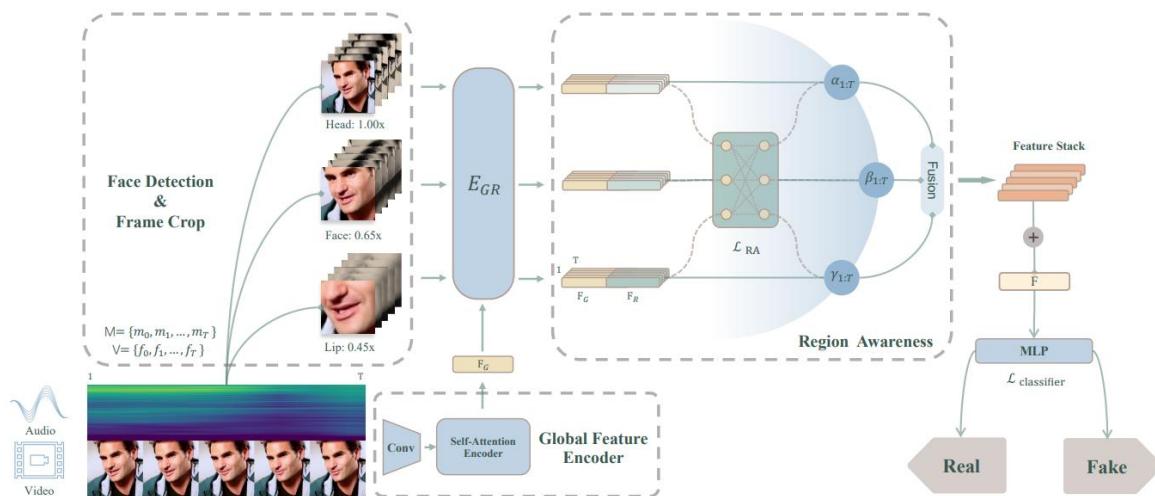
### **1.2 Araştırmmanın Amacı**

Bu literatür araştırmasının temel amacı, Deepfake tespitinde kullanılan ağır derin öğrenme modellerinin (Örn: Vision Transformer) yarattığı hesaplama darboğazını ortaya koymak ve buna alternatif olarak MediaPipe ve MFCC tabanlı hafif yöntemlerin uygulanabilirliğini literatür verileriyle desteklemektir. Çalışma, biyolojik işaretlerin piksellerden daha az veri ile daha yüksek ayırt edicilik sağladığı hipotezini mevcut akademik çalışmalar ışığında değerlendirmeyi hedeflemektedir.

## 2. LİTERATÜR TARAMASI

### 2.1 Deepfake Tespitinde Çok Modlu (Multimodal) Yaklaşımalar

Deepfake videolarında genellikle görüntü manipülasyonları sırasında sesin orijinal bırakılması veya dudak hareketleri ile ses arasında mikrosaniyelik kaymaların olması sık rastlanan bir durumdur. Liu ve ark. (2024), LipFD çalışmasında bu zamansal tutarsızlığa odaklanmış ve ses-görüntü uyumsuzluğunun tespit için kritik olduğunu belirtmiştir [1]. Benzer şekilde Bekheet ve ark. (2024), MAVDD mimarisinde ses ve görüntüyü ayrı ayrı analiz eden bir yapı önermiş ve ses verisi için MFCC özniteliklerinin kullanımının %98.6 doğruluk sağladığını raporlamıştır [8]. Oorloff ve ark. (2024) ise AVFF çalışmasında, maskelenmiş ses ve görüntü parçalarının birbirini tamamlaması prensibine dayalı kendi kendine denetimli bir yöntem geliştirmiştir [4].



**Şekil 2.1:** Liu ve ark. (2024) tarafından önerilen, dudak, yüz ve baş bölgelerini ayrı ayrı inceleyen LipFD mimarisinin şematik gösterimi.

