Sürdürülebilir Mühendislik Uygulamaları veTeknolojik Gelişmeler Dergisi
Journal of Sustainable Engineering Applications and Technological Developments
2025, 8(2): 163-172

Video Tabanlı Sınıf Yoklamasının Derin Öğrenme ve Makine Öğrenmesi Temelli Hibrit Bir Yaklaşımla Gerçek Zamanlı Olarak Elde Edilmesi

Real-Time Acquisition of Video-Based Classroom Attendance Using a Hybrid Approach Based on Deep Learning and Machine Learning

Pınar İplikçi Ekincioğlu [1], Serkan Keser [2*]

[1*] Kırşehir Ahi Evran Üniversitesi, Fen Bilimleri Enstitüsü, Elektrik-Elektronik Mühendisliği, Kırşehir, Türkiye

[2] Kırşehir Ahi Evran Üniversitesi, Mühendislik Mimarlık Fakültesi, Elektrik-Elektronik Mühendisliği, Kırşehir, Türkiye

ÖZET

Bu çalışma, sınıf içi video akışından gerçek zamanlı yoklama üretmek üzere yüz algılama, yüz tanıma, öznitelik çıkarımı, çoklu sınıflandırıcı ve çoğunluk oyu yaklaşımını temel alan hibrit bir sistem önermektedir. İlk aşamada Viola–Jones tabanlı kademeli (cascade) algılayıcı ile yüz adayları belirlenir ve bir Evrişimsel Sinir Ağı (CNN) doğrulayıcı model ile "yüz/yüz değil" olarak sınıflandırılarak yanlış pozitifler elenir. Doğrulanan yüzler üzerinde Yönlendirilmiş Gradyan Histogramı (HOG), Evrişimsel Sinir Ağı (CNN) ve AlexNet-fc7 (LEX: *Layer Extraction from AlexNet fc7*) öznitelikleri çıkarılır. Sınıflandırmada Destek Vektör Makineleri (SVM), En Yakın Komşu (KNN), Rastgele Orman (RF), Çift Yönlü Uzun Kısa Süreli Bellek (BiLSTM), Kapılı Tekrarlayan Birim (GRU) ve Evrişimsel Sinir Ağı (CNN) modelleri değerlendirilmiştir. Ayrıca tüm sınıflayıcıların hibrit olarak kullanıldığı ve kararın çoğunluk oyu ile verildiği bir çalışma yapılmıştır. Farklı öğrenci sayıları (4–12) ve çekim senaryolarında önerilen yapı yüksek doğruluk üretmiş; özellikle hibrit, GRU ve BiLSTM modelleri istikrarlı sonuçlar vermiştir. Sistem, ek donanım gerektirmeden yalnızca kamera görüntüsü ve bir bilgisayar yardımı ile müdahalesiz ve hızlı yoklama sağlamaktadır.

Anahtar Kelimeler: Yüz tespiti, yüz tanıma, HOG, LEX, sınıf yoklaması

ABSTRACT

This study proposes a hybrid system based on face detection, face recognition, feature extraction, multi-classifier, and majority voting approach to generate real-time attendance from in-class video streaming. In the first stage, face candidates are identified using a Viola–Jones-based cascade detector, and classified as "face/not face" using a Convolutional Neural Network (CNN) validator model to eliminate false positives. From the verified faces, Histogram of Oriented Gradients (HOG) features, Convolutional Neural Network (CNN) features, and AlexNet-fc7 (LEX: Layer Extraction from AlexNet fc7) representations are extracted. For classification, Support Vector Machine (SVM) with radial basis kernel (RBF), k-Nearest Neighbour (KNN), Random Forest (RF), Bidirectional Long Short-Term Memory (BiLSTM), Gated Recurrent Unit (GRU), and Convolutional Neural Network (CNN) models were evaluated. A hybrid configuration combining all classifiers with majority voting was also implemented.The proposed structure achieved high accuracy under different student counts (4–12) and classroom scenarios, with the hybrid, GRU, and BiLSTM models in particular yielding stable results. The system provides unobtrusive and rapid attendance acquisition using only a camera and a computer, without requiring any additional hardware.

Keywords: Face detection, face recognition, HOG, LEX, class attendance.

Başvuru: 02.10.2025 Son Revizyon: 24.10.2025 Kabul: 26.10.2025
Doi: 10.51764/smutgd.1795569

[2]*Sorumlu yazar: Kırşehir Ahi Evran Üniversitesi, Elektrik Elektronik Mühendisliği, Kırşehir, Türkiye;
[1] E-mail: pinariplikci@gmail.com; ORCID: 0009-0007-8296-5517
[2]* E-mail: skeser@ahievran.edu.tr; ORCID: 0000-0001-8435-0507

1. GİRİŞ

Sınıf ortamlarında yoklama süreci eğitim yönetimi açısından kritik olsa da birçok kurumda halen elle yürütülmektedir. Bu da zaman kaybı, veri hatası ve suistimale açıklık gibi sorunları beraberinde getirmektedir. Video tabanlı ve temassız bir yoklama sistemi, yalnızca bir kamera ile müdahalesiz biçimde öğrenci varlığını kayda geçirerek sürecin güvenilirliğini ve verimliliğini artırabilir. Ancak böyle bir sistemin gerçek zamanlı çalışması değişken aydınlatma, poz ve ifade farklılıkları, kısmi örtüşmeler, ölçek değişimleri, kamera yerleşimi ve hatta CPU/GPU kısıtları gibi bir dizi zorluğu aynı anda ele almayı gerektirir. Erken dönem çözümler elle çıkarılmış öznitelikler (LBP, HOG) ve klasik sınıflayıcılar (SVM, KNN) üzerine kuruluyken, derin öğrenmenin yaygınlaşmasıyla CNN tabanlı yaklaşımlar (VGGFace, FaceNet, ArcFace/InsightFace) kimlik tanımada belirgin ilerleme sağlamıştır.

Bu çalışma, yanlış pozitifleri azaltıp zaman-mekân tutarlılığını güçlendiren hibrit bir yaklaşım önermektedir. Kademeli (kaskad) tespit hattına eklenen bir CNN yüz doğrulama modeli ile yüz-yüz değil ayrımı yapılarak hatalar erkenden elenmiştir. Doğrulanan yüzlerden HOG, CNN ve AlexNet-fc7 öznitelikleri çıkarılır ve SVM, KNN, Random Forest, CNN, BiLSTM ve GRU gibi farklı sınıflayıcıların çıktıları majority/soft voting ile birleştirilir. Amaç, ders süresince alınan video akışında kadrajda görünen yüzleri gerçek zamanlı olarak doğru kişiye atamak ve ders sonunda güvenilir bir yoklama listesi üretmektir. Tasarım tercihleri bu hedefe hizmet edecek şekilde seçilmiştir.

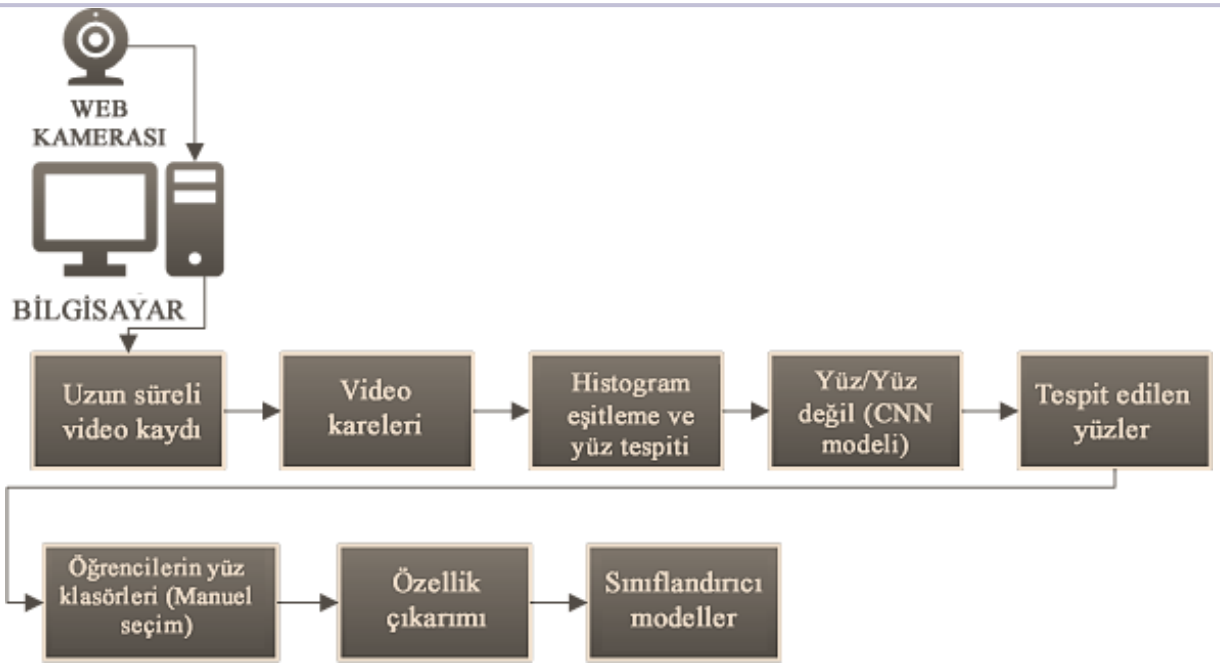
1.1 İlgili Çalışmalar

Sınıf içi video verisinden yoklama çıkarımı, yüz algılama ve kimliklendirme literatürünün iki yönüne birden yaslanır: bir yanda algılama/yerelleme, diğer yanda temsiller ve sınıflandırma bulunur. Erken dönem çalışmalar, gerçek zamanlılık avantajı nedeniyle Viola–Jones tabanlı cascade algılayıcıları tercih etmiş, ancak zorlu aydınlatma ve kalabalık sahnelerde artan yanlış-pozitifler önemli bir pratik sorun olarak raporlanmıştır (Viola & Jones, 2001). Benzer dönemde HOG (Dalal ve Triggs, 2005) ve Local Binary Pattern (LBP, Ahonen vd., 2006) gibi elle çıkarılan öznitelikler, SVM ve KNN gibi klasik sınıflayıcılarla birlikte kullanılmış; küçük-orta ölçekli veri ve sınırlı donanım koşullarında dengeli sonuçlar vermiştir. Derin öğrenmenin yaygınlaşmasıyla birlikte AlexNet (Krizhevsky vd., 2012) türü ağlardan alınan ara katman vektörlerinin "genel amaçlı" temsil gücüne dayanan transfer öğrenme yaklaşımları benimsenmiştir. Özellkle yüz tanımada VGGFace (Parkhi vd., 2015), FaceNet (Schroff vd., 2015) ve açısal-marjine dayalı ArcFace (Deng vd., 2019) gibi yöntemler aydınlatma değişimlerine daha dayanıklı ayrışması kuvvetli temsiller üreterek doğruluğu belirgin biçimde artırdı. Algılama tarafında ise çok-aşamalı MTCNN (Zhang vd., 2016) ve tek-aşamalı RetinaFace (Deng vd., 2020) gibi yöntemler, zorlayıcı sahnelerdeki başarımı yükseltirken, işlem maliyeti uygulama tercihlerini belirler hâle geldi. Bu nedenle son yıllarda literatürde iki eğilim dikkat çekmektedir. İlki hızlı ama hataya açık algılayıcıların üstüne bir CNN doğrulayıcı ekleyerek yanlış pozitifleri elemek ve ikinci olarak derin öznitelikleri klasik sınıflayıcılarla veya küçük CNN/RNN mimarileriyle birleştirerek gerçek zamana yakın çalışan hibrit modeller kurmaktır (Patil ve Shukla, 2014).

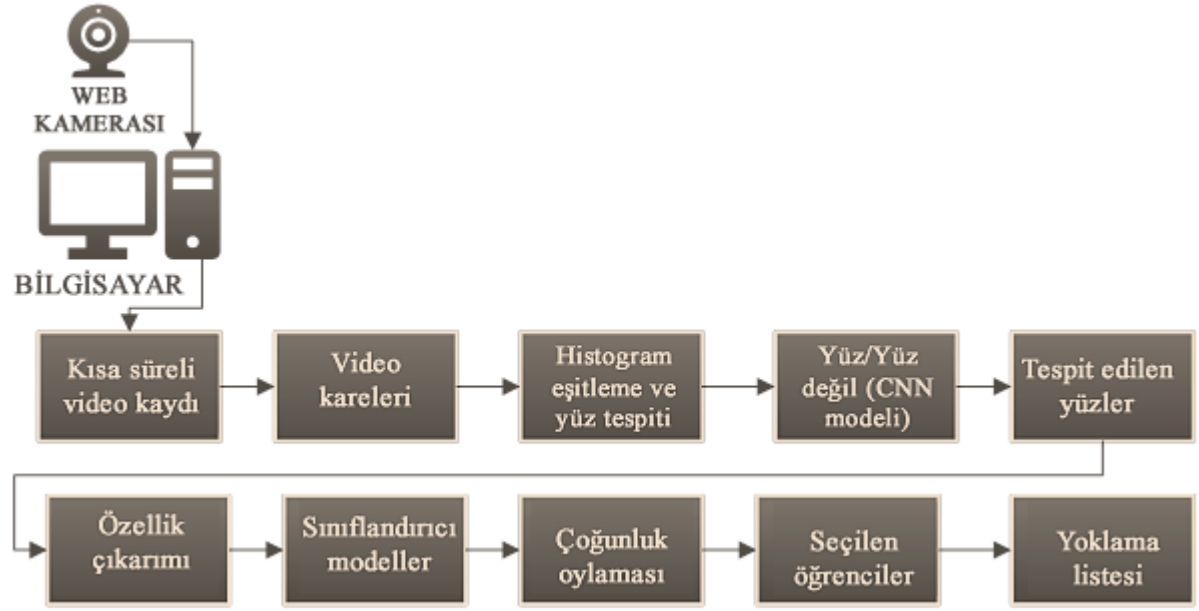
Video tabanlı senaryolarda tek-kare tahminlerin dalgalanması ve kısa süreli tıkanıklık etkileri kararların zamansal bağlam ile dengelenmesini gerektirir. Bu amaçla iki yol yaygındır. İlki basit ama etkili oylama/pencereleme stratejileri ve sıralı modellerdir. Oylama yaklaşımları, ardışık karelerdeki sınıf olasılıklarını birleştirerek anlık hataları sönümler; sıra düzensel modeller ise doğrudan zaman bilgisini öğrenir. Özellikle BiLSTM (Hochreiter ve Schmidhuber, 1997) ve GRU (Cho vd., 2014), yüz temsillerini kısa sekanslar hâlinde tüketerek mikro-hareketler ve poz geçişlerinden kaynaklanan hataları azaltır; böylece video-düzeyi doğruluk tek-kare doğruluğunun üzerine çıkabilir. Tüm bu durumlar pratik kısıtları (tek kamera, CPU ağırlıklı işlem) ve gizlilik gereksinimlerini (yerinde işleme, asgari veri saklama) gözeten "hibrit" çözümleri öne çıkarır. Bu çalışmanın katkısı da bu bağlamda açığa çıkmaktadır. Bu katkılar yoklama sırasında hızlı bir kaskad modelin üstüne yerleştirilen CNN yüz doğrulama ile yanlış pozitiflerin erkenden elenmesi, öznitelik olarak HOG, CNN ve AlexNet-fc7'nin kullanılması, karar katmanında sınıflayıcıların çıktılarının majority/soft voting ile birleştirilmesi olarak sıralanabilir.

2. MATERYAL VE METOD

Aşağıda Şekil 1(a)'da modellerin eğitilmesi ve oluşturulması için kullanılan eğitim blok diyagramı verilmiştir. Şekil 1(b)'de ise yoklama sonucunun elde edildiği test blok diyagramı verilmiştir. Şekil 1(a)'da sınıf içi yoklama sürecinin uçtan uca nasıl yürütüldüğünü tek hat üzerinde göstermiştir. Web kamerasından alınan görüntü akışı bilgisayarda uzun süreli video kaydına dönüştürülmüş, kayıt karelere ayrıştırılmış, her karede histogram eşitleme ile parlaklık/kontrast dengesi iyileştirilmiş ve yüz adayları tespit edilmiştir. Bulunan adaylar hafif bir CNN sinir ağı ile "yüz/yüz değil" olarak doğrulanmış, yüz olmayan bölgeler elenmiş ve geçerli yüz kırpımları elde edilmiştir. Başlangıçta öğrencilerin örnek yüzleri ilgili klasörlere manuel olarak ayrılmış ve referans kümesi oluşturulmuştur. Geçerli yüzlerden ayırt edici öznitelikler çıkarılmış, bu öznitelikler eğitimli sınıflandırıcı modeller tarafından değerlendirilmiş ve her yüz doğru öğrenci kimliğine atanmıştır. Böylece tüm işlem cihaz üzerinde tamamlanmış ve ders sonunda güvenilir bir yoklama listesi üretilmiştir.



(a)



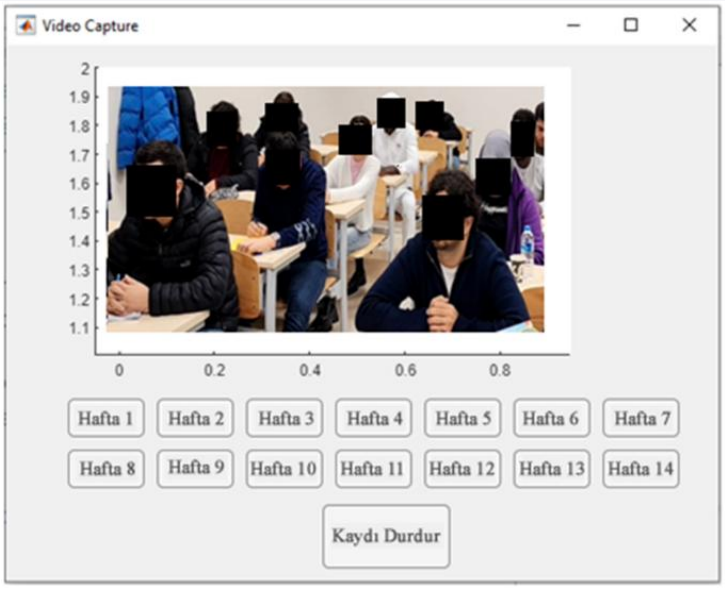
(b)

Şekil 1. Sınıf yoklama sisteminin eğitim (a) ve test (b) aşamaları

Şekil 1(b)'de ise web kamerasından alınan kısa süreli video kaydının bilgisayarda karelere ayrıştırıldıktan sonra histogram eşitleme ile iyileştirilip yüz tespiti yapılmasıyla başlayan boru hattını göstermiştir. Bulunan adaylar hafif bir CNN ile "yüz/yüz değil" olarak doğrulanmış, böylece yalnızca geçerli yüzler sonraki aşamalara aktarılmıştır. Geçerli yüzlerden öznitelikler çıkarılmış, bu öznitelikler eğitimli sınıflandırıcı modellerce değerlendirilmiş ve ardışık karelerde çoğunluk oylaması uygulanarak anlık hatalar sönümlenmiştir. Oylama sonucunda seçilen öğrenciler kimlikleriyle eşleştirilmiş ve süreç sonunda otomatik yoklama listesi üretilmiştir.

2.1. Deneysel Kurulum ve Ortam

Çalışmayı tipik bir derslik düzeninde, yalnızca bir webcam ve bir notebook kullanarak yürüttük. Kamera, tahtayı ve ilk iki-üç sırayı birlikte görecek şekilde sınıfın ön bölgesine konumlandırıldı. Böylece öğrencilerin yüzleri çok ağır yakın plan olmadan, doğal bir açıyla görüntülendi. Kayıtlar 1280×720 çözünürlükte ve 30 fps hızında alındı. Notebook tarafında CPU tabanlı bir sistem kuruldu. Böylece okullarda ek donanım yatırımı gerektirmeden sistemin uygulanabilirliğini göstermek istedik. GPU mevcut olduğunda özellikle AlexNet'in fc7 katmanından öznitelik çıkarma süresi belirgin düşüyor; ancak tüm deneylerimiz CPU ile de gerçek zamana yakın hızlarda sürdürülebilir oldu. Şekil 2'de 14 haftalık ders listesi alma butonlarını içeren MATLAB GUI verilmiştir.



Şekil 2. Webcam kamera kullanılan yoklama alma GUI programı

2.2. Veri Toplama ve Etiketleme

Veri farklı ders seanslarından elde edilen yaklaşık 25 dakikalık video kesitlerinden oluşuyor. Her seans öncesinde, öğrencilerden kısa referans kareleri alarak kimlik eşlemesini hazırladık; bu karelerden üretilen yüz kırpımları, daha sonra sınıf içi görüntülerde gözlenen yüzlerle eşleştirildi. Eğitim ve test bölümleri kişi düzeyinde ayrı oluşturularak bir öğrencinin görüntülerinin aynı anda hem eğitim hem test setlerinde yer almaması sağlandı. Bu düzen, modelin "aynı kişiyi görmüş olmanın" getirdiği yapay performans artışını önledi. Eğitim için bir öğrenciye ait tüm görüntülerin %95'i, test için ise %5'i kullanıldı. Aşağıdaki Tablo 1'de on iki öğrenciye ait görüntülerin eğitim ve test sayıları belirtilmiştir.

Tablo 1: Öğrencilere ait yüz resimleri sayıları

Alt Klasör	Toplam Yüz Sayısı	Eğitim Seti (%95)	Test Seti (%5)
sa	1612	1531	81
sb	2000	1900	100
sc	1472	1398	74
sd	2712	2576	136
se	1483	1409	74
sf	3046	2894	152
sg	1427	1356	71
sh	1684	1600	84
si	1041	989	52
sj	1924	1828	96
sk	1369	1301	68
sl	1374	1305	69

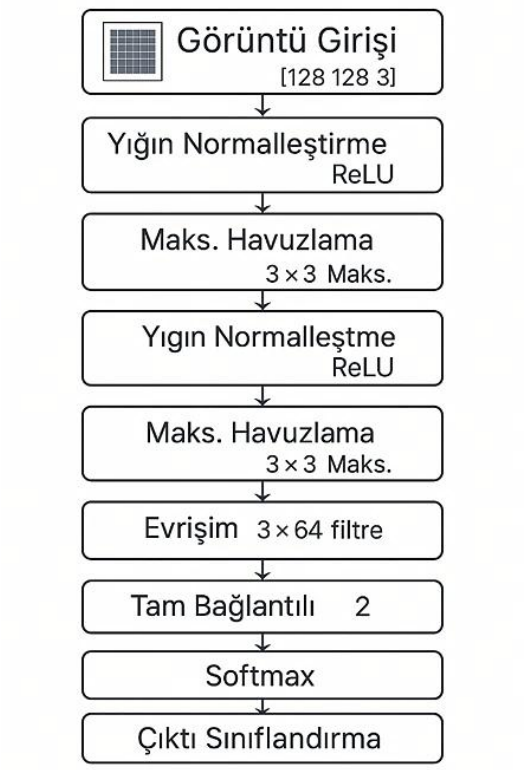
2.3. Ön-işleme Adımları

Videodan ayrılan her kare üzerinde bir ön-işleme işlemi uygulanmıştır. Aydınlatma dalgalanmalarına karşı nispeten daha tutarlı davrandığı için gri tonlu görüntüler kullanılmıştır. Ayrıca global histogram eşitleme kullanımı anlamlı iyileşmeler sağlamıştır. Renk uzayı veya kontrast müdahalesinin ardından, kırpılacak yüz yamasını 128×128 boyutuna normalize edilmiştir. Agresif ön-işlemenin özellikle HOG üzerinde olumsuz etki yaptığını, hafif ve tutarlı adımların daha iyi sonuç verdiği görülmüştür.

2.4. Yüz Adaylarının Bulunması ve Doğrulanması

Yüz algılayıcı iki aşamalı olarak planlanmıştır. İlk aşamada hızlı bir Viola–Jones taramasıyla kare üzerinde olası yüz bölgeleri çıkarılmıştır. Bu algoritmanın tercih edilme sebebi, CPU üzerinde gerçek zamana yakın çalışabilmesi ve geniş bir görüş alanında hızlı tarama yapabilmesidir. Ancak bu hızın doğal bir yan etkisi olarak yanlış pozitif sayısı artabilmiştir. İkinci aşamada, her bir aday bölgeyi küçük ve hızlı bir CNN yüz doğrulama ile "yüz/yüz değil" olarak doğruladık. Bu CNN modelinde üç konvolüsyon bloğu (16–32–64 filtre, BatchNorm + ReLU) ve iki MaxPool (2×2) bulunmaktadır. Son kısımda global average pooling ve iki sınıflı bir tam bağlı katman bulunmaktadır. Eğitimde çapraz entropi kaybı ve Adam optimizasyon (öğrenme oranı=1e−3) kullanılmıştır. Batch boyutu 16 ve 20 epok yeterli sonucu vermiştir. Bu doğrulayıcı katman, özellikle arka plandaki nesnelerden kaynaklanan sahte yüzleri etkin biçimde eledi ve alt aşamalardaki sınıflandırıcıların hata yükünü kayda değer ölçüde azaltmıştır.

Şekil 3'de yüz doğrulama için kullanılan CNN modeli gösterilmiştir. Kullanılan konvolüsyonel sinir ağı mimarisi, 128×128 boyutunda ve 3 kanallı (RGB) giriş görüntüleriyle başlayarak üç adet evrişim bloğu içermektedir. Her blokta sırasıyla 3×3 boyutunda filtrelere sahip konvolüsyon katmanı, ardından batch normalization ve ReLU aktivasyon fonksiyonu yer almakta; her iki blok sonrasında 2×2 boyutlu ve 2 adım atlamalı (stride) maksimum havuzlama katmanı uygulanmaktadır. İlk blokta 16, ikinci blokta 32 ve üçüncü blokta 64 filtre kullanılmıştır. Son evrişim bloğunun ardından gelen tam bağlantılı katman, iki nöronla sınıflandırma görevini gerçekleştirmek üzere yapılandırılmıştır. Bu iki nöron yüz/yüz değil sınıflarına karşılık gelmektedir. Bunu takip eden softmax katmanı, sınıflandırma olasılıklarını hesaplayarak nihai kararın verildiği çıktı katmanına veri aktarmaktadır. Bu yapı, düşük hesaplama maliyetiyle etkili bir özellik çıkarımı ve sınıflandırma sağlamaktadır.



Şekil 3. CNN yüz doğrulama modeli yapısı

Modelin değerlendirme ilk aşamada yüz algılama doğrulama modeli (net_face_find) yüklenerek başlatılmaktadır. Algılanan yüzlerin saklandığı detected_faces klasöründen tüm yüz görüntüleri okunmakta ve her biri işlenmek üzere sisteme alınmaktadır. Görsellerin renk formatı kontrol edilerek renkli olmayanlar üç kanallı (RGB) hale dönüştürülmektedir. Daha sonra, her yüz görüntüsü, eğitilmiş derin öğrenme modeli kullanılarak "isRealFace" fonksiyonu aracılığıyla doğrulanmaktadır. Yüz doğrulama işleminde, her görüntü 128×128 piksele ölçeklendirilerek sinir ağı modeline giriş olarak verilmiştir. Model tarafından "real_face" olarak sınıflandırılan yüzler "output_faces" klasörüne kaydedilmektedir. Geçersiz veya tanınmayan yüzler, işlem sırasında filtrelenmekte ve saklanmamaktadır. Çalışmada Viola–Jones'un tipik false-positive oranı ~%15 civarı iken, CNN doğrulayıcı sonrası bu oran yaklaşık %1–1.5'e düşmüştür. Ayrıca, kodun hata yönetimi mekanizması sayesinde, işlenemeyen resimler tespit edilerek kullanıcıya raporlanmaktadır. Bu yöntem, yüz tanıma sürecinde yanlış pozitiflerin önüne geçmek, düşük kaliteli veya hatalı algılanmış yüzleri sistemden temizlemek ve doğruluk oranını artırmak için geliştirilmiştir. Böylece, öğrenci yoklama sistemine yalnızca güvenilir yüz görüntüleri dahil edilerek daha sağlam ve tutarlı bir yoklama listesi oluşturulması sağlanmıştır.

2.5. Öznitelik Çıkarımı (HOG, CNN ve LEX)

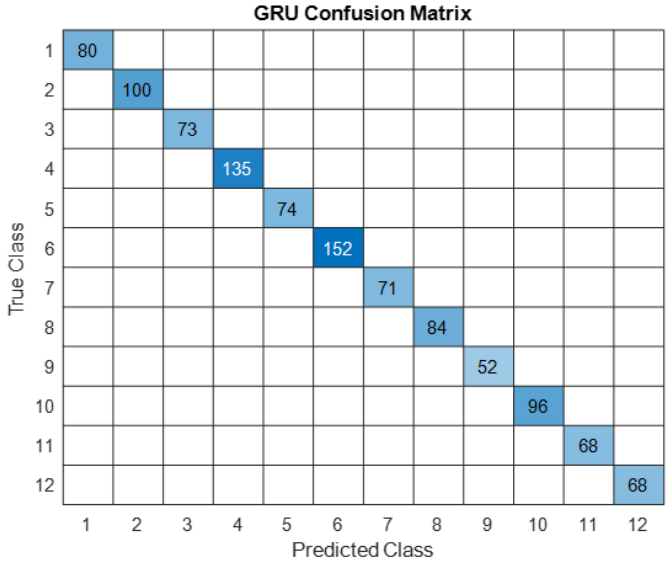
Sınıf içi yüz temsili için bir CNN modeli oluşturulmuştur. Model 128×128 gri seviye girişleri alacak biçimde tasarlanmış, ardışık üç konvolüsyon bloğu (3×3 filtreler; 16-32-64 kanal), her blokta BatchNorm ve ReLU ile desteklenmiş, iki blok sonunda 2×2 max-pooling uygulanmıştır. Konvolüsyon katmanlarını izleyen 48 nöronlu tam bağlantılı katmandan alınan "fc_1" ara vektörleri öz nitelik olarak dışa aktarılmış, son katmanda softmax ile sınıflandırma yapılmıştır. Eğitim Adam ile 20 epoch ve mini-batch=16 ayarlarında yürütülmüş, süreç training-progress ile izlenmiş ve özellik çıkarımı uyumlu biçimde entegre edilmiştir.

Ayrıca doğrulanan yüz yaması üzerinde başka bir öznitelik yapısı olan HOG'da kullanılmıştır. Yama 128×128'e ölçeklenmiş, hücre boyutu 8×8, blok 2×2 hücre, 9 yönelim bin'i ve L2-Hys normalizasyonu ile yaklaşık 8100 boyutlu bir vektör elde edilmiştir. Bu yapı, kenar ve tekstür bilgilerini kararlı biçimde temsil etmiş, aydınlatma değişimlerine karşı dayanıklılığı artırmış ve klasik sınıflayıcıların girişinde düşük maliyetle yüksek ayırt edicilik sağlamıştır.

Üçüncü öznitelik çıkarma yöntemi olarak derin temsil için ön-eğitimli AlexNet'in fc7 katmanı da kullanılmıştır. Yüz yaması, ağın standart ön-işleme adımları ile beslenmiş, fc7'den 4096 boyutlu bir vektör çıkarılmıştır. Bu vektör, poz ve aydınlatma değişimlerine karşı daha esnek bir ayrım gücü sunmuş, özellikle sınıf çeşitliliği arttığında genelleme başarımını yükseltmiştir.

2.6. Sınıflandırıcıların Eğitimi ve Ayarı

Sınıflandırma için hem "klasik" hem de "derin" yöntemler değerlendirilmiştir. SVM (RBF) için küçük bir ızgarada arama yapılmış, $C \in \{1, 10, 100\}$ ve $\gamma \in \{1e{-}3, 1e{-}4\}$ aralıkları sınanmıştır. KNN'de $k \in \{1, 3, 5\}$ ve Euclidean mesafe kullanılmıştır. Random Forest için n_estimators ∈ {100, 300} denenmiş, class_weight=balanced ayarının küçük ve dengesiz alt kümelerde fayda sağladığı görülmüştür. Derin tarafta, doğrudan yüz yamalarından sınıflayan bir CNN yüz doğrulama modeli ile zamansal bağlamı modellemek üzere BiLSTM/GRU uygulanmıştır. RNN girişleri, kare başına öznitelik vektörlerinden oluşan kısa sekanslar olarak hazırlanmıştır. Gizli boyut 128, dropout=0.3, Adam (öğrenme oranı=2e−4) ve yaklaşık 25 epoch ile eğitilmiştir. Tüm modellerde hiper parametre seçimi validation setiyle yapılmış ve early stopping ile ezberleme sınırlandırılmıştır. Aşağıdaki Şekil 4'te eğitilmiş bir GRU modelin test görüntüleri ile elde edilmiş karışıklık matrisi verilmiştir. Test sonucu %100 doğruluk oranına sahiptir. Ancak bu test görüntüleri önceden dosyalanmış öğrenci görüntülerinden alınmıştır, yani videodan anlık alınan görüntülerin sonucu değildir. Yani eğitilen modelin ne kadar iyi eğitildiği, ileride anlık yoklama için kullanılacak bu modelin test edilmesiyle anlaşılabilmektedir. Bu yüzden bu karışıklık matrisleri büyük önem taşımaktadır.



Şekil 4. GRU test karışıklık matrisi

Sınıf sahnesi doğası gereği akıcı; tek bir karede yapılacak hata, birkaç kare sonra kendini düzeltebiliyor. Bu nedenle kare bazlı tahminleri kayan pencere mantığıyla bir araya getirip video-düzeyi bir karara dönüştürdük. Pencere boyutunu W=21 kare (~0.7 s @30 fps) civarında tuttuk; daha küçük pencereler gürültüye açık, daha büyük pencereler ise tepki süresini artırma eğilimindedir.

Çalışmadaki gecikmeyi bir bütçeyle izledik: face proposal CPU'da tipik olarak 8–12 ms, CNN validator 2–4 ms, HOG çıkarımı 2–3 ms, fc7 çıkarımı CPU'da 8–12 ms, sınıflandırma çoğu klasik model için 1 ms'in altında, RNN/CNN tarafında 2–5 ms civarında gerçekleşti. Voting işlemi önemsiz bir yük getiriyor (<1 ms). Bu durum, CPU üzerinde

~25–35 ms bandında (yaklaşık 28–40 fps eşleniği) bir uçtan uca gecikmeye karşılık geliyor ve pratikte gerçek zamana oldukça yakın bir deneyim sağlıyor. Deneylerimiz, fc7 maliyetinin GPU ile düşürülmesinin sistemi "ferahlatmakla" birlikte, doğru kurgulanmış bir validator + HOG akışıyla CPU'da da yeterli performans alınabildiğini gösterdi. Çalışma Intel Core i7-9750H @ 2.60 GHz (6 çekirdek), 16 GB RAM ve Windows 11 (64-bit) üzerinde, MATLAB R2023b sürümü kullanılarak yürütülmüştür. Görüntü işleme ve öznitelik çıkarımı için Image Processing ve Computer Vision Toolbox işlevleri (ör. extractHOGFeatures) kullanılmıştır. Klasik sınıflandırıcılar Statistics and Machine Learning Toolbox ile kurulmuş; SVM için fitcsvm, KNN için fitcknn, Random Forest için TreeBagger uygulanmıştır. Derin öğrenme tabanlı katmanlar Deep Learning Toolbox ile tasarlanmış; CNN mimarisi imageInputLayer, convolution2dLayer, batchNormalizationLayer, reluLayer, maxPooling2dLayer, fullyConnectedLayer zinciriyle oluşturulmuş, eğitim trainingOptions('adam') ve trainNetwork ile yürütülmüştür. Zamansal modelleme gereken koşullarda BiLSTM/GRU katmanları (sequenceInputLayer, bilstmLayer, gruLayer) ile kısa sekanslar üzerinde öğrenme yapılmış; çıktı olasılıkları softmaxLayer ile elde edilmiştir. Hibrit yaklaşımında her sınıflandırıcının olasılık skorları aynı aralığa ölçeklenmiş ve aritmetik ortalama ile birleştirilmiştir; nadir eşitliklerde güveni yüksek olan yolun kararı öne alınmıştır. Bu düzen özellikle öğrenci sayısı arttığında kare bazlı dalgalanmaları yatıştırmış ve daha kararlı nihai kararlar üretmiştir.

3. BULGULAR VE TARTIŞMA

Bu dosya, tezde raporlanan sonuçları özetleyen doğruluk tablolarını içerir. Tablolar, küçük sınıf (4–6 öğrenci) ve daha büyük sınıf (8–12 öğrenci) senaryoları için frame-level ve video-level (voting) doğruluk aralıklarını göstermektedir.

Tablo 2. Küçük Sınıf (4–6 öğrenci) için doğruluk (%) oranları

Yaklaşım	Çıktı	Doğruluk (%)
HOG + SVM	Frame-level	96–99
	Video-level (voting)	98–100
LEX (AlexNet-fc7) + SVM/CNN	Frame-level	97–100
	Video-level (voting)	99–100
CNN (uçtan uca)	Frame-level	96–99
	Video-level (voting)	98–100
KNN	Frame-level	96–100
	Video-level (voting)	98–100
Random Forest (RF)	Frame-level	94–99
	Video-level (voting)	97–100
BiLSTM	Frame-level	95–99
	Video-level (voting)	98–100
GRU	Frame-level	96–100
	Video-level (voting)	99–100
Hibrit	Frame-level	97–100
	Video-level (voting)	99–100

Tablo 3. Daha büyük sınıf (7–12 öğrenci) için doğruluk (%) oranları

Yaklaşım	Çıktı	Doğruluk (%)
HOG + SVM	Frame-level	90–94
	Video-level (voting)	93–96
LEX (AlexNet-fc7) + SVM/CNN	Frame-level	92–96
	Video-level (voting)	95–98
CNN (uçtan uca)	Frame-level	91–95
	Video-level (voting)	94–97
KNN	Frame-level	91–95
	Video-level (voting)	94–97
Random Forest (RF)	Frame-level	88–93
	Video-level (voting)	91–95
BiLSTM	Frame-level	92–95
	Video-level (voting)	95–98
GRU	Frame-level	93–96
	Video-level (voting)	96–98
Hibrit	Frame-level	93–97
	Video-level (voting)	96–99

Tablo 2 ve 3'den de görüldüğü gibi çoğunluk oylama uygulanınca çoğu senaryoda doğruluk değerleri 2–4 puan artmıştır. LEX (fc7) temsili, aydınlatma ve poz değişimlerinde HOG'a göre daha esnek davranmış, özellikle büyük sınıflarda birkaç puan öne geçmiştir. GRU ve BiLSTM, kısa sekanslardaki dalgalanmaları bastırmış ve video düzeyi doğruluğu kalıcı biçimde yükseltmiştir. KNN küçük sınıflarda güçlü bir taban sunmuş, veri çeşitliliği arttıkça kararlılığını korumuştur. Random Forest, bazı dengesiz örneklemelerde değişkenlik göstermiş, kalabalık sınıflarda alt bantta kalmıştır. Uçtan uca CNN, veri çeşitliliği yeterli olduğunda kuvvetli sonuçlar üretmiş; veri az olduğunda HOG/LEX + klasik yöntemlere göre daha hassas kalmıştır. Hibrit yaklaşımı, tekil modellerin zayıflıklarını dengelemiş ve en kararlı üst bant sonuçlarını tutarlı biçimde üretmiştir. Küçük sınıflarda tüm yöntemler tavana yakın sonuçlar vermiş; sınıf büyüdükçe LEX ve RNN tabanlı yaklaşımlar belirgin biçimde avantaj sağlamıştır.

Tablo 4'te, AlexNet-fc7 (LEX) öznitelikleri kullanıldığında hemen tüm modellerin K1–K2 koşullarında "tavana" yakın performans verdiğini göstermiştir. Burada K1 on iki öğrenciden ilk altı kişiyi, K2 ise geriye kalan 6 kişiyi göstermektedir. Böylece tüm sınıf test edilmiş olmaktadır. CNN, SVM, RF ve hibrit satırlarında her hücrenin 6/6 olması, her bir koşulda tüm öğrencilerin doğru tanınabildiğini ve fc7 temsilinin sınıflar arasını net biçimde ayırabildiğini ortaya koymuştur. Bu durum, fc7'nin yüksek seviyeli semantik bilgiyi taşıdığı ve küçük sınıf senaryosunda karar sınırlarını belirginleştirdiği için lineer/karar ağaçlı yöntemlerle dahi kolay ayrım sağlanabildiğini düşündürmüştür.

Tablo 4. Altı öğrenci için Alexnet (Fc7) öznitelikleri ile bulunan doğruluk oranları

	K1	**K2**	**K1**	**K2**
CNN	6/6	6/6	6/6	6/6
Bilstm	6/6	6/6	5/6	6/6
GRU	6/6	6/6	5/6	6/6
SVM	6/6	6/6	6/6	6/6
RF	6/6	6/6	6/6	6/6
KNN	5/6	6/6	5/6	6/6
Hibrit	6/6	6/6	6/6	6/6

BiLSTM ve GRU için bazı hücrelerde görülen 5/6, zamansal modellemenin nadiren tek bir öğrencide sapma ürettiğine işaret etmiştir. Bunun tipik nedeni, kısa süreli occlusion, pencere hizasındaki kaymalar ya da sekansın düşük varyanslı parçasında modelin "yakın kimlik"e kayması olabilmiştir; buna rağmen diğer koşullarda 6/6 korunmuş ve genel tablo yüksek doğrulukta kalmıştır. KNN tarafındaki 5/6 durumları ise en yakın komşu yaklaşımının küçük örnek kümelerinde sınır noktalarına duyarlı olmasından kaynaklanmış; tek bir öğrencinin komşuluk yapısı değişince sınıf etiketi kayabilmiştir. Hibrit sonuçlarının bütünde 6/6 gelmesi, tekil modellerin hata kalıplarını birbirini telafi edecek biçimde birleştirmenin işe yaradığını göstermiştir. Bu sayede RNN/KNN kaynaklı tekil kayıplar SVM/CNN/RF'in oylarıyla dengelenmiş ve nihai karar istikrarlı biçimde tavana taşınmıştır. Özetle, fc7 temsili küçük sınıflarda yüksek ayırt edicilik sağlamış; SVM, RF, CNN ve hibrit ile birlikte tam isabet sürekli korunmuştur. RNN'lerdeki nadir 5/6 durumlarının pencere uzunluğu, oylama ve sekans seçimi iyileştirmeleriyle giderilebileceği değerlendirilmiştir.

Tablo 5, öznitelik türleri ile kullanılan sınıflayıcıların frame-level düzeyinde elde ettikleri ortalama doğruluk ve F1 skorlarını göstermektedir. Görüldüğü üzere, HOG öznitelikleri klasik yöntemler (SVM, KNN) ile birlikte kullanıldığında dahi %96–99 doğruluk üretmiş, AlexNet'in fc7 katmanından çıkarılan derin öznitelik (LEX) ise daha yüksek genelleme kapasitesiyle %97–100 aralığında tavana yaklaşmıştır. RNN-tabanlı modeller (BiLSTM, GRU) zamansal tutarlılığı iyileştirerek doğruluk ve F1 skorlarını bir miktar artırmıştır. Tüm bu özniteliklerin hibrit biçimde çoğunluk oylamasıyla birleştirildiği durumda en yüksek ve kararlı sonuçlar (%99 F1) elde edilmiştir. Bu bulgular, farklı öznitelik-sınıflayıcı çiftlerinin birbirini tamamladığını ve hibrit yaklaşımın sistemin genel kararlılığını artırdığını göstermektedir.

Tablo 5. Öznitelik–Sınıflayıcı eşleştirmeleri ve başarımlar (Frame-level)

Öznitelik	**Sınıflayıcı**	**Doğruluk (%)**	**F1 (%)**
HOG	SVM (RBF)	96–99	98
HOG	KNN (3)	96–99	98
LEX (fc7)	SVM	97–100	99
LEX (fc7)	CNN	97–100	99
CNN özellikleri	Random Forest	94–99	96
CNN özellikleri	BiLSTM	95–99	97
CNN özellikleri	GRU	96–100	98
Tümü (hibrit voting)	—	97–100	99

4. SONUÇLAR

Bu çalışma, sınıf içi video akışından yoklama çıkarmak için tasarlanan sistemin tek bir kamera ve sıradan bir dizüstü bilgisayarla dahi gerçeğe yakın zamanda ve güvenilir biçimde çalışabildiğini göstermiştir. İki aşamalı algılama yapısı, hızlı yüz aday üretimini korumuş; hafif bir yüz doğrulayıcı sinir ağı ile yanlış pozitifler daha başta elenmiş ve tüm sürecin dengesi belirgin biçimde iyileştirilmiştir. Yönlendirilmiş Gradyan Histogramları, AlexNet ağının fc7 katmanından alınan derin öznitelikler ve CNN özniteliklerinin kullanımı, küçük veri ve değişken ışık/poz koşullarında tek başına yöntemlerin zayıf kaldığı noktaları kapatmış ve daha dengeli bir genelleme sağlamıştır. Karar katmanında destek vektör makineleri, en yakın komşu, rastgele orman, evrişimsel sinir ağı ve zamansal yinelenen ağlardan (BiLSTM/GRU) gelen çıktılar çoğunluk oylaması ve olasılık ortalaması ile birleştirilmiş; özellikle kalabalık ve hareketli sahnelerde kare bazlı dalgalanmalar bastırılmış ve video düzeyi doğruluk yükseltilmiştir. Uygulanabilirlik bakımından, uçtan uca gecikme bütçesi yalnızca işlemci (CPU) kullanıldığında dahi kabul edilebilir aralıkta kalmıştır. LEX öznitelik çıkarımı grafik işlem birimi ile hızlanmıştır; ancak doğru ayarlanmış bir "yüz doğrulayıcı + HOG" akışı ile grafik işlem birimi olmadan da akıcı bir deneyim elde edilebilmiştir. Bu bulgular, okullarda ek donanım gerektirmeden bir çözümün mümkün olduğunu göstermiştir. Bununla birlikte sert yönlü ışık, uzun süreli örtülme (occlusion) ve çok küçük ya da eğik yüz görünümleri hata olasılığını artırmıştır. Genel olarak, önerilen hibrit yaklaşım; yüksek doğruluk odaklı derin yöntemlerle kurumların maliyet ve altyapı gerçekliği arasında pratik bir denge kurmuştur. Bu çalışmanın gerçek sınıf koşullarında sürdürülebilir doğruluk ve hız dengesini daha da sağlamlaştıracağı öngörülmüştür.

Çıkar Çatışması Beyanı

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

Yazar Katkısı Beyanı

Yazarlar makale hazırlanmasındaki tüm aşamalarda ortak katkı sağlamışlardır.

KAYNAKLAR

Ahonen, T., Hadid, A., & Pietikäinen, M. (2006). Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(12), 2037–2041.

Cho, K., van Merriënboer, B., Gulcehre, C., et al. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. EMNLP, 1724–1734.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE., doi: 10.1109/CVPR.2005.177.

Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. CVPR, 4690–4699, doi: 10.1109/CVPR.2019.00482.

Deng, J., Guo, J., Zhou, Y., et al. (2020). RetinaFace: Single-shot multi-level face localisation in the wild. CVPR, 5202–5211.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. NeurIPS, 1097–1105.

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. BMVC.

Patil, A. R., & Shukla, A. (2014). Automated attendance using face recognition. International Journal of Computer Applications, 103(16), 6–10.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. CVPR, 815–823 , doi: 10.1109/CVPR.2015.7298682.

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. CVPR, 511–518 , doi: 10.1109/CVPR.2001.990517.

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multi-task cascaded convolutional networks. Signal Processing Letters, 23(10), 1499–1503.