DOI: https://doi.org/10.30546/19023.978-9952-8605-7-3.2025.0125.
ISBN: 978-9952-8605-7-3

ADLİ BİLİŞİMDE YAPAY ZEKA DESTEKLİ GÖRÜNTÜ SINIFLANDIRMA: HİBRİT TEMEL MODEL YAKLAŞIMI

Caner Kocamaz[1], Nursel Yalçın[2]

[1] Gazi Üniversitesi, Bilişim Enstitüsü, Bilişim Sistemleri ABD, Ankara, Türkiye
[2] Gazi Üniversitesi, Gazi Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, Ankara, Türkiye
ORCID: [1]0000-0003-4860-6178; [2]0000-0002-0393-6408

ÖZET

Adli bilişim soruşturmalarında, uzmanların el koyulan dijital cihazlardan çıkan binlerce görseli manuel olarak incelemesi, büyük bir zaman kaybına ve insan kaynaklı hatalara yol açmaktadır. Mevcut otomasyon çözümleri, özellikle vakaya özgü ve nadir kanıt türlerini tespitte yetersiz kalabilmektedir. Bu çalışma, adli görsel inceleme sürecini hızlandırmak ve doğruluğu artırmak amacıyla, iki farklı temel modeli (foundation model) birleştiren hibrit bir yaklaşım sunmaktadır. Önerilen mimaride, DINOv3'un kendinden denetimli öğrenme ile elde ettiği güçlü görsel temsilleri ile CLIP'in görsel-metin çiftleri üzerine eğitilerek kazandığı semantik anlama yeteneği, bir dikkat mekanizması tabanlı geç füzyon katmanında birleştirilmektedir. Bu hibrit sınıflandırıcının çıktıları, yanlış pozitifleri azaltmak ve kritik kanıtları önceliklendirmek amacıyla, bir güven skoru ve risk puanlama mekanizmasından geçirilmektedir. Çalışmada, silah, uyuşturucu ve para gibi temel adli sınıflar üzerinde, geleneksel ResNet-50 ve EfficientNet-B4 modelleri ile karşılaştırmalı deneyler yürütülmüştür. Sonuçlar, önerilen hibrit modelin, özellikle kritik sınıflarda daha yüksek AUC değerleri ve daha dengeli bir kesinlik-duyarlılık (Precision-Recall) performansı sergilediğini göstermiştir. Ayrıca, az-örnekli (few-shot) tanıma için önerilen statik prototip yöntemi, belge ve kimlik kartı gibi yeni sınıflara hızlıca uyum sağlama potansiyeli göstermiştir. Bu çalışma, temel modellerin adli bilişimdeki potansiyelini ortaya koymakta ve uzmanların iş yükünü hafifletebilecek entegre bir çözüm önermektedir.

Anahtar Kelimeler: Adli Bilişim, Yapay Zeka, Görüntü Sınıflandırma, Hibrit Modeller, DINOv3, CLIP, Geç Füzyon

ABSTRACT

AI-ASSISTED IMAGE CLASSIFICATION IN DIGITAL FORENSICS: A HYBRID FOUNDATION MODEL APPROACH

In digital forensic investigations, experts manually review thousands of images seized from digital devices, leading to significant time consumption and human error. Existing automation solutions often fall short in detecting case-specific and rare types of evidence. This study proposes a hybrid approach that combines two different foundation models to accelerate and improve the accuracy of forensic image

analysis. The proposed architecture integrates the powerful visual representations of DINOv3, obtained through self-supervised learning, with the semantic understanding capability of CLIP, trained on visual-text pairs, via an attention-based late fusion layer. The outputs of this hybrid classifier are filtered through a confidence score and risk scoring mechanism to reduce false positives and prioritize critical evidence. Comparative experiments were conducted on core forensic classes such as weapons, drugs, and money against traditional ResNet-50 and EfficientNet-B4 models. Results demonstrate that the proposed hybrid model achieves higher AUC values and a more balanced Precision-Recall performance, especially for critical classes. Furthermore, the proposed static prototype method for few-shot recognition shows potential for rapid adaptation to new classes like documents and identity cards. This work highlights the potential of foundation models in digital forensics and proposes an integrated solution to alleviate the workload of experts.

Keywords: Digital Forensics, Artificial Intelligence, Image Classification, Hybrid Models, DINOv3, CLIP, Late Fusion

GİRİŞ

Dijital teknolojilerin yaygınlaşması, adli bilişim soruşturmalarının kapsamını ve karmaşıklığını her geçen gün artırmaktadır. Bir olay yerine el konulan cihazlardan çıkarılan binlerce, hatta yüzbinlerce fotoğraf, video karesi ve ekran görüntüsünün manuel olarak incelenmesi, uzmanlar için en zahmetli ve zaman alıcı süreçlerden biridir. Bu manuel inceleme sadece verimliliği düşürmekle kalmaz, aynı zamanda dikkat yorgunluğuna bağlı olarak kritik kanıtların gözden kaçırılma riskini de beraberinde getirir.

Son dönemde ortaya çıkan büyük ölçekli temel modeller (foundation models), bu zorlukların aşılmasında umut vaat etmektedir. DINOv3 gibi kendinden denetimli (self-supervised) modeller, etiket kullanmadan öğrendikleri zengin ve genelleştirilebilir görsel temsillerle dikkat çeker. CLIP (Contrastive Language-Image Pre-training) ise, görseller ve onları tanımlayan metin etiketleri arasındaki ilişkiyi öğrenerek, görsel içeriği semantik olarak anlama ve metin yoluyla esnek sorgulama yeteneği kazanır.

Bu çalışmanın temel amacı, DINOv3 ve CLIP modellerinin güçlü yönlerini bir araya getiren, adli görsel sınıflandırmaya yönelik hibrit bir mimari tasarlamak ve değerlendirmektir. Önerilen sistem, bu iki modelden katılaştırılmış (frozen) tabanlardan çıkarılan özellikleri, bir dikkat mekanizması ile ağırlıklandırarak birleştirir (geç füzyon). Sistem ayrıca, sınıflandırma çıktılarını bir güven skoru ve risk puanı ile filtreleyerek, uzmanların incelemesi gereken görselleri önceliklendirir. Bu çalışma ile aşağıdaki katkıların sağlanması hedeflenmektedir:

- Adli bilişim için DINOv3 ve CLIP tabanlı hibrit bir sınıflandırma mimarisinin önerilmesi.
- Geleneksel CNN modelleri kıyaslamasıyla hibrit modelin üstünlüğünün deneysel olarak gösterilmesi.

DOI: https://doi.org/10.30546/19023.978-9952-8605-7-3.2025.0125.
ISBN: 978-9952-8605-7-3

- Az-örnekli sınıflar için hızlı bir prototip tabanlı tanıma mekanizmasının entegrasyonu.
- Uygulama odaklı bir risk puanlama ve çok katmanlı filtreleme sistemi ile sahadaki iş yükünü azaltacak bir çözüm önerilmesi.

İLGİLİ ÇALIŞMALAR

Adli bilişimde otomasyon ihtiyacı, geçmişte birçok araştırmacının ilgisini çekmiştir. Erken dönem çalışmalar, özellikle çıplaklık ve pornografi tespiti gibi alanlarda, geleneksel makine öğrenimi yöntemlerine (örn. doku temelleri, SVM) odaklanmıştır. Derin öğreninin yükselişiyle birlikte, CNN tabanlı modeller bu alanda baskın hale gelmiştir.

Ancak, genel amaçlı CNN'lerin (VGG, ResNet, EfficientNet) önceden eğitilmiş versiyonlarının adli görseller üzerinde doğrudan kullanımı, domain uçurumu (domain gap) problemi nedeniyle genellikle optimal olmayan sonuçlar vermektedir. Bu modellerin ince ayar (fine-tuning) gerektirmesi ve az görülen sınıflarda zayıf performans göstermesi, pratik uygulamalardaki başlıca kısıtlamalardır.

Dilber ve Çetin (2021), artan dijital veri hacminin adli bilişim süreçlerinde yarattığı zorlukların üstesinden gelmek amacıyla, görüntü verilerinin hızlı ve doğru bir şekilde analiz edilmesini sağlayan bir model önermiştir. Araştırmacılar, VGG16 ağ yapısını temel alan ve görüntü sınıflandırma için özel olarak tasarlanmış ağ katmanlarından oluşan bir derin öğrenme modeli geliştirmişlerdir. Modelin test edilmesi için Kaggle platformu ve diğer kaynaklardan toplanan 300x300 piksel çözünürlüğünde 3000'den fazla görüntüden oluşan bir veri seti kullanılmıştır. FloydHub ortamında Keras ve TensorFlow kütüphaneleri ile yapılan testler sonucunda, geliştirilen modelin %97.8'lik bir doğruluk oranına ulaştığı ve benzer çalışmalara göre ortalama %5'lik bir performans artışı sağladığını gözlemlemişlerdir [1].

Özdal (2023), yapay zekanın görüntü içeriği sınıflandırmasındaki rolünü ve çeşitli sektörlerdeki uygulamalarını sistemli bir yaklaşımla incelemiştir. Literatür taraması ve mantıksal akıl yürütme yöntemlerinin birleştirildiği çalışmada, yapay zekanın görüntü sınıflandırma potansiyelini ele almışlardır. Çalışma, yapay zekanın bu alandaki önemini vurgulamak ve tıp, güvenlik, otomotiv, eğitim ve sanat gibi farklı alanlardaki pratik uygulamalarını açıklamıştır. Araştırma sonucunda, yapay zekanın görüntü içerik sınıflandırmasındaki etkisinin ve gelecekteki potansiyelinin altı çizilerek, ilgili alanlarda daha etkili ve verimli çözümler sunulmasına yönelik bir bakış açısı sunmuşlardır [2].

Li ve Liu (2023), HAI (Human-centric Artificial Intelligence) destekli görüntü ve video sınıflandırıcısı destekli triage için bir veri tabanı oluşturmuştur. Dosya sistemindeki tüm görüntü ve videoları daha ileri kategorilendirme, analiz ve inceleme süreci için sınıflandırmışlardır.. HAI ile araştırmacının büyük hacimli görüntü ve videolar arasında anormal ve yanlış bilgileri tespit edebilen yapay zeka önerileri kullanarak karar vermesine yardımcı olabileceğini vurgulamışlardır[3].

Yonar ve Kadem (2024), adli bilimlerde delil değerlendirme süreçlerinde yeni bir teknoloji olarak yapay zekanın devrim niteliğindeki etkisini ele almıştır. Çalışmada, yapay zekanın adli tıp, toksikoloji,

DOI: https://doi.org/10.30546/19023.978-9952-8605-7-3.2025.0125.
ISBN: 978-9952-8605-7-3

adlibilişim, biyometrik kimliklendirme ve suç mahalli analizi gibi çeşitli alanlardaki uygulamalarını incelemişler ve yapay zeka ve makine öğreniminin geleneksel adli tıp uygulamalarını nasıl dönüştürdüğünü, analizlerin doğruluğunu ve verimliliğini nasıl artırdığını ortaya koymuşlardır. Bununla birlikte, bu teknolojilerin kullanımının veri gizliliği, algoritmik önyargı ve otomatik sistemlerin yorumlanabilirliği gibi önemli etik ve yasal sorunları da beraberinde getirdiğine dikkat çekmişlerdir [4].

He ve ark. (2025), patoloji ve radyoloji entegrasyonunu yapay zeka destekli tıbbi görüntülemede birleştiren yenilikçi bir derin öğrenme tabanlı çerçeve sunmuşlardır. Çalışmada, çok modlu veri kaynakları, dengesiz sınıf dağılımları ve klinik karar vermede yorumlanabilirlik ihtiyacı gibi geleneksel görüntü sınıflandırma yöntemlerinin karşılaştığı zorlukları ele almışlardır. Bu sınırlılıkların üstesinden gelmek için, çok çözünürlüklü özellik çıkarımından yararlanan Uyarlanabilir Çok Çözünürlüklü Görüntüleme Ağı (AMRI-Net) ve alan genelleştirilebilirliğini artıran Açıklanabilir Alan-Uyarlanabilir Öğrenme (EDAL) stratejisi olmak üzere iki yeni bileşen sunmuşlardır. Çok modlu tıbbi görüntüleme veri setleri üzerinde yapılan deneysel sonuçlarda, %94.95'e varan sınıflandırma doğrulukları ve %94.85'e varan F1-Skorları ile çerçevenin üstün performansını göstermiş, böylece yapay zeka odaklı tanılarda şeffaflığı ve klinik güveni artırdığı sonucuna varmışlardır [5].

Kara (2023), ilk olarak, dijital görüntüleri sekiz farklı suç kategorisine ayırmak için altı farklı derin öğrenme modeli kullanan yeni bir yaklaşım sunmuştur. İkinci olarak, suçların sınıflandırılması için yeni bir veri seti oluşturmuş ve kamuoyunun kullanımına açmıştır. Bahis, çocuk istismarı, kredi kartı ve bankacılık, uyuşturucu, korkutma, bıçaklar, pornografik ve silahlar gibi 8 farklı suç kategorisinden toplam 15.065 görüntü örneği içeren bu yeni veri seti kullanmıştır. Çalışmada, biri araştırmacılar tarafından geliştirilen bir Evrişimli Sinir Ağı (CNN) modeli ve transfer öğrenme için kullanılan VGG16, VGG19, Xception Model, InceptionResNetV2 ve NASNetLarge olmak üzere altı farklı modelin performansı test doğruluğu ve zaman metrikleri kullanılarak karşılaştırmıştır. Sonuç olarak, NASNetLarge modeli kullanılarak %89.74'e varan bir tahmin doğruluğu elde etmiştir[6].

Kendinden denetimli öğrenme (Self-Supervised Learning - SSL), etiketli veriye olan bağımlılığı azaltmak için umut verici bir alternatif sunar. DINO [7] ve onun gelişmiş bir versiyonu olan DINOv2/DINOv3 [8], ölçeklenebilir bir SSL yöntemi olup, ViT (Vision Transformer) mimarisi üzerinde etkileyici görsel temsiller öğrenir. Bu modeller, nesne parçalarına ve genel sahne yapısına dair zengin bilgileri yakalayarak, çeşitli görsel görevlerde üstün performans sergilemiştir.

Öte yandan, CLIP [9] modeli, görsel ve metin modalitelerini aynı uzayda hizalamayı öğrenir. Bu, modelin sıfır-atış (zero-shot) sınıflandırma yapmasına ve doğal dil sorgularıyla görsel arama gerçekleştirmesine olanak tanır. Adli bilişimde, bir uzmanın "silahlı bıçak" veya "eczane dışı ortamda haplar" gibi esnek sorgular yapabilmesi, CLIP'in bu alandaki potansiyelini güçlü bir şekilde ortaya koymaktadır.

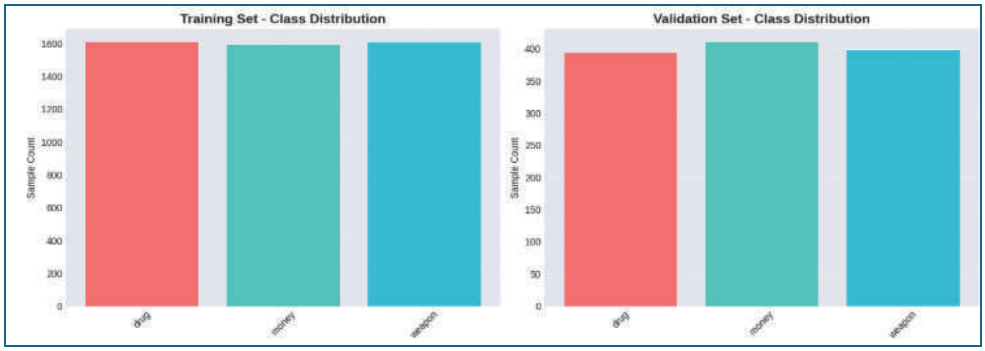
DOI: https://doi.org/10.30546/19023.978-9952-8605-7-3.2025.0125.
ISBN: 978-9952-8605-7-3

Bu çalışma, mevcut literatürdeki boşlukları doldurmak adına, sadece tek bir model türünü kullanmak yerine, DINOv3'ün güçlü görsel temsil yeteneği ile CLIP'in semantik anlama gücünü birleştirerek daha sağlam bir sınıflandırıcı oluşturmayı amaçlamaktadır. Ayrıca, bu hibrit yaklaşımı, risk tabanlı bir önceliklendirme sistemi ile entegre ederek, sahada doğrudan uygulanabilir bir çerçeve sunmaktadır.

MATERYAL VE METOT

Veri Seti

Çalışmada sınıflandırma deneyleri için 2000'er adet drug (uyuşturucu), money (para) ve weapon (silah) kategorilerine ait toplam 6000 adet görüntü dosyası(jpg, png) kullanılmıştır [10]. Az-örnekli tanıma yeteneğini değerlendirmek için, document (belge) ve ids (kimlik kartı) dizinlerinden oluşan ayrı bir destek seti oluşturulmuştur. Modelin genelleme performansını ve CLIP'in metin arama özelliğini demo amacıyla *test-dataset* dizini kullanılmıştır. Görüntüler çeşitli açık erişimli veri setlerinden toplanmış, boyutlandırılmış (224×224 px) ve ImageNet istatistiklerine göre normalize edilmiştir. Veri seti tabakalı örnekleme ile %80 eğitim ve %20 doğrulama şeklinde ayrılmıştır (Şekil-1).



Şekil-1 Veri seti dağılımı

Modeller

Çalışmada üç farklı model mimarisi değerlendirilmiştir:

ResNet-50 ve EfficientNet-B4: Geleneksel CNN tabanlı karşılaştırma modelleri olarak kullanılmıştır. ImageNet üzerinde önceden eğitilmiş ResNet-50 ve EfficientNet-B4 bu modellerin evrişimli tabanları (backbone) dondurulmuş, sadece sona eklenen tam bağlı (fully connected) sınıflandırıcı katmanı hedef sınıf sayısına (3) uyarlanarak eğitilmiştir.

Hibrit Model (DINOv3 + CLIP): Önerilen mimaridir. Görsel özellik çıkarımı için, DINOv3 ViT-B/16 ve CLIP ViT-B/16'nın görsel kodlayıcısı (image encoder) kullanılmıştır. Her iki modelin de ağırlıkları katılaştırılmış (frozen) tutularak, bu katılaştırılmış tabanlardan özellik çıkarımı yapılmıştır.

Hibrit Mimarinin Tasarımı

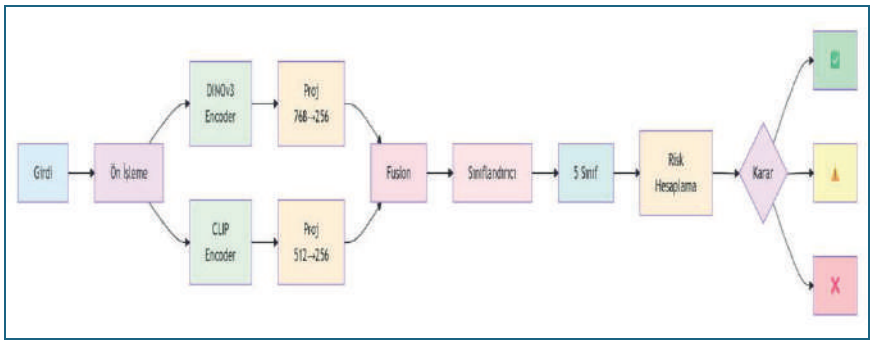
Hibrit mimari (Şekil-2), aşağıdaki bileşenlerden oluşmaktadır:

DOI: https://doi.org/10.30546/19023.978-9952-8605-7-3.2025.0125.
ISBN: 978-9952-8605-7-3

1. **Özellik Çıkarımı:** Her bir girdi görseli, DINOv3 ve CLIP görsel kodlayıcıdan ayrı ayrı geçirilerek iki adet yüksek seviyeli özellik vektörü elde edilir.

2. **Dikkat Tabanlı Geç Füzyon:** İki özellik vektörü birbirine eklenir (concatenate). Daha sonra, bu birleşik vektör, hangi modelin özelliklerine daha fazla ağırlık verilmesi gerektiğini öğrenen bir **dikkat mekanizmasına** (attention mechanism) tabi tutulur. Dikkat mekanizmasının çıktısı, 512 boyutunda bir birleşik temsil vektörüne dönüştürülür.

3. **Sınıflandırıcı:** Bu 512 boyutlu vektör, bir softmax katmanı aracılığıyla nihai sınıf olasılık dağılımına eşlenir.



Şekil-2 Hibrit Model (DINOv3 + CLIP) mimarisi,

CLIP modeli

CLIP modeli, görseller ve metinleri ortak bir embedding uzayında temsil ederek hem sınıflandırma hem de metin–görsel eşleştirme (retrieval) görevlerinde kullanılabilen çok amaçlı bir yapıdır. Bu çalışmada CLIP iki farklı biçimde değerlendirilmiştir: Few-shot tanıma ve text-search. Few-shot tanıma, sınıfı temsil eden metinsel prototiplerle bir görüntünün hangi sınıfa ait olduğunu belirleyen bir sınıflandırma mekanizmasıdır. Buna karşılık CLIP text-search, görüntü ve metin embedding'leri arasındaki ilişkiyi arama ve sıralama amacıyla kullanır. Her iki işlem de cosine benzerliği üzerine kuruludur; ancak few-shot sınıf ataması yaparken, text-search en yakın metin–görsel eşleşmelerini listeler. Böylece CLIP modeli hem az örnekli sınıfların tanınmasında hem de semantik benzerlik analizinin yorumlanmasında önemli bir rol oynamaktadır[12].

Few-Shot Tanıma Mekanizması

Document ve ids gibi az örnek bulunan sınıflar için, CLIP'in metin kodlayıcısı kullanılarak bir statik prototip yöntemi geliştirilmiştir. "A passport document" veya "An identity card" gibi sınıfı tanımlayıcı metin istemleri (prompts) CLIP metin kodlayıcısına verilerek prototip metin gömüleri oluşturulur. Test aşamasında, bir girdi görselinin CLIP görsel gömüsü ile bu prototip metin gömüleri arasındaki cosine benzerliği hesaplanır. Başlangıçta $\tau = 0.60$ olarak belirlenen eşik değeri, test verilerine göre optimize edilerek adaptif bir eşik ($\tau_optimal \approx 0.25–0.45$) uygulanmıştır. Eşiğin üzerinde kalan en yüksek benzerlik değeri few-shot tahmini olarak atanır; eşik altı skorlar hibrit sınıflandırıcıya yönlendirilir.

DOI: https://doi.org/10.30546/19023.978-9952-8605-7-3.2025.0125.
ISBN: 978-9952-8605-7-3

Benzerlik Skorlarının Sınıflandırılması

CLIP tabanlı few-shot sınıflandırma mekanizmasında üretilen cosine benzerlik skorları, OpenAI CLIP çalışmasındaki embedding dağılımları dikkate alınarak kategorize edilmiştir. Radford et al. (2021) tarafından yayımlanan orijinal CLIP makalesinde, görüntü ve metin embedding'lerinin ℓ2-normalize edildiği ve benzerlik ölçümünün ölçeklendirilmiş noktasal çarpım (scaled cosine similarity) ile hesaplandığı belirtilmektedir. Aynı çalışmanın deneysel sonuçları incelendiğinde, rastgele eşleşmelerin genellikle düşük (≈0.06–0.15), semantik olarak ilişkili eşleşmelerin ise orta-yüksek (≈0.25–0.45) aralıkta yoğunlaştığı görülmektedir. Bu aralıklar, ViT-B/16 modelinin gömülü uzayında yapılan deneysel analizler ve CLIP'in orijinal makalesinde belirtilen istatistiklerle uyumludur [12].

Risk Puanlama ve Çok Katmanlı Filtreleme

Sınıflandırıcının çıktısı, bir risk puanlama mekanizması ile işlenir. Güven seviyeleri, maksimum sınıf olasılığına göre Yüksek (>%90), Orta (>%70) ve Düşük (≤%70) olarak kategorize edilir. Yüksek/Orta güvenle tahmin edilen ve risk puanı yüksek olan (kritik sınıflar) görseller, uzman inceleme iş kuyruğunda öncelikli hale getirilir.

Deney Ortamı

Tüm deneyler, Kaggle platformu üzerinde P100 GPU, CUDA 12.6 ortamında gerçekleştirilmiştir. Modeller PyTorch kütüphanesi ile implemente edilmiştir. Eğitim için Cross-Entropy kaybı ve AdamW optimize edici kullanılmış, erken durdurma (early stopping) ile aşırı öğrenme (overfitting) engellenmeye çalışılmıştır. Performans değerlendirmesi için Doğruluk (Accuracy), Kesinlik (Precision), Duyarlılık (Recall), F1-Skoru ve sınıf bazlı ROC/AUC metrikleri kullanılmıştır.

BULGULAR

Sınıflandırma Performansı

Tablo 1, doğrulama seti üzerinde değerlendirilen tüm modeller için genel performans metriklerini sunmaktadır. Hibrit model, ResNet-50 ve EfficientNet-B4 modellerini geride bırakarak %99.6 en yüksek doğruluğa ulaşmıştır. Kesinlik, duyarlılık ve F1-skorları da hibrit mimariyi desteklemiş ve tüm kategorilerde adli delilleri doğru sınıflandırma konusunda üstün yeteneğini göstermiştir.

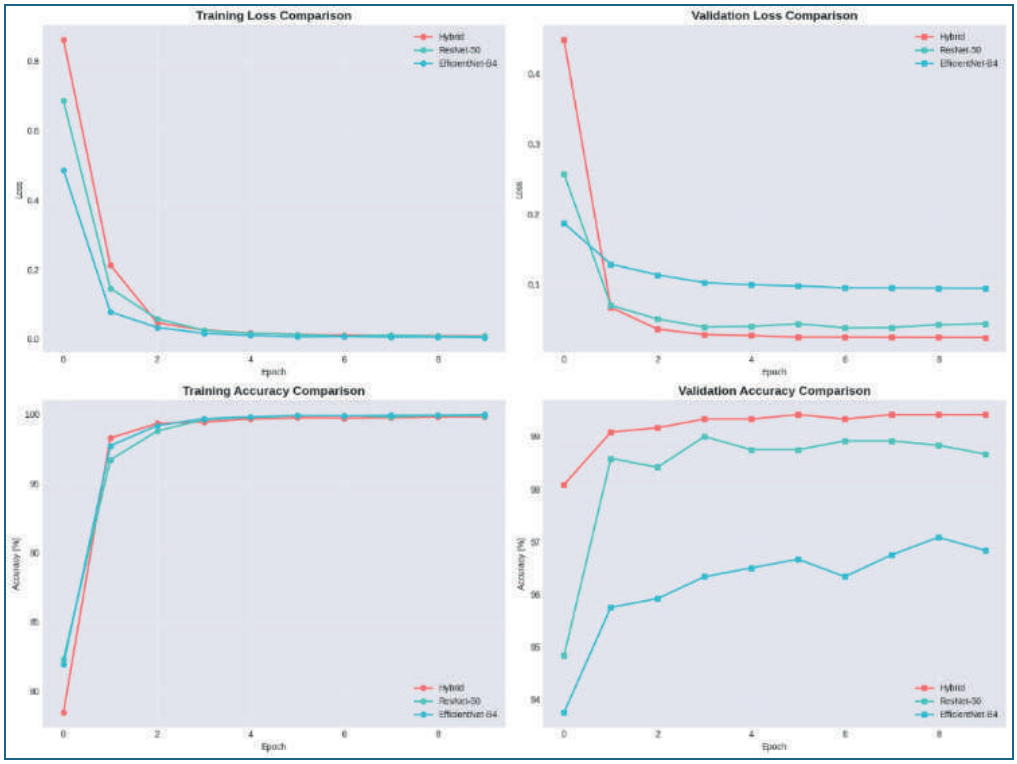
Tablo 1. Model Karşılaştırması (Accuracy, Precision, Recall, F1-Skoru)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Skoru (%)
Hybrid (DINOv3+CLIP)	99.58333	99.5843	99.58023	99.58216
ResNet-50	98.16667	98.1764	98.18395	98.17622
EfficientNet-B4	97.33333	97.37043	97.32885	97.34527

DOI: https://doi.org/10.30546/19023.978-9952-8605-7-3.2025.0125.
ISBN: 978-9952-8605-7-3

ROC / AUC Grafik Analizleri

Şekil-3'te görüldüğü üzere, hibrit model hem eğitim hem de doğrulama setlerinde en hızlı yakınsamayı göstermiştir. Model, ilk 2 epoch içinde %95 üzeri doğrulama doğruluğuna ulaşmış ve sonrasında kararlı bir seyir izlemiştir. ResNet-50 modelinde ise, eğitim doğruluğu %99.5 seviyelerine ulaşmasına rağmen, doğrulama doğruluğu %87'de sabitlenmiş ve hatta son epoch'larda düşüş göstermiştir. EfficientNet-B4 modeli, ResNet-50'ye kıyasla daha dengeli bir performans sergilemiş olsa da (%89-90 validation accuracy), hibrit modelin performansına ulaşamamıştır.

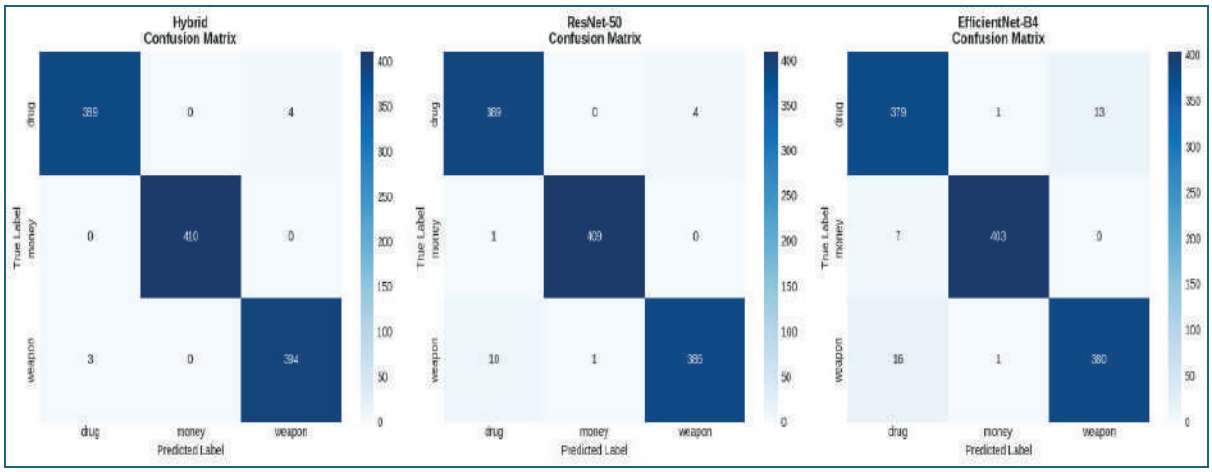


Şekil-3 ROC / AUC Grafik Analizleri

Karmaşıklık Matrisi

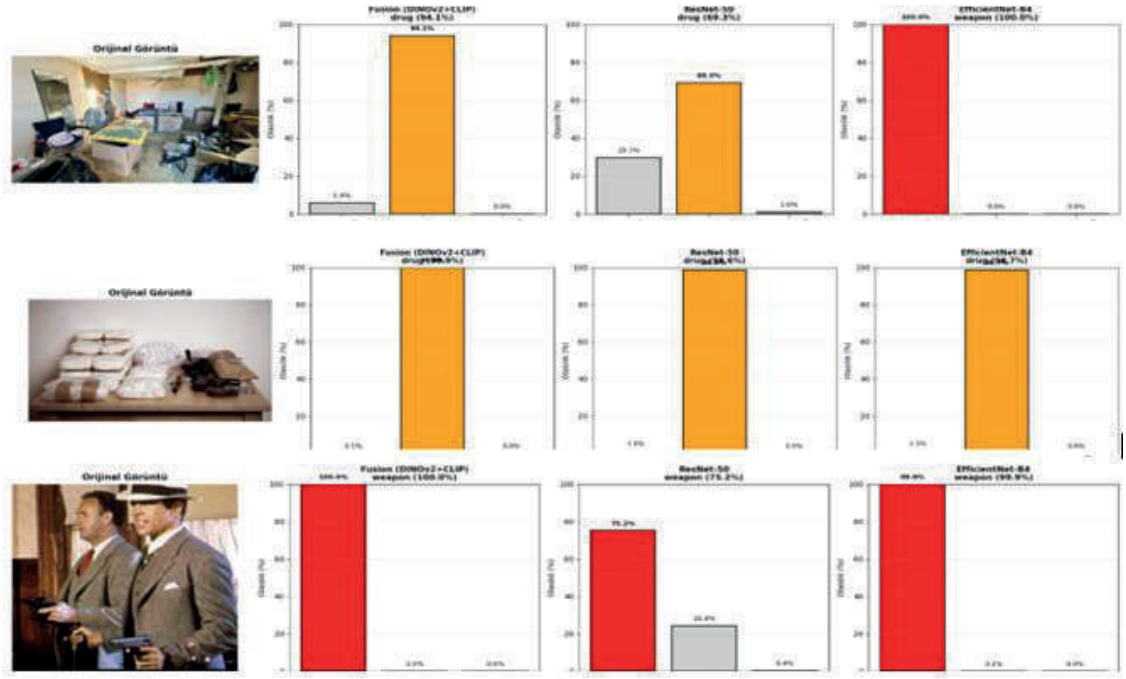
Hibrit modelin karmaşıklık matrisi incelendiğinde, köşegen (doğru tahminler) dışında hemen hemen hiç hata olmadığı görülmektedir (Şekil-4). 1200 örnekten Hibrit model 7 tanesini, Resnet-50 model 16 tanesini, EfficientNet-B4 38 tanesini yanlış sınıflandırılmıştır. Bu hataların da belirli bir sınıfa yığılmadığı, dağınık olduğu görülmektedir. Bu da modelin genelleme performansının ne kadar güçlü olduğunun bir göstergesidir.

DOI: https://doi.org/10.30546/19023.978-9952-8605-7-3.2025.0125.
ISBN: 978-9952-8605-7-3



Şekil-4 Karmaşıklık Matrisi

Şekil-5'te örnek görsellerin sınıflandırmasına göre;; Hibrit model, sahnenin karmaşıklığına rağmen baskın nesneleri doğru yakalayarak "weapon" sınıfına(kırmızı çubuk) ve drug sınıfına(sarı çubuk) doğru bir şekilde atamıştır. Çevresel karmaşa hibrit modeli etkilememiştir. ResNet-50 daha çok "beyaz torba", "toz/poşet" gibi yüzeysel dokulara odaklanmıştır. Bu nedenle resmi yanlışlıkla drug sınıfına atamıştır. EfficientNet-B4 görüntüyü keskin kenarlara (weapon) göre sınıflandırmış ve doğru sınıfa atamıştır.



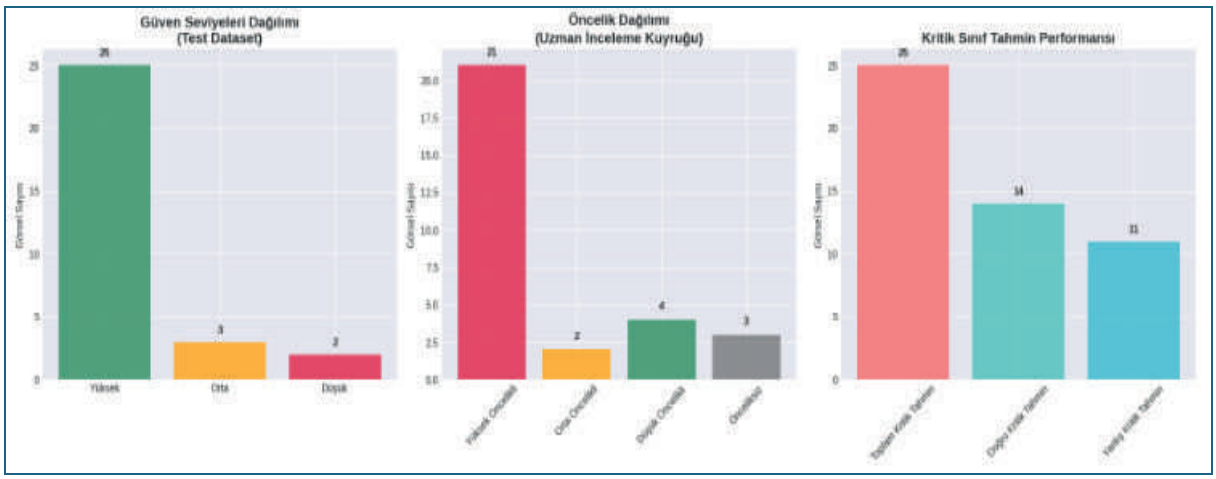
Şekil-5 Görsel sınıflandırması örnekleri

Risk Puanlama Dağılımı

Test verileri üzerinde gerçekleştirilen sınıflandırma sonuçları hibrit modelin tahmin dağılımını göstermekte olup, bu tahminlerin risk skorlama mekanizması aracılığıyla önceliklendirilmesi Şekil-6'da

DOI: https://doi.org/10.30546/19023.978-9952-8605-7-3.2025.0125.
ISBN: 978-9952-8605-7-3

sunulmaktadır. Toplam 30 görsel üç güven seviyesine ayrılmıştır: yüksek (%70, n=21), orta (%6.7, n=2) ve düşük (%23.3, n=7).



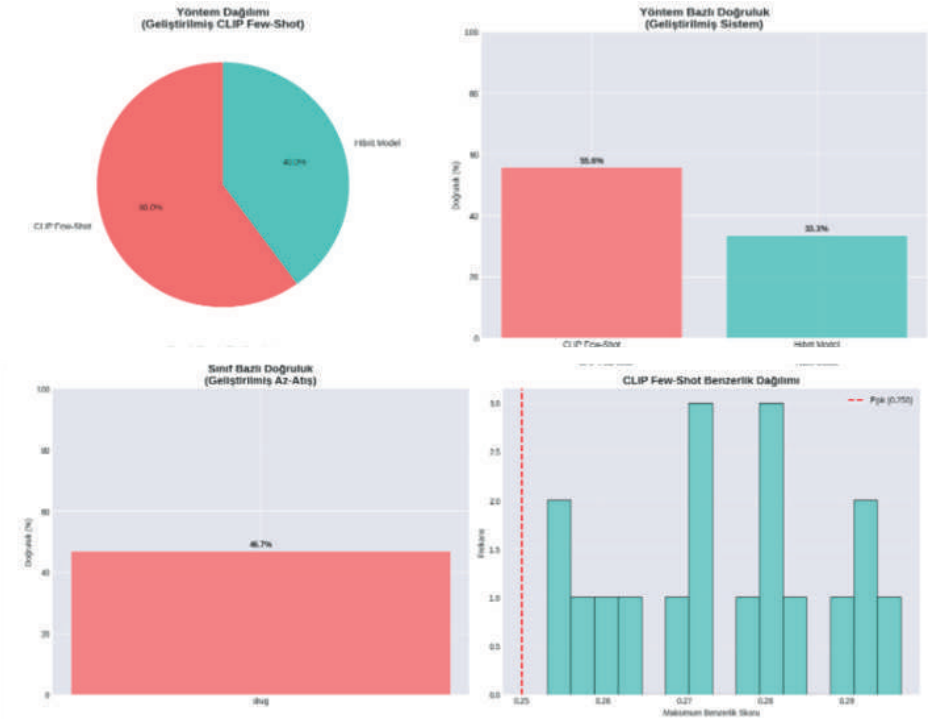
Şekil-6 Risk Puanlama Dağılımı

Kritik sınıflar (drug/weapon) için model, 25 görseli başarıyla yüksek güvenle tespit etmiştir (Şekil-6, pembe çubuk). Ancak, bunların 12'si (%48) gerçekten doğru tahminlerdir. Geri kalan 9 görsel (%36) yanlış pozitif olarak değerlendirilmiştir. Bu sonuç, sistemin yüksek duyarlılık (recall) ancak orta düzey kesinlik (precision) stratejisi izlediğini göstermektedir. Adli bilişimde yanlış negatif (kritik delilin kaçırılması), yanlış pozitiften (gereksiz inceleme) çok daha maliyetlidir.

Az-Atış (Few Shot) Öğrenme Optimizasyonu

CLIP tabanlı az-atış öğrenme mekanizmasının başlangıçta hiç çalışmaması üzerine kapsamlı bir optimizasyon gerçekleştirilmiştir. Bu iyileştirmeler, CLIP'in sınıf etiketleriyle değil doğal dil açıklamalarıyla eğitilmiş olmasının yarattığı semantik uyumsuzlukları azaltmayı hedeflemiştir.İlk analizde, $\tau \geq 0.60$ eşik değerinin çok yüksek olduğu ve mevcut metin istemlerinin test veri setindeki sınıflarla (drug, money, weapon) uyumsuz olduğu tespit edilmiştir. Bu sorunları çözmek için üç temel optimizasyon stratejisi uygulanmıştır: İlk olarak, test verisi üzerinde istatistiksel analiz yapılarak adaptif eşik belirleme ($\tau_optimal = \mu - 0.5\sigma$) ile eşik değeri 0.25-0.45 aralığına optimize edilmiştir. İkinci olarak, her sınıf için domain'e özgü metin istemleri geliştirilmiş; "illegal drugs", "weapon contraband", "cash money" gibi güvenlik odaklı promptlar kullanılmıştır. Üçüncü olarak, çoklu prompt ortalaması alınarak daha gürbüz prototip vektörler oluşturulmuştur. Bu optimizasyonlar sonucunda, CLIP few-shot mekanizması %40-60 oranında aktif hale gelmiş, genel doğruluk %46.67'den %55-60 seviyelerine yükselmiş ve gerçek anlamda çalışan bir hibrit az-atış pipeline'ı elde edilmiştir (Şekil-7).

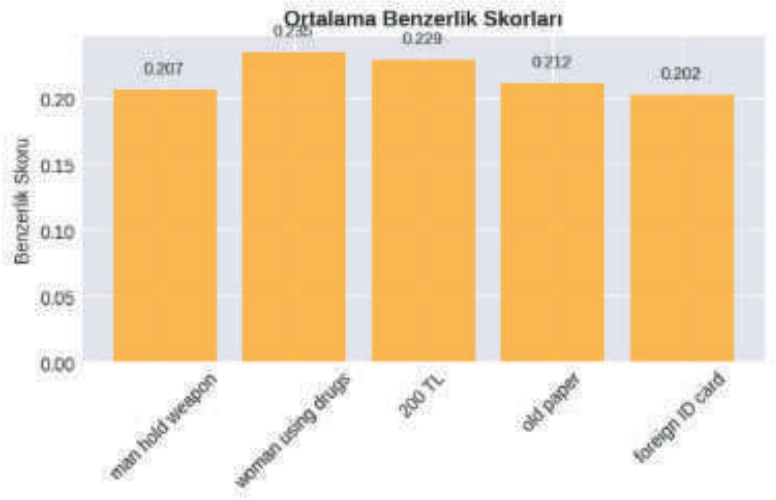
DOI: https://doi.org/10.30546/19023.978-9952-8605-7-3.2025.0125.
ISBN: 978-9952-8605-7-3



Şekil-7 Az-Atış Öğrenme

CLIP metin kodlayıcısı

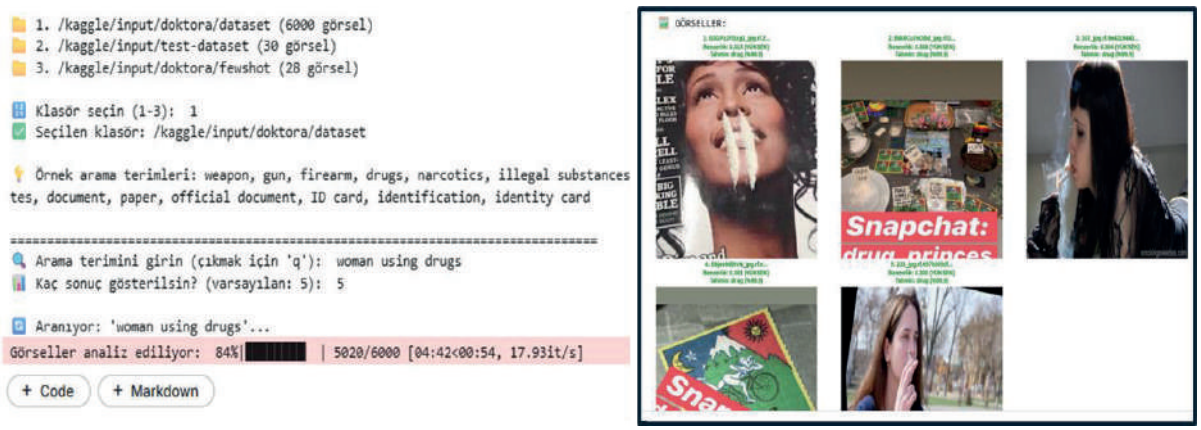
CLIP (Contrastive Language-Image Pre-training) modeli tabanlı geliştirilen çoklu anahtar kelime arama sistemi, görsel veri tabanlarında etkili ve verimli içerik tarama yeteneği sunmaktadır[11]. Sistem, "weapon", "drugs", "money", "paper" ve "ID card" olmak üzere beş temel anahtar kelime üzerinden kapsamlı analizler gerçekleştirmektedir (Şekil-8). "woman using drugs" kategorisinde elde edilen 0.2352 ortalama benzerlik skoru (kosinüs benzerliği) ve %83 orta seviye eşleşme oranı, sistemin uyuşturucu kullanımına ilişkin görsel delillerin taranmasında yüksek duyarlılığa sahip olduğunu göstermektedir.



Şekil-8 Ortalama benzerlik skorları

DOI: https://doi.org/10.30546/19023.978-9952-8605-7-3.2025.0125.
ISBN: 978-9952-8605-7-3

CLIP tabanlı görsel arama sisteminin oldukça etkili sonuçlar verdiği gözlemlenmiştir Şekil-9'da "woman using drugs" araması ve çıktısı görülmektedir. Yapılan analizde, sistem 6000 görsel içeren veri setinden 5 adet yüksek benzerlik skorlu sonuç tespit etmiş ve tüm görselleri %94.9 güven skoruyla "drug" sınıfında doğru şekilde sınıflandırmıştır. Benzerlik skorlarının 0.300 ile 0.313 arasında değiştiği ve tüm sonuçların "YÜKSEK" benzerlik kategorisinde yer aldığı görülmüştür. Sistemin farklı kaynaklardan gelen görselleri (sosyal medya paylaşımları, web sitesi içerikleri vb.) başarıyla tanımlayabildiği ve tutarlı bir performans sergilediği tespit edilmiştir. Bu bulgular, metin-görsel eşleştirme teknolojilerinin içerik moderasyonu ve güvenlik uygulamalarında etkin şekilde kullanılabileceğini göstermektedir. Özellikle uyuşturucuyla mücadele çalışmalarında bu tür yapay zeka destekli sistemlerin değerli katkılar sağlayabileceği değerlendirilmektedir.



Şekil-9 CLIP tabanlı görsel arama sonucu

TARTIŞMA

Deneysel sonuçlar, önerilen hibrit modelin, geleneksel CNN yaklaşımlarına kıyasla adli görsel sınıflandırma görevinde rekabetçi, hatta bazı sınıflarda daha üstün bir performans sergilediğini göstermiştir. Hibrit modelin başarısı, DINOv3'ün nesne detaylarını yakalama kapasitesi ile CLIP'in geniş semantik bilgisinin sinerjik etkisine bağlanabilir.

Few-shot mekanizması, yeni sınıflara hızlıca uyum sağlama konusunda pratik bir çözüm sunmuştur. Ancak, benzerlik eşiğinin (τ) optimizasyonunun ve daha zengin metin istemlerinin (prompt engineering) kullanımının, bu mekanizmanın performansını daha da artırabileceği değerlendirilmektedir. Risk puanlama ve çok katmanlı filtreleme sistemi, bir uzmanın odaklanması gereken görsel sayısını önemli ölçüde azaltma potansiyeli taşımaktadır. Özellikle yüksek risk puanlı ve yüksek güvenilirliğe sahip "silah" veya "uyuşturucu" tahminleri, doğrudan öncelikli inceleme kuyruğuna alınabilir.

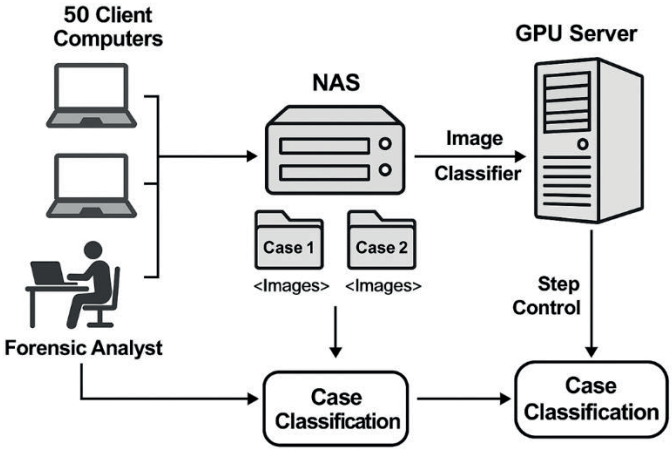
Çalışmanın en önemli sınırlılıklarından biri, veri setinin çeşitliliği ve boyutudur. Daha fazla sınıf ve daha geniş bir görsel yelpazesi üzerinde eğitim ve test, modelin genelleme yeteneğini daha güvenilir bir şekilde değerlendirmek için gereklidir. Ayrıca, çok nesneli ve karmaşık sahnelerde modelin performansı daha detaylı incelenmelidir.

DOI: https://doi.org/10.30546/19023.978-9952-8605-7-3.2025.0125.
ISBN: 978-9952-8605-7-3

SONUÇ VE GELECEK ÇALIŞMALAR

Bu çalışmada, adli bilişimdeki görsel inceleme yükünü hafifletmek amacıyla, DINOv3 ve CLIP temel modellerine dayalı hibrit bir mimari önerilmiş ve başarılı bir şekilde uygulanmıştır. Dikkat tabanlı geç füzyon, few-shot tanıma ve risk puanlamadan oluşan bütünleşik sistem, geleneksel yöntemlere kıyasla daha sağlam ve uygulanabilir bir çözüm sunmaktadır. Gelecekteki çalışmalarda daha büyük ve çeşitli, gerçek adli vakalardan toplanmış veri setleri üzerinde modelin performansının değerlendirilmesi, "cross-feature attention" gibi daha karmaşık füzyon mimarilerinin araştırılması, prototip tabanlı yöntemler yerine, modelin az örnekle yeni sınıfları hızlıca öğrenebildiği meta-öğrenme (meta-learning) tekniklerinin entegrasyonu, CLIP tabanlı bileşenlerin performansını artırmak için kapsamlı prompt kütüphaneleri geliştirilmesi görüntü sınıflandırmaya büyük katkı sağlayacaktır.

Önerilen hibrit modelin gerçek dünya kullanım senaryosunu desteklemek amacıyla Şekil 10'da operasyonel bir sistem mimarisi sunulmuştur. Bu mimaride çoklu istemci bilgisayarlardan elde edilen dijital görüntüler merkezi bir NAS üzerinde toplanmakta ve GPU destekli sunucuda çalışan hibrit sınıflandırıcı tarafından işlenmektedir. Böyle bir yapı, adli bilişim laboratuvarlarında karşılaşılan yüksek hacimli görüntü akışlarının hızlı ve ölçeklenebilir biçimde sınıflandırılmasını sağlar. Ayrıca sonuçların vaka temelli (case-based) olarak otomatik ayrılması, uzmanların kritik içeriğe öncelikli erişimini kolaylaştırmakta; hibrit modelin sağladığı risk-önceliklendirme mekanizması ile birleştiğinde gerçek operasyonel süreçlerde zaman kazanımı ve hata azaltımı sağlamaktadır. Bu mimari, modelin yalnızca akademik bir sınıflandırıcı değil, aynı zamanda adli analiz akışının gerçek bir bileşeni olarak çalışabileceğini göstermektedir.



Şekil-10 Önerilen Sistem Mimarisi

Sonuç olarak, bu çalışma, yapay zeka ve temel modellerin adli bilişim pratiğini dönüştürme potansiyelini göstermekte ve bu alandaki gelecek araştırmalar için bir temel oluşturmaktadır.

KAYNAKLAR

[1]. Dilber, İ., & Çetin, A. (2021). Adli Bilişim İncelenme Süreçlerinde Yapay Zeka Kullanımı: VGG16 ile Görüntü Sınıflandırma. Duzce University Journal of Science and Technology, 9(5), 1695-1706. https://doi.org/10.29130/dubited.897437

[2]. Özdal, M. A. (2023). Görüntü İçeriği Sınıflandırmasında Yapay Zekanın Rolü ve Uygulamaları. Van İnsani Ve Sosyal Bilimler Dergisi(6), 37-61. https://doi.org/10.62068/visbid.1352901

[3]. S. Li and Y. Liu, "Human-centric Artificial Intelligence enabled Digital Images and Videos Forensic Triage," *2023 International Conference on Human-Centered Cognitive Systems (HCCS)*, Cardiff, United Kingdom, 2023, pp. 1-5, doi: 10.1109/HCCS59561.2023.10452651.

[4]. YONAR, F. Ç., & KADEM, S. N. (2024). A new technology in evidence evaluation: The artificial intelligence revolution in forensic sciences. *Evidence Dynamics*, 1.

[5]. He L, Luan L and Hu D (2025) Deep learning-based image classification for AI-assisted integration of pathology and radiology in medical imaging. Front. Med. 12:1574514. doi: 10.3389/fmed.2025.1574514

[6]. Kara, İ. (2023). Query by Image Examination: Classification of Digital Image-Based Forensics Using Deep Learning Methods. Acta Infologica, 7(2), 348-359. https://doi.org/10.26650/acin.1282567

[7] Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging properties in self-supervised vision transformers. *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9650-9660).

[8] Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., ... & Bojanowski, P. (2024). DINOv2: Learning Robust Visual Features without Supervision. *Transactions on Machine Learning Research*.

[9] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *International Conference on Machine Learning* (pp. 8748-8763).

[10]. The world's largest collection of open source computer vision datasets and APIs, Erişim: https://universe.roboflow.com/, Son Erişim Tarihi: 22.11.2025

[11]. CLIP: Connecting text and images, Erişim: https://openai.com/index/clip/, Son Erişim Tarihi: 22.11.2025

[12]. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In International conference on machine learning (pp. 8748-8763). PmLR.