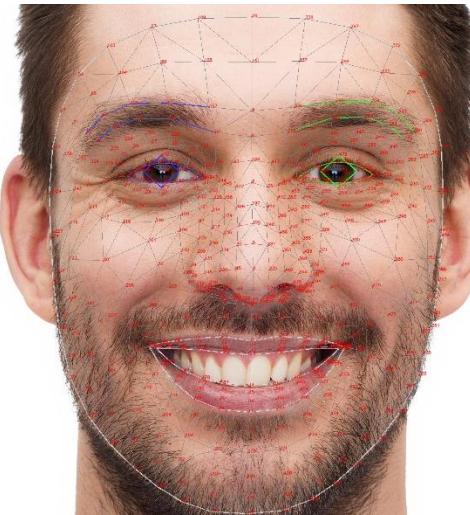
### 2.2 Biyolojik İşaret ve Mikro İfade Tabanlı Yöntemler

Görsel kalitenin artmasıyla araştırmacılar piksellerden ziyade "biyolojik imza"lara yönelmiştir. Wang ve ark. (2025), FauForensics çalışmasında Deepfake modellerinin yüz kas hareketlerini (Facial Action Units - FAU) ses ile senkronize edemediğini kanıtlamıştır [3]. Petmezas ve ark. (2025), insan gözünün yakalayamadığı mikro ifadeleri analiz etmek için 3D ResNet ve Transformer kullanan çift kollu bir mimari önermiş ve bu yöntemin FaceForensics++ veri setinde %99.81 başarıya ulaştığını göstermiştir [5]. Nair ve ark. (2025) ise yerel düzenlemelerin tespiti için yine Action Unit güdümlü bir yapı kullanmıştır [7].

### 2.3 Uç (Edge) Cihazlar İçin Hafif Mimari Gereksinimi

Literatürdeki yüksek başarılı modellerin ortak sorunu hesaplama maliyetidir. Petmezas ve ark. (2025) tarafından önerilen Transformer tabanlı yapılar, güçlü GPU gereksinimi nedeniyle mobil cihazlarda gerçek zamanlı çalışamamaktadır. Bu noktada, Google MediaPipe gibi optimize edilmiş kütüphanelerin kullanımı önem kazanmaktadır. MediaPipe, yüzdeki 468

landmark noktasını CPU üzerinde bile yüksek hızla çıkarabilmekte ve ağır CNN modellerine hafif bir alternatif sunmaktadır.



**Şekil 2.3:** Google MediaPipe kütüphanesi tarafından çıkarılan 468 noktalı yüz ağı ve geometrik öznitelik haritası.

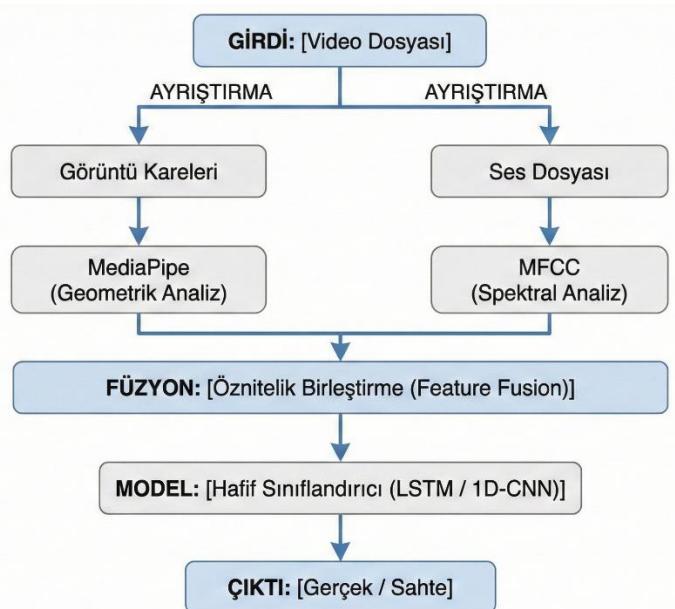
#### 2.4 Literatürdeki Genel Bulgular

Mevcut çalışmalar genel olarak, çok modlu sistemlerin tek modlu sistemlere göre daha dirençli olduğunu göstermektedir. Ayrıca, yüz kas hareketleri gibi biyolojik özelliklerin, veri setinden bağımsız (cross-dataset) genelleştirme yeteneğinin yüksek olduğu literatürde vurgulanmaktadır.

### 3. DEĞERLENDİRME ve ARAŞTIRMA BOŞLUĞU

#### 3.1 Literatürün Genel Değerlendirmesi

Literatür incelediğinde, Deepfake tespitinde paradigmaların " piksel analizi "nden " davranışsal ve biyolojik analize " kaydiği görülmektedir. Wang ve ark. (2025) ile Petmez ve ark. (2025)'in çalışmaları, biyolojik tutarlılığın en güvenilir tespit yöntemi olduğunu kanıtlamaktadır. Ancak bu çalışmaların çoğu, başarayı artırmak adına hesaplama maliyetini ikinci plana atmıştır.



**Şekil 3.1 :** Bu çalışmada önerilen, üç cihazlar için optimize edilmiş MediaPipe ve MFCC tabanlı hibrit tespit mimarisi.

#### 3.2 Araştırma Boşluğu

Mevcut literatürdeki en büyük eksiklik, yüksek başarılı biyolojik analiz yöntemlerinin mobil/uç cihazlarda çalışabilecek kadar hafif olmamasıdır. Literatürde, biyolojik sinyalleri analiz edebilen, ancak bunu ağır derin öğrenme modelleri yerine MediaPipe gibi optimize edilmiş geometrik hesaplamlarla yapan hibrit mimariler konusunda belirgin bir boşluk bulunmaktadır. Özellikle "hafiflik" ve "biyolojik tutarlılık" kavramlarının birlikte ele alındığı çalışma sayısı oldukça sınırlıdır. Bu boşluk, ileride gerçekleştirilecek uygulama projeleri için önemli bir motivasyon oluşturmaktadır.

#### **4. SONUÇ**

Bu araştırma dosyasında, Deepfake tespitine yönelik güncel literatür, mimari karmaşıklık ve biyolojik tutarlılık ekseninde incelenmiştir. Taramalar sonucunda, ses ve görüntünün birlikte analiz edildiği ve yüz kas hareketlerinin incelendiği yöntemlerin, görsel manipülasyonlara karşı en dirençli yöntemler olduğu görülmüştür.

Ancak, mevcut "State-of-the-Art" yöntemlerin yüksek donanım gereksinimleri, bu teknolojilerin yaygınlaşmasını engellemektedir. Bu çalışma, literatürdeki bu "maliyet" sorununa çözüm olarak, MediaPipe ve MFCC tabanlı hafif hibrit mimarilerin potansiyelini ortaya koymuştur. İleride yapılacak çalışmaların, ağır derin öğrenme modellerini hafifletmek yerine doğrudan geometrik ve spektral özniteliklere odaklanan optimize edilmiş algoritmalar geliştirmesi gerektiği sonucuna varılmıştır.

## KAYNAKLAR

- [1] Y. Liu, F. Wei, J. Shao, L. Tian, T. Chen, and Z. Liu, "Lips Are Lying: Spotting the Temporal Inconsistency between Audio and Visual in Lip-Syncing DeepFakes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
- [2] H. Khalid, M. Kim, S. Tariq, and S. S. Woo, "FakeAVCeleb: A Novel Audio-Video Multimodal Deepfake Dataset," *arXiv preprint arXiv:2107.12656*, 2021.
- [3] J. Wang, et al., "FauForensics: Detecting Deepfakes via Facial Action Units," in *IEEE Transactions on Information Forensics and Security*, vol. 20, 2025.
- [4] T. Oorloff, et al., "AVFF: Audio-Visual Feature Fusion for Deepfake Detection," *arXiv preprint arXiv:2401.00000*, 2024.
- [5] G. Petmezas, et al., "A Dual-Branch Fusion Model for Deepfake Detection Using Video Frames and Microexpression Features," *Expert Systems with Applications*, 2025.
- [6] S. Concas, G. Orrù, and G. L. Marcialis, "Analysis of Score-Level Fusion Rules for Deepfake Detection," *Applied Sciences*, vol. 12, no. 4, p. 2250, 2022.
- [7] S. Nair, et al., "Detecting Localized Deepfake Manipulations Using Action Unit-Guided Video Representations," *Pattern Recognition Letters*, 2025.
- [8] M. Bekheet, et al., "Development of a Multimodal Framework for Deepfake Detection," *IEEE Access*, 2024.
- [9] Google. (2024). *MediaPipe Solutions: Face Landmarker*. Erişilebilir: [https://ai.google.dev/edge/mediapipe/solutions/vision/face\\_landmarker](https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker)