

KOCAELİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ
ANABİLİM DALI

YÜKSEK LİSANS TEZİ

GÖRÜNTÜ İŞLEME VE YAPAY ZEKÂ İLE ORTAM DUYGU
YOĞUNLUK ANALİZİ

İSAK YILDIRIM

KOCAELİ 2025

KOCAELİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ
ANABİLİM DALI

YÜKSEK LİSANS TEZİ

GÖRÜNTÜ İŞLEME VE YAPAY ZEKÂ İLE ORTAM DUYGU
YOĞUNLUK ANALİZİ

İSAK YILDIRIM

Dr. Öğr. Üyesi Hikmetcan ÖZCAN

Danışman, Kocaeli Üniv.

.....

Prof. Dr. Sevinç İLHAN OMURCA

Jüri Üyesi, Kocaeli Üniv.

.....

Prof. Dr. Nevcihan DURU

Jüri Üyesi, Kocaeli Sağlık ve Teknoloji Üniv.

.....

Tezin Savunulduğu Tarih: 15.01.2025

ETİK BEYAN VE ARAŞTIRMA FONU DESTEĞİ

Kocaeli Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez/proje çalışmada,

- Bu tezin/projenin bana ait, özgün bir çalışma olduğunu,
- Çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı,
- Bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi,
- Bu çalışmanın Kocaeli Üniversitesi'nin abone olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü'nün belirlemiş olduğu ölçütlere uygun olduğunu,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Tezin/Projenin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez/proje çalışması olarak sunmadığımı,

beyan ederim.

☒ Bu tez/proje çalışmasının herhangi bir aşaması hiçbir kurum/kuruluş tarafından maddi/alt yapı desteği ile desteklenmemiştir.

☐ Bu tez/proje çalışması kapsamında üretilen veri ve bilgiler tarafından no'lu proje kapsamında maddi/alt yapı desteği alınarak gerçekleştirilmiştir.

Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

İsak YILDIRIM

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI

Fen Bilimleri Enstitüsü tarafından onaylanan lisansüstü tezimin/projemin tamamını veya herhangi bir kısmını, basılı ve elektronik formatta arşivleme ve aşağıda belirtilen koşullarla kullanıma açma izninin Kocaeli Üniversitesi'ne verdiğimi beyan ederim. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet hakları bende kalacak, tezimin/projemin tamamının ya da bir bölümünün gelecekteki makale, kitap, tebliğ, lisans, patent gibi çalışmalarda kullanımı bana ait olacaktır.

Tezin/projenin kendi özgün çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin/projenin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim kurulu tarafından yayınlanan **“Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge”** kapsamında tezim aşağıda belirtilen koşullar haricinde YÖK Ulusal Tez Merkezi/ Kocaeli Üniversitesi Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- ☐ Enstitü yönetim kurulu kararı ile tezimin/projemin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir.
- ☐ Enstitü yönetim kurulu gerekçeli kararı ile tezimin/projemin erişime açılması mezuniyet tarihinden itibaren 6 ay ertelenmiştir.
- ☒ Tezim projem ile ilgili gizlilik kararı verilmemiştir.

İsak YILDIRIM

ÖNSÖZ VE TEŞEKKÜR

Görüntü işleme ve yapay zekâ tekniklerinin hızla geliştiği günümüzde, duygusal yoğunluk analizi alanında yapılan çalışmalar da büyük bir ivme kazanmıştır. Çalışmamız kapsamında, toplu ortamlarda duygu yoğunluğunun analiz edilerek çeşitli alanlarda fayda sağlanabilmesi amacıyla uçtan uca bir sistem tasarlanmış ve geliştirilmiştir. Bu çalışma, duygusal analizlerin toplum yararına kullanılabileceği yeni bakış açıları sunmayı hedeflemektedir.

Bu önemli konuda yürüttüğümüz çalışmada, desteğini esirgemeyen danışmanım Sayın Dr. Öğr. Üyesi Hikmetcan Özcan'a, ayrıca eğitim hayatım boyunca bana yol gösteren tüm hocalarıma içten teşekkürlerimi sunuyorum.

Ayrıca hayatım boyunca desteklerini her an hissettiğim çok değerli aileme, MetropolCard ailesine ve her daim yanımda olan çok kıymetli hayat arkadaşıma da sonsuz teşekkürlerimi sunuyorum.

Ocak-2025

İsak YILDIRIM

İÇİNDEKİLER

ETİK BEYAN VE ARAŞTIRMA FONU DESTEĞİ	i
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI	ii
ÖNSÖZ VE TEŞEKKÜR	iii
İÇİNDEKİLER	iv
ŞEKİLLER DİZİNİ	vi
TABLOLAR DİZİNİ	viii
SİMGELER VE KISALTMALAR DİZİNİ	ix
ÖZET	x
ABSTRACT	xi
1. GİRİŞ	1
1.1. Duygu Yoğunluğu Analizi Problemi	1
1.2. Duygu Yoğunluğu Analizi Amacı	5
1.3. Duygu Yoğunluğu Analizi Önemi	5
1.4. Araştırma Kapsamı	6
1.5. Organizasyon	6
2. LİTERATÜR	7
2.1. Yüz Tespiti	7
2.1.1. Geleneksel Yöntemler	8
2.1.2. Modern Yöntemleri	11
2.1.3. Yüz Tespitinde Kullanılan Veri Setleri	15
2.1.3.1. FERET Veri Seti	15
2.1.3.2. FDDB Veri Seti	16
2.1.3.3. Wider Face Veri Seti	16
2.1.3.4. AFW Veri Seti	16
2.2. Duygu Tespiti	17
2.2.1. Temel Duygu Kategorileri	19
2.2.2. Duygu Tespit Modelleri	22
2.2.2.1. EfficientNetB7	22
2.2.2.2. CNN-LSTM	23
2.2.2.3. CNN	24
2.2.3. Duygu Tespitinde Kullanılan Veri Setleri	25
2.2.3.1. AffectNet Veri Seti	26
2.2.4. Duygu Tespitinde Karşılaşılan Zorluklar	29
2.3. Ortam Duygu Yoğunluğu Analizi	30
2.3.1. Yöntemlerin Karşılaştırılması	31
2.3.2. Ortam Uygulamaları	32
2.4. Mevcut Çalışmaların Değerlendirilmesi	34
3. MATERYAL VE YÖNTEMLER	36
3.1. Kullanılan Veri Setleri	36
3.1.1. FER-2013 Veri Seti	36
3.1.2. CK+48 Veri Seti	37
3.2. Yüz Tespiti	38
3.2.1. OpenCV	39
3.2.2. MTCNN Algoritması	42
3.3. Duygu Sınıflandırma Model Eğitimleri	46
3.3.1. EfficientNet-B7	47

3.3.2. ConvLSTM	49
3.3.3. CNN	51
3.4. Model Parametrelerinin ve Sonuçlarının Değerlendirilmesi	53
3.5. Ortam Duygu Yoğunluğu Analizi	54
3.6. Ortam Duygu Yoğunluğu Analiz Yöntemleri	58
3.6.1. Tüm Yüzler ve Ortalama Yoğunluk Analizi	59
3.6.2. Duygu Değişim Süreci ile Ortam Duygu Analizi	61
3.6.3. Eşik Değeri Belirlenmesi ile Ortam Duygu Analizi	63
3.6.4. Anlık Görüntülerden Ortam Duygu Analizi	65
3.6.5. Yüz Lokasyonundaki Değişiklerin Takibi ile Ortam Duygu Analizi	66
4. BULGULAR	69
4.1. Model Performans Sonuçları	69
4.1.1. EfficientNet-B7	69
4.1.2. ConvLSTM	70
4.1.3. CNN	71
4.1.4. Gerçek Zamanlı Test Sonuçları	72
4.2. Ortam Duygu Yoğunluğu Analizi Sonuçları	73
4.2.1. Farklı Yöntemlerin Karşılaştırılması	73
4.2.2. Uygulama Alanlarına Göre Bulgular	75
5. TARTIŞMA	77
5.1. Elde Edilen Sonuçların Değerlendirilmesi	77
5.2. Araştırma Kapsamındaki Sınırlılıklar	80
5.3. Gelecek Çalışmalar İçin Öneriler	81
6. SONUÇLAR VE ÖNERİLER	83
6.1. Araştırmanın Genel Sonuçları	83
6.2. Öneriler ve Uygulamalar	83
KAYNAKLAR	85
EKLER	91
KİŞİSEL YAYIN VE ESERLER	95
ÖZGEÇMİŞ	96

ŞEKİLLER DİZİNİ

Şekil 1.1.	Yüz algılama için özellik tabanlı yaklaşımlar	3
Şekil 2.1.	RFB mimarisi	13
Şekil 2.2.	Inception-V2 mimarisi	13
Şekil 2.3.	MTCNN P-Net mimarisi	14
Şekil 2.4.	MTCNN R-Net mimarisi	14
Şekil 2.5.	MTCNN O-Net mimarisi	14
Şekil 2.6.	Boyutsal modelleme	21
Şekil 2.7.	Değerlik uyarılma döngüsündeki örnek görüntüler	22
Şekil 2.8.	EfficientNetB7 mimarisi	23
Şekil 2.9.	Örnek bir LSTM mimarisi	24
Şekil 2.10.	AffectNet eğitim seti dağılımı	27
Şekil 2.11.	AffectNet doğrulama seti dağılımı	27
Şekil 2.12.	FER pazar payı	34
Şekil 3.1.	FER-2013 veri setine ait temel duygulara örnek görüntüler	37
Şekil 3.2.	FER-2013 Doğrulama ve eğitim veri seti örnek dağılımı	37
Şekil 3.3.	CK+48 veri setine ait temel duygulara örnek görüntüler	37
Şekil 3.4.	CK+48 Doğrulama ve eğitim veri seti örnek dağılımı	38
Şekil 3.5.	Open CV sürüm zaman çizelgesi	39
Şekil 3.6.	Open CV Kullanım yöntemi şeması	40
Şekil 3.7.	MTCNN Üç aşamalı mimarisi	41
Şekil 3.8.	MTCNN Yüz bölgesi belirleme	42
Şekil 3.9.	MTCNN ile yüz tespiti süreci	44
Şekil 3.10.	MTCNN ile kalabalık ortam yüz tespiti test çalışması	45
Şekil 3.11.	MTCNN ile kalabalık ortam yüz tespiti test çalışması-2	45
Şekil 3.12.	MTCNN ile kalabalık ortam yüz tespiti test çalışması-3	46
Şekil 3.13.	Modellerin genel eğitim aşaması	46
Şekil 3.14.	EfficientNet-B7 Katman mimarisi	47
Şekil 3.15.	EfficientNet-B7 FER-2013 sonuç grafiği	48
Şekil 3.16.	ConvLSTM Katman mimarisi	49
Şekil 3.17.	ConvLSTM FER-2013 sonuç grafiği	50
Şekil 3.18.	CNN Katman mimarisi	51
Şekil 3.19.	CNN Modeli FER-2013 sonuç grafiği	52
Şekil 3.20.	Ortam duygu yoğunluğu analizi işlem hattı	55
Şekil 3.21.	MetropolCard toplantısı tüm model test görüntüsü	56
Şekil 3.22.	Halı saha röportajı tüm model test görüntüsü	57
Şekil 3.23.	Metropolcard iş toplantısı CNN tüm duygular için ortam tahmini	57
Şekil 3.24.	Metropolcard toplantısının ortam duygu analizi grafiği	58
Şekil 3.25.	Tüm yüzler ve ortalama yoğunluk analizi akış diyagramı	60
Şekil 3.26.	Tüm yüzler ve tahminler için ortalama yoğunluk yöntemi grafiği	60
Şekil 3.27.	Duygu değişim süreci ile ortam duygu analizi akış diyagramı	62
Şekil 3.28.	Duygu değişim süreci ile ortam duygu analizi yöntemi grafiği	62
Şekil 3.29.	Eşik değeri belirlenmesi ile ortam duygu analizi akış diyagramı	64
Şekil 3.30.	Eşik değeri belirlenmesi ile ortam duygu analizi yöntemi grafiği	64
Şekil 3.31.	Anlık görüntülerden ortam duygu analizi akış diyagramı	66
Şekil 3.32.	Anlık görüntü alınarak ortalama yoğunluk analizi yöntemi grafiği	66

Şekil 3.33.	Yüz lokasyonundaki deęişiklerin takibi ile ortam duygu analizi akış diyagramı	67
Şekil 3.34.	Yüz bölgesindeki deęişikliklerin takibi ile ortalama yoğunluk analizi yöntemi grafięi	68
Şekil 4.1.	Toplantı sonrası oluşan duygu grafięi	76



TABLÖLER DİZİNİ

Tablo 2.1.	HOG, LBP ve GFC yüz tanıma yöntemlerinin karşılaştırması	9
Tablo 2.2.	2DPCA ve şablon eşleştirme ile önden tek yüz algılama sonuçları	10
Tablo 2.3.	2DPCA ve şablon eşleştirme ile önden çoklu yüz algılama sonuçları.....	10
Tablo 2.4.	Yüz algılama dedektörlerinin karşılaştırılması	12
Tablo 2.5.	RFE-MTCNN'nin Wider Face veri seti ile tespit performansı karşılaştırması	15
Tablo 2.6.	Yüz ifadesi tanıma için kullanılan ana kodlar	18
Tablo 2.7.	Yüz ifadesi tanıma için kullanılan yardımcı kodlar.....	18
Tablo 2.8.	Yüz ifadesi tanıma için kullanılan yardımcı kodlar-2	18
Tablo 2.9.	Yedi temel duygunun "Action Unit" kodlarına göre açıklaması	19
Tablo 2.10.	Duygu Sınıflandırması.....	20
Tablo 3.1.	EfficientNet-B7 FER-2013 veri seti ile sınıflandırma sonuçları	48
Tablo 3.2.	ConvLSTM FER-2013 veri seti ile sınıflandırma sonuçları	50
Tablo 3.3.	CNN FER-2013 veri seti ile sınıflandırma sonuçları	52
Tablo 3.4.	Farklı derin öğrenme modellerinin performans ve hesaplama karmaşıklığı karşılaştırması	54
Tablo 4.1.	Yöntemlerin performans karşılaştırmaları.....	75

SİMGELER VE KISALTMALAR DİZİNİ

A	: Uyarılma
D	: Dominans
i	: Zaman dilimini temsil eden indeks değeri
K	: Her zaman dilimindeki kare sayısı
T	: Tolerans katsayısı
V	: Değerlik

Kısaltmalar

AFEW	: Audio-Visual Emotion Challenge (Görsel-İşitsel Duygu Mücadelesi)
AU	: Action Unit (Eylem Birimi)
BCE	: Binary Cross-Entropy (İkili Çapraz Entropi)
CK+	: Extended Cohn-Kanade Dataset (Genişletilmiş Cohn-Kanade Veri Seti)
CNN	: Convolutional Neural Network (Evrşimsel Sinir Ağı)
DISFA	: Dynamic Static Facial Expressions (Dinamik Statik Yüz İfadeleri)
DOG	: Difference of Gaussian (Gauss Farkı)
FACS	: Facial Action Coding System (Yüz Hareketi Kodlama Sistemi)
FER	: Facial Expression Recognition (Yüz İfadesi Tanıma)
FLOP	: Floating Point Operations per Second (Saniye Başına Kayan Nokta)
GPU	: Graphics Processing Unit (Grafik İşleme Birimi)
JAFFE	: Japanese Female Facial Expression (Japon Kadın Yüz İfadesi)
KNN	: K-Nearest Neighbors (K-En Yakın Komşu algoritması)
LBP	: Local Binary Pattern (Yerel İkili Desen)
LSTM	: Long Short-Term Memory (Uzun Kısa Süreli Bellek)
mAP	: Mean Average Precision (Ortalama Doğruluk)
MMI	: Multi-Modal Interaction Database (Çok Modlu Etkileşim Veri Tabanı)
NMS	: Non-Maximum Suppression (Maksimum Olmayan Bastırma)
O-Net	: Output Network (Çıktı Ağı)
OpenCL	: Open Computing Language (Açık Hesaplama Dili)
P-Net	: Proposal Network (Öneri Ağı)
PNG	: Portable Network Graphics (Taşınabilir Ağ Grafikleri)
R-Net	: Refinement Network (İyileştirme Ağı)
RMSE	: Root Mean Squared Error (Kök Ortalama Kare Hatası)
SSD	: Single Shot Multibox Detector (Tek Atışlı Çoklu Kutu Tespit Cihazı)
SVM	: Support Vector Machine (Destek Vektör Makinesi)
VAD	: Valence-Arousal-Dominance (Değerlik-Uyarılma-Dominans)
YOLO	: You Only Look Once (Sadece Bir Kez Bak)

GÖRÜNTÜ İŞLEME VE YAPAY ZEKÂ İLE ORTAM DUYGU YOĞUNLUK ANALİZİ

ÖZET

Günümüzde görüntü işleme ve yapay zekâ alanında hızla gelişen teknikler, özellikle insan davranışlarının ve duygularının analiz edilmesinde önemli katkılar sağlamaktadır. İş toplantıları, eğitim, sağlık ve sosyal etkileşim gibi çeşitli alanlarda duygusal durumun izlenmesi ve analiz edilmesi, bireylerin ihtiyaçlarının anlaşılmasına, süreçlerin iyileştirilmesine ve karar alma mekanizmalarının güçlendirilmesine katkı sağlamaktadır. Duygu yoğunluğu analizi, toplu ortamlarda belirli bir duygunun baskın hale geldiği anların belirlenmesi ve bu anların değerlendirilmesi için önemli bir araç olarak ortaya çıkmaktadır. Bu bağlamda, görüntü işleme ve yapay zekâ teknikleri kullanarak duyguların doğru bir şekilde analiz edilmesi; seminerler, iş toplantıları, eğitim ve doktor-danışan görüşmeleri gibi alanlarda katılımcıların duygu durumlarının daha iyi anlaşılmasına imkân tanımaktadır.

Duygu tanımlama modellerinin eğitimi için FER-2013 ve CK+48 gibi geniş veri setleri kullanılarak, çeşitli derin öğrenme modelleriyle analiz gerçekleştirilmiştir. EfficientNet-B7, ConvLSTM ve Derin CNN gibi karmaşık modeller kullanarak duygu tespitinde farklı doğruluk oranları elde edilmiştir. Bu analizler doğrultusunda en iyi performans gösteren model seçilerek MTCNN yüz dedektörü ve OpenCV kütüphaneleri ile gerçek zamanlı ortam duygu yoğunluğu analizi gerçekleştirilmiştir.

Çalışmanın bir diğer katkısı, ortam duygu yoğunluğu analizinde kullanılan farklı hesaplama yöntemleridir. İlk olarak, saniye başına düşen karelerdeki tüm yüzlerdeki duyguların tanımlanması ve bu duyguların ortalamasının alınması yöntemi kullanılmıştır. İkinci olarak, ortamda belirgin duygu değişimlerinin analiz edilmesi amacıyla ortalama sürenin hesaplanması ele alınmıştır. Üçüncü yöntem, belirli bir eşik değerinin üzerindeki tahmin oranlarının hesaba katılmasına yoğunlaşmıştır. Dördüncü yöntem ise anlık görüntüler üzerinden belirli bir zaman aralığında yakalamalar yaparak hesaplamalarımız gerçekleştirilmesidir.

Anahtar Kelimeler: Duygu Sınıflandırması, Duygusal Yoğunluk, Gerçek Zamanlı Takip, Görüntü İşleme, Yüz Tespiti.

EMOTION INTENSITY ANALYSIS IN ENVIRONMENTS USING IMAGE PROCESSING AND ARTIFICIAL INTELLIGENCE

ABSTRACT

The rapidly advancing techniques in image processing and artificial intelligence today make significant contributions, especially in analyzing human behavior and emotions. Monitoring and analyzing emotional states across various fields such as business meetings, education, healthcare, and social interactions help understand individual needs, improve processes, and strengthen decision-making mechanisms. Emotion intensity analysis emerges as an essential tool for identifying and evaluating moments when a particular emotion becomes dominant in a crowd. In this context, accurately analyzing emotions using image processing and artificial intelligence techniques allows for a deeper understanding of participants' emotional states in areas such as seminars, business meetings, education, and doctor-client consultations.

By using extensive datasets like FER-2013 and CK+48, various deep learning models have been applied to analyze emotions. Different accuracy rates have been achieved in emotion detection using complex models such as EfficientNet-B7, ConvLSTM and Deep CNN. Based on these analyses, the best-performing model was selected to perform real-time environment emotion intensity analysis using the MTCNN face detector and OpenCV libraries.

Another contribution of this study is the different calculation methods used in environmental emotion intensity analysis. The first method involves identifying emotions in all faces in each frame per second and averaging these emotions. The second method focuses on calculating the average duration to analyze significant emotional changes in the environment. The third method emphasizes the consideration of prediction rates above a certain threshold. The fourth method captures snapshots over a specific time interval using real-time images to perform calculations.

Keywords: Emotion Classification, Emotional Intensity, Real-Time Tracking, Image Processing, Face Detection.

1. GİRİŞ

Görüntü işleme ve duygu tanıma teknolojilerinin gelişimi, bilgisayarla görme alanında önemli bir yer tutmaktadır. Özellikle yüz ifadelerinin analizine dayalı duygu tespiti, insan-makine etkileşiminin iyileştirilmesinde ve çeşitli sosyal uygulamalarda etkili bir araç haline gelmiştir. Bu çalışma, yüz tespiti ve duygu analizi yöntemlerini kullanarak, bir ortamda bulunan kişilerin duygu yoğunluklarını gerçek zamanlı olarak belirleyip ortamın genel duygu durumunu analiz etmeyi amaçlamaktadır.

1.1. Duygu Yoğunluğu Analizi Problemi

Günümüzde duygu tespiti ve yoğunluk analizinin gerçek zamanlı ve doğru bir şekilde yapılabilmesi, çeşitli teknik zorlukları beraberinde getirmektedir. Özellikle ortamda bulunan tüm yüzlerin algılanabilmesi ve bu yüzler üzerinden duygu durumlarının doğru bir şekilde analiz edilebilmesi, kaynak tüketimi gibi iç etkenlerin, çözünürlük, ışıklandırma, açı gibi dış etkenlerin sebep olduğu problemlerden etkilenmektedir.

Görüntü işleme teknolojileri ve bu teknolojilerin çeşitli uygulama alanları uzun yıllara dayanan bir geçmişe sahiptir. Bu süreçte kullanılan yöntemler çeşitlenmiş, kullanım amaçlarına göre farklılaşmış ve önemli ilerlemeler kaydetmiştir. Görüntü işleme alanındaki yapay zekâ çalışmalarının önemli bir bölümü görüntüler üzerinden tespit ve analize odaklanmıştır. Duygu tespitinin temelinde nesne tespiti yöntemleri yer almakta olup, bu yöntemlerin üstünde yüz tespiti bulunmaktadır. Yüz tespiti, özellikle son yıllarda gelişen derin öğrenme yöntemleri sayesinde en çok araştırılan ve yüksek verim elde edilen alanlardan biri haline gelmiştir (Guo ve Zhang, 2018). Bu gelişmelere bağlı olarak, yüz tanımlama sürecinden elde edilen veriler üzerinde çeşitli çıkarımlar yapılarak bu çalışmaların kapsamı genişletilmiştir.

Görüntüler üzerinden yüz ve duygu tespiti gerçek zamanlı olarak gerçekleştirildiğinde, verimlilik ve doğruluk açısından istenilen sonuçlara ulaşmak genellikle zordur. Bu tür çalışmalarda, görüntü kalitesi ve yüzün açısı gibi girdiler ortak zorluklar olarak öne çıkmakta ve nihai sonuçları önemli ölçüde etkilemektedir. Özellikle, yoğun ortamlarda duygu analizinin doğru yapılabilmesi için yüzlerin güvenilir şekilde tespit edilmesi,

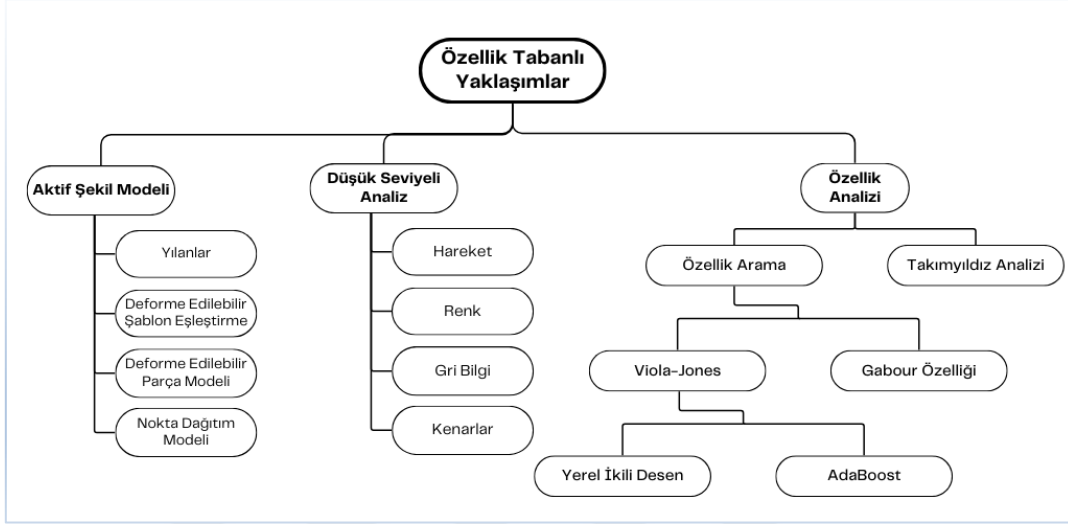
alışmanın verimlilięi ve doęruluęu aısından kritik bir parametre olarak karřımıza ıkmaktadır.

Yüz tespiti alanındaki ilk alışmalardan bazıları, günümüzde daha ilkel yöntemler olarak deęerlendirilmekle birlikte, bu alandaki gelişimin temelini oluşturmıştır. Başlangı aşamalarında, önerilere dayalı olarak geliştirilen bu yöntemler, yüz tespiti üzerine yapılan arařtırmalarla ilerleyerek bugünkü modern yaklaşımlara yol amıştır. Erken dönem alışmalardan biri, "Low-dimensional procedure for the characterization of human faces" başlıklı alışmadır. Bu alışmada, yüz tespiti piksellere dayalı olarak, her pikseli içeren temsili bir matris aracılığıyla dijital hale getirilmiştir. "Eigenpictures" olarak adlandırılan özdeęer görüntüleri, yüzlerin varyans matrisinin özvektörlerinden türetilmiştir. Bu görüntüler, yüzlerin en uygun koordinat sistemini sağlayarak veri sıkıştırma işlevi görmektedir. Ayrıca, gri tonlama kullanılarak temsil edilen bu piksellerin, Fourier dönüşümü gibi alternatif yöntemlerle dijital olarak ifade edilebileceęi belirtilmiştir (Sirovich ve Kirby, 1987).

Yüz tespiti, insan yüzlerinin görüntülerde bulunması işlemi olup, yüz doęrulama, modelleme, yař ve cinsiyet tahmini gibi daha gelişmiş teknolojilere temel oluşturan önemli bir adımdır. İnsanların kolayca yapabildięi bu görev, makineler için karmařık ve yoğun hesaplama gücü gerektirir. Literatürde yüz tespiti üzerine eřitli alışmalar yapılmış olsa bile, kullanılan yöntemlerin kapsamlı bir incelemesinin eksik olduęu vurgulanmaktadır. Yüz tespiti algoritmalarını tarihesi, prensipleri, avantajları, sınırlamaları ve dięer alanlardaki kullanımları bağlamında beř bölümde incelemekte ve yöntemleri sistematik olarak sınıflandırmaktadır (Hasan ve dię., 2021). Özellik tabanlı ve görüntü tabanlı yaklaşımlar olarak ikiye ayrılan yöntemler, ışık, engel ve karmařık arka plan gibi faktörler aısından deęerlendirilmektedir. Şekil 1.1'de özellik tabanlı yaklaşımlar şeması görölmektedir.

Ortam yoğunluęundaki duygu analizi söz konusu olduęunda, mevcut ortamda bulunan tüm bireylerin duygu durumlarının kapsamlı bir şekilde hesaplanması gerekmektedir. Bu hesaplamayı mümkün kılmanın temel şartı ise ortamda bulunan her yüzün güvenilir bir şekilde tespit edilmesidir. Bu alanda yeni bir dönemin kapısını aralayan en önemli alışma, Paul Viola ve Michael Jones'un 2001 yılında yayımladıkları "Rapid Object Detection using a Boosted Cascade of Simple Features" başlıklı alışmadır. Viola ve

Jones, yüz tespiti algoritmalarında çığır açan bir yaklaşım olan AdaBoost algoritmasını kullanarak basit özelliklerin güçlendirilmiş bir kaskad sistemi ile nesne tespiti gerçekleştirmiştir (Viola ve Jones, 2001).



Şekil 1.1. Yüz algılama için özellik tabanlı yaklaşımlar (Hasan ve diğ., 2021)

Bu çalışmanın sunduğu yöntem, o zamana kadar kullanılan diğer yüz tespiti yöntemlerine göre çok daha hızlı ve verimli bir alternatif sunmuş ve yüz tespitinde yeni bir çağ başlatmıştır. AdaBoost algoritması ile güçlendirilen bu yaklaşım, temel özelliklerin hızlı bir şekilde işlenmesiyle yüzleri etkili bir şekilde tespit edebilmiş ve görüntü işleme alanında standart bir yöntem haline gelmiştir. Günümüzde kullanılan birçok modern yüz ve duygu tanıma teknolojisi, bu çalışmada önerilen yöntemlerden ve elde edilen bulgulardan ilham alarak geliştirilmiştir. Viola-Jones algoritması, yüz tespitinde öncü ve devrim niteliğinde bir ilerleme sağlamış ve alanın ilerlemesine önemli katkılarda bulunmuştur.

Yüz ifadeleri, bireylerin içinde bulundukları duygusal durumları ve aktarmak istedikleri mesajları iletmekte en güçlü, doğal ve yaygın sinyallerden birini oluşturmaktadır. Günümüz teknolojisi, görüntülerdeki bu ifadelerin ve temel duyguların tespit edilmesine olanak sağlamaktadır. Özellikle son yirmi yılda, görüntü işleme yoluyla duygu tespiti çalışmaları çeşitli yöntemlerle dikkat çekici bir araştırma alanı haline gelmiştir (Jain ve diğ., 2019). Yüz ifadelerinin sınıflandırılması ve analiz edilmesine yönelik çalışmalarda, çocuklar ve yaşlılar gibi belirli grupların duygusal durumlarının değerlendirilmesine dair çalışmaların da geliştiği gözlemlenmektedir. Bununla birlikte, belirli yaş gruplarına veya

özel gruplara odaklanan çalışmaların sayısı sınırlı kalmaktadır. Farklı alanlardaki bu çalışmalar arasında sonuçların tutarsız ve değişken olması ise genellikle kullanılan yöntemlerin ve veri çeşitliliğinin yarattığı farklılıklara dayanmaktadır (Durand ve diğ., 2007).

Geleneksel yüz tespiti yöntemleri, yüz tanımlama, duygu analizi, yaş ve cinsiyet tahmini gibi üst düzey bilgisayarla görme uygulamalarında belirli bir başarı elde etse de günümüzde otonom sürüş ve robot teknolojileri gibi karmaşık alanlarda yetersiz kalmaktadır. Bu yetersizliklerin temel sebepleri arasında, görüntü kalitesindeki düşüklük, açısal sapmalar ve yetersiz ışıklandırma gibi faktörler yer almaktadır. Bu çalışmada, duygu tespiti sürecinde karşılaşılan bu tür zorlukları aşabilmek için derin öğrenme tabanlı yaklaşımlar tercih edilmiştir ve yaygın olarak kullanılan derin sinir ağı yöntemleri ele alınacaktır.

Duygu analizi, yüz tespiti temelinde şekillenen ve bilgisayarla görme alanında gelişen bir araştırma konusudur. Bu bağlamda, duygu tespitinin önemi, amacı ve potansiyel uygulama alanları da değerlendirilmiştir.

Duygu çıkarımı, yüz ifadelerindeki mikro hareketlerin ve kas aktivasyonlarının yanı sıra, ses tonu ve metin analizi gibi çok sayıda farklı girdiyle gerçekleştirilebilmektedir. Çalışmamız, yüz tespiti ve bu tespitten elde edilen duygu analizine odaklanmaktadır. Duygu tespiti için yüz ifadelerinin hangi duygu durumunu temsil ettiğini anlamak önemlidir. Bu konuda öncü çalışmalardan biri olan Paul Ekman'ın çalışması, insan yüzündeki ifadelerin evrensel olduğunu ve yedi temel duyguyu (mutluluk, üzüntü, öfke, korku, şaşkınlık, tiksinti ve küçümseme) içerdiğini ortaya koymuştur (Ekman ve Friesen, 1978). Ekman'ın geliştirdiği Yüz Hareketi Kodlama Sistemi (FACS), İsveçli anatomist Carl-Herman Hjortsjö'nün çalışmalarına dayanmaktadır (Hjortsjö, 1970). Ekman, bu çalışmasını 2002 yılında yayımladığı kitabında büyük ölçüde güncellemiştir (Ekman ve diğ., 2002).

Yüz tespiti için geleneksel yöntemlerden Haar Cascade gibi algoritmalar ya da modern derin öğrenme tabanlı yöntemler kullanılabilmektedir. Yüz ifadelerinden duygu çıkarımı (FER) için derin sinir ağları ve özellikle evrişimli sinir ağları (CNN) kullanımı son yıllarda yaygınlaşmıştır. İlk dönemde yüz tanıma sistemlerinde destek vektör makineleri

(SVM), Bayes sınıflayıcılar, bulanık mantık teknikleri, öznitelik seçimi ve yapay sinir ağları (ANN) gibi geleneksel yöntemler tercih edilmiştir. Ancak, düşük çözünürlük, poz değişikliği, ölçek ve ışıklandırma farklılıkları gibi kısıtlamalar bu yöntemlerin doğruluğunu sınırlamıştır. CNN'in yetenekleri, FER doğruluğunu önemli ölçüde artırmış ve FER için en etkin yöntemlerden biri olarak öne çıkmıştır (Ezerceli ve Eskil, 2022).

Duygu analizi modellerinin başarısında kullanılan veri setlerinin kalitesi büyük önem taşır. FER2013, CK+, JAFFE ve FERG gibi veri setleri yüz tespiti ve duygu analizi çalışmalarında sıklıkla kullanılmaktadır. Bu veri setleri içinde FER2013 veri tabanı özellikle öne çıkar; 2013 yılında ICML Temsil Öğrenimindeki Zorluklar konferansında tanıtılan bu veri seti, yüz örtüşmeleri, düşük görüntü kalitesi, kontrast farklılıkları ve gözlük kullanımı gibi çeşitli zorlukları barındırarak daha kapsamlı ve çeşitlilik sunmaktadır (Ezerceli ve Eskil, 2022).

Bu bağlamda, çalışmamız derin sinir ağları yardımıyla yüz ifadelerinin daha doğru tespit edilmesini hedeflemekte ve zorlu koşullarda dahi etkin sonuçlar verebilecek bir duygu tespiti altyapısı sunmayı amaçlamaktadır.

1.2. Duygu Yoğunluğu Analizi Amacı

Bir ortamda tespit edilen yüzler üzerinden duygu yoğunluğu analizini gerçek zamanlı olarak gerçekleştirmektir. Çalışma kapsamında, yüz tespiti algoritmaları ve duygu analiz teknikleri kullanılarak, ortamdaki bireylerin duygu yoğunluklarının hesaplanması ve bu yoğunluğun genel ortam atmosferine etkisinin belirlenmesi hedeflenmektedir. Ayrıca, yüz tespiti sırasında karşılaşılan zorluklar analiz edilerek duygu yoğunluğu analizine olan etkileri değerlendirilecektir. Ek olarak Analiz ve hesaplama yöntemleri için farklı yöntemler test edilerek uygun analiz yöntemleri incelenecektir. Kaynak tüketimi ve performans değerlendirmeleri yapılarak öneriler verilecektir. Farklı ortamlarda duygu yoğunluğu analizi yaparak sistemimizin performansı ölçülecektir.

1.3. Duygu Yoğunluğu Analizi Önemi

Yüz ifadelerine dayalı duygu tespiti, günümüzde sosyal medya, güvenlik, eğitim ve sağlık gibi pek çok alanda kullanılmaktadır. Bu araştırma, duygu yoğunluklarının toplu analizinin, örneğin bir etkinlik ortamında katılımcıların duygu durumlarını anlamak veya

bir eğitim ortamında öğrenci memnuniyetini değerlendirmek gibi uygulamalara katkı sağlayabileceğini göstermektedir. Bu tür analizler, ortamlardaki duygu dinamiklerini daha iyi anlamamızı ve bu bilgilere dayalı olarak çeşitli iyileştirmeler yapmamızı mümkün kılmaktadır.

1.4. Araştırma Kapsamı

Bu çalışma, yüz tespiti ve duygu analizine dayalı olarak ortam duygu yoğunluğunu belirlemeyi hedefleyen derin öğrenme tabanlı bir sistem tasarlamaktadır. Çalışmada, temel olarak derin sinir ağları, evrişimli sinir ağları (CNN) ve veri setleri (FER-2013, CK+ gibi) kullanılacaktır. Ayrıca, geleneksel yöntemlerle derin öğrenme yöntemleri arasındaki farklar ve yüz tespiti sırasındaki çeşitli zorlukların duygu yoğunluğu analizi üzerindeki etkileri kapsam dahilinde incelenecektir.

1.5. Organizasyon

Bu tez yedi ana bölümden oluşmaktadır. Birinci bölüm, çalışmanın konusunun tanıtımını, problemin tanımını, araştırmanın amacını, önemini, kapsamını ve organizasyonunu içermektedir.

İkinci bölüm, yüz tespiti ve duygu tespiti konularında yapılan literatür taramasını, bu alandaki veri setlerini, kullanılan yöntemleri ve mevcut çalışmaların değerlendirilmesini kapsamaktadır.

Üçüncü bölüm, araştırmada kullanılan veri setleri, yüz tespiti, özellik çıkarma, duygu sınıflandırma ve ortam duygu yoğunluğu analiz yöntemleri hakkında detaylı bilgi sunmaktadır.

Dördüncü bölüm, model performans sonuçlarını ve ortam duygu yoğunluğu analizine dair bulguları detaylı bir şekilde açıklamaktadır.

Beşinci bölüm, elde edilen sonuçların değerlendirilmesini, araştırmanın sınırlılıklarını ve gelecekte yapılabilecek çalışmalar için önerileri içermektedir. Tez ayrıca altıncı bölümde genel sonuçlara ve uygulama önerilerine yer vermekte, yedinci bölümde ise kaynakları sunmaktadır. Ek olarak, tezde algoritmaların teknik detayları, yazarın kişisel yayın ve eserleri ile özgeçmişi de bulunmaktadır.

2. LİTERATÜR

Bu çalışma kapsamında, ortam duygu yoğunluğunun hesaplanması için yüz tespiti, tespit edilen yüz üzerinde duygu tanımlaması (Facial Expression Recognition - FER) ve tespit edilen duyguların ortamın yoğunluğunun değerlendirilmesinde kullanılması olmak üzere üç temel aşama ele alınmıştır. Literatür incelememiz, bu üç aşamanın her birine yönelik önceki çalışmaları kapsayacak şekilde yapılandırılmıştır.

İlk aşama olan yüz tespiti, görüntü işleme ve makine öğrenimi alanlarında gerçekleştirilen çalışmaların değerlendirilmesi ile en etkili yöntemlerin belirlenmesini hedeflemektedir. İkinci aşamada, duygu tanımlaması (FER) ile ilgili literatürdeki mevcut yaklaşımlar incelenerek, farklı algoritmaların doğruluk ve verimlilik açısından karşılaştırması yapılacaktır. Son olarak, güncel literatür çalışmaları doğrultusunda elde edilen yöntemlerin, önerilen sistemin kapsamı ve katkıları çerçevesinde analiz edilmesi sağlanacaktır.

Bu bağlamda, literatür incelemesi üç ana başlık altında detaylandırılacak olup, yüz tespiti ve duygu tanımlaması alanlarındaki gelişmeler ile güncel yöntemlerin incelenmesi suretiyle çalışmamızın metodolojik temelini oluşturulması amaçlanmıştır.

2.1. Yüz Tespiti

Yüz tespiti, bilgisayarlı görü alanında nesne tespitinin özel bir alt dalı olarak değerlendirilir. Nesne tespiti, bir görüntüdeki belirli nesnelerin yerlerini belirleme ve sınıflandırma sürecini ifade ederken, yüz tespiti bu süreci insan yüzlerine odaklanarak gerçekleştirir. Teknik olarak, yüz tespiti algoritmaları, nesne tespiti için geliştirilen temel prensiplere dayanır ve bunları yüzlerin karakteristik özelliklerini tanımlamak için uyarlamaktadır.

Güncel yüz tespiti yöntemleri, nesne tespiti yaklaşımlarından türetilen iki temel yöntemi içerir: özellik tabanlı ve görüntü tabanlı yöntemler. Özellik tabanlı yöntemler, nesne tespiti için kullanılan kenar, köşe veya doku gibi temel özellik çıkarımı tekniklerini insan yüzüne özgü özelliklerle zenginleştirir. Örneğin, yüz tespitinde gözler, burun ve ağız gibi belirgin yüz yapıları, özellik tabanlı tespitin temel bileşenlerini oluşturur.

Görüntü tabanlı yöntemler ise derin öğrenme tabanlı nesne tespit tekniklerinden türetilmiştir ve genellikle geniş veri kümeleri üzerinde eğitilen Convolutional Neural Networks (CNN) modellerine dayanır. Bu yöntemler, nesne tespiti için kullanılan genel modelleri yüzlerin karmaşıklığına adapte ederek, farklı pozlar, ışık koşulları ve yüz ifadeleri altında yüksek performans sergiler. Örneğin, Faster R-CNN ve YOLO gibi nesne tespit modelleri, yüz tespiti için özelleştirilerek geniş bir uygulama yelpazesinde kullanılmaktadır.

2.1.1. Geleneksel Yöntemler

Yüz tespiti için en önde gelen çalışmalardan biri Viola-Jones algoritmasıdır. Geleneksel yöntem olarak en çok kullanılan yöntemlerin başında gelmektedir. Viola-Jones Algoritması, 2001 yılında tanıtıldığında düşük çözünürlüklü görüntülerde gerçek zamanlı algılama ihtiyaçlarını karşılayabilecek şekilde tasarlanmıştır. Ancak, günümüzün yüksek çözünürlükteki görüntü ve video verileri karşısında bu algoritma performans açısından yetersiz kalmaktadır. Bu soruna yönelik bir çözüm olarak, Weiyan Wang ve arkadaşları algoritmanın temel yapısını değiştirmek yerine, genel amaçlı GPU'lar (GPGPU) ve OpenCL standardı gibi modern donanım ve yazılım teknolojilerinden faydalananak performans optimizasyonuna odaklanmıştır.

Bu bağlamda, algoritmanın GPU platformlarında daha verimli çalışmasını sağlamak için çeşitli teknikler uygulanmıştır. Örneğin, bellek erişimini hızlandırmak amacıyla yerel bellek ve manuel ön bellekleme mekanizmaları kullanılmış, aynı zamanda paralel iş yükü yönetimiyle hesaplama kaynaklarının verimli bir şekilde kullanımı sağlanmıştır. Bu optimizasyonlar sayesinde hem AMD hem de Nvidia platformlarında, CPU tabanlı uygulamalara kıyasla 13 kata kadar hız artışı elde edilmiştir. Bununla birlikte, bu çalışma yalnızca Viola-Jones algoritmasının performansını iyileştirmekle kalmayıp, aynı zamanda GPU platformlarına yönelik genel optimizasyon prensiplerine de ışık tutmaktadır (Wang ve diğ., 2012).

Bir diğer çalışmada yüz tanıma için yönlendirilmiş gradyanların yoğun ızgara histogramlarına (HOG) dayalı bir yöntem önermektedir. Yöntemde, yüz görüntüsü yoğun ızgaralara bölünerek her bir ızgaradan HOG özellikleri çıkarılmakta ve bu özellikler

birleştirilerek yüzün genel temsili oluşturulmaktadır. Daha sonra, yüz tanıma işlemi en yakın komşu sınıflandırıcı kullanılarak gerçekleştirilmektedir.

Çalışmada, FERET yüz veri tabanı kullanılarak aydınlatma, zaman ve ortam değişiklikleri altında HOG parametreleri optimize edilmiştir. Gama aydınlatma düzeltmesi, uzaysal gradyan yönü, blok boyutu ve çözünürlük gibi faktörler analiz edilmiştir.

Önerilen yöntem, Gabor dalgacığı ve yerel ikili model (LBP) yöntemleriyle karşılaştırılmıştır. Deneysel sonuçlar, yoğun ızgaralı HOG yönteminin zaman ve ortam değişikliklerine daha uygun olduğunu ve LBP'den daha az boyutla daha iyi bir tanıma oranı sağladığını göstermektedir. Ayrıca, yoğun ızgara HOG özelliğı çıkarma süresi, Gabor dalgacığı yöntemine kıyasla çok daha hızlıdır. HOG ve LBP'nin işlem süreleri ise benzerdir. Şekil 2.1'de yapılan çalışmanın diğer yöntemler ile karşılaştırılması gösterilmektedir.

Tablo 2.1. HOG, LBP ve GFC yüz tanıma yöntemlerinin karşılaştırması

Method	Çözünürlük	Özellik Boyutu	Fb	Fc	Dup1	Dup2
LBP	80×64	4720	95,5	40,7	61,6	47,4
LBP	128×128	3304	93,0	51,0	61,0	50,0
LBP	160×128	18880	95,3	65,5	65,9	57,7
HOG	80×64	2560	92,6	72,2	66,3	61,1
GFC	64×64	250	96,3	81,4	68,8	49,1

Şekil 2.1'de Fb; ilk test seti için tanıma başarıım yüzdesini, Fc; aydınlatma ve kamera değişikliklerine karşı dayanıklılığı ölçen test seti başarıımını, Dup1; zaman içinde değişen veriler için test edilen başarı yüzdesini, Dup2; daha büyük zaman aralığında ve farklı çevresel koşullarda test edilen başarı yüzdesini temsil etmektedir. HOG yöntemi, zaman içindeki değişikliklere en dayanıklı olanı. Özellikle dup2 testinde %12 daha iyi performans göstermesi, HOG'un çevresel değişikliklere ve zaman faktörüne daha uyumlu olduğunu kanıtlıyor.

Sonuç olarak, yoğun ızgaralı HOG yöntemi, hızlı ve dayanıklı bir özellik çıkarma yöntemi olarak yüz tanıma alanında etkili bir alternatif sunmaktadır (Xiang ve diğ., 2018).

Jizeng Wang ve arkadaşları yaptıkları çalışmada, yüz tanıma teknolojileriyle ilgili bir yüz tespiti yöntemini ele almakta ve önerilen bir hiyerarşik yüz tespit algoritmasını tanıtmaktadır. Çalışmanın amacı, özellikle karmaşık arka planlara sahip görüntülerde yüzlerin hızlı ve doğru bir şekilde tespit edilmesidir. Önerilen yöntem, şablon eşleştirme ve 2DPCA (İki Boyutlu Ana Bileşen Analizi) algoritmalarını birleştirerek önden görünüm yüzlerin tespit edilmesini sağlamaktadır (Yang ve diğ., 2021). Testler, ORL veri tabanından alınan 30 yüz, normal kimlik fotoğraflarından 50 yüz, günlük hayattan 20 fotoğraf ve internetten rastgele seçilen çoklu yüz içeren 20 görüntü ile gerçekleştirilmiştir. Ayrıca MIT-CMU veri seti de çoklu yüz testleri için kullanılmıştır. İki göz şablonu ve yüz şablonları oluşturulmuş, bu şablonlar yardımıyla tespit pencereleri sıralanmıştır. İlk aşamada, iki göz şablonuna göre yüz adayları belirlenmiş, ardından yüz şablonları ile daha fazla eleme yapılmıştır. Şablon eşleştirme sonrası kalan görüntüler, 2DPCA algoritması ile işlenmiştir. Bu algoritma, doğrudan görüntü matrisleri üzerinde çalışarak kovaryans matrisinden özdeğer ve özvektörleri hesaplamış, böylece tespit hızını artırmıştır.

Tablo 2.2. 2DPCA ve şablon eşleştirme ile önden tek yüz algılama sonuçları

Tespit Yöntemi	Tespit Sayısı	Doğru Sayısı	Hatalı Sayısı	Alev Algılama Sayısı	Tespit Oranı (%)
2DPCA	100	92	8	11	92
ŞE & 2DPCA	100	92	8	5	92

Deney sonuçlarına göre, şablon eşleştirme ve 2DPCA yöntemlerinin birleştirilmesi, yalnızca 2DPCA kullanan yönteme kıyasla hem tespit oranını artırmış hem de yanlış tespit sayısını azaltmıştır. Tablo 2.2’de tekli yüz testlerinde %92 tespit oranı sağlanırken, Tablo 2.3’de çoklu yüz testlerinde tespit oranı %85,96’ya ulaşmıştır. Ayrıca, hiyerarşik yapı sayesinde işlem hızında da belirgin bir artış sağlanmıştır (Wang ve Yang, 2008).

Tablo 2.3. 2DPCA ve şablon eşleştirme ile önden çoklu yüz algılama sonuçları

Tespit Yöntemi	Tespit Sayısı	Doğru Sayısı	Hatalı Sayısı	Alev Algılama Sayısı	Tespit Oranı (%)
2DPCA	983	812	171	212	82,60
ŞE& 2DPCA	983	845	138	103	85,96

2.1.2. Modern Yöntemleri

Yüz tespiti alanında derin öğrenme tabanlı yaklaşımlar, son yıllarda geleneksel yöntemlerin yerini almıştır. Bu bağlamda, YOLO mimarisi üzerine inşa edilen YFaces sistemi, gerçek zamanlı yüz tespiti için önemli bir gelişmeyi temsil etmektedir (Chan ve diğ., 2024).

Geleneksel yüz tespit yöntemleri, HAAR özellikleri (Viola ve Jones, 2001) ve HOG (Dalal ve Triggs, 2005) gibi tasarlanmış algoritmalar temel alınarak geliştirilmiştir. Ancak, derin öğrenmenin hızlı ilerleyişi ile birlikte, özellikle tek aşamalı tespit mekanizmaları olan SSD (Souaidi ve Ansari, 2022) ve YOLO (Huang ve Kang, 2024) gibi modeller, gerçek zamanlı yüz tespiti için önemli avantajlar sunmuştur.

YOLO mimarisinin yüz tespiti bağlamındaki kullanımı, YOLOv3-face modelinin ortaya çıkışıyla başlamış ve bu alanda ciddi ilerlemeler kaydedilmiştir. Daha sonraki modellerden YOLO-face, darknet-53 omurgasını optimize ederek küçük boyutlu yüzlerin tespitinde performansı artırmış ve özel kayıp fonksiyonlarının entegrasyonu ile daha başarılı sonuçlar elde etmiştir. YOLO-FaceV2 ise, ölçekler arası farklılıklar ve yüzlerin örtüşmesi gibi sorunları ele alarak, ölçek farkındalığını artıran Scale-Aware RFE modülü ve Repulsion Loss (itme kaybı) gibi yenilikçi yöntemler sunarak yüz tespitinde daha yüksek doğruluk sağlamıştır.

Aifian Adi Sufian Chan ve arkadaşlarının YFaces mimarisi ve Wider Face (Yang ve diğ., 2016) data setini kullanarak başarılı sonuçlar aldılar. Easy subset: >94% mAP, Medium subset: 92-95% mAP, ve Hard subset: 83-88% mAP bu sonuçlar, özellikle zorlu koşullarda YFaces'in önceki YOLO tabanlı sistemlere göre daha başarılı olduğunu göstermektedir (Chan ve diğ., 2024).

Xin Li ve çalışma arkadaşları tarafından 2021 yılında gerçekleştirilen bir çalışmada, yüz algılama alanında önerilen modellerin performansı birden fazla veri seti üzerinde test edilmiştir. Bu çalışmada, çift dallı bir merkez yüz algılama modeli olan, basit ancak etkili bir saf evrişimli sinir ağı (P-CNN) önerilmiştir. Önerilen model, AFW (Zhu ve Ramanan, 2012), PASCAL FACE (Yan ve diğ., 2014), FDDB (Jain ve Learned-Miller, 2010) ve WIDER FACE (Yang ve diğ., 2016) veri setleri üzerinde değerlendirilmiştir.

Modelin performansını artırmak ve işlem hızını optimize etmek amacıyla, Özellik Piramidi Toplama (Feature Pyramid Aggregation, FPA) modülü ve Ölçeğe Uyarlanabilir Gauss Maskesi (Scale-Adaptive Gaussian Mask, SAGM) önerilmiştir. Bu yaklaşım, hızlı ve verimli bir yüz algılama dedektörü geliştirmeyi hedeflemiştir. Çalışmada, önerilen modelin performansı, daha önce yapılan çalışmalarla karşılaştırılarak tüm veri setleri üzerinde test edilmiştir.

Özellikle WIDER FACE veri setinde, kolay, orta ve zor kategorilerde yaklaşık %2 ila %6 arasında bir performans artışı sağlandığı rapor edilmiştir. Ayrıca, geliştirilen modelin, çok küçük yüzler ve karmaşık yapılar gibi zorlu durumlarda da yüksek bir başarı sergilediği belirtilmiştir. İşlem süresi açısından yapılan karşılaştırmalarda, önerilen modelin SRN ve PyramidBox ile kıyaslandığında farklı boyutlardaki girişler üzerinde ortalama iki kat daha hızlı çalıştığı ifade edilmiştir (Li ve diğ., 2021).

Sumit Tariyal ve arkadaşları bilgisayarlı görü alanındaki devrimsel gelişmeler sonucunda yüz tespiti en önemli görevlerden biri haline geldiğini belirtmiştir. Yapılan çalışmada doğruluk ve verimlilik olarak gerçek zamanlı olarak dört dedektörü karşılaştırmak için çalışma yaptılar.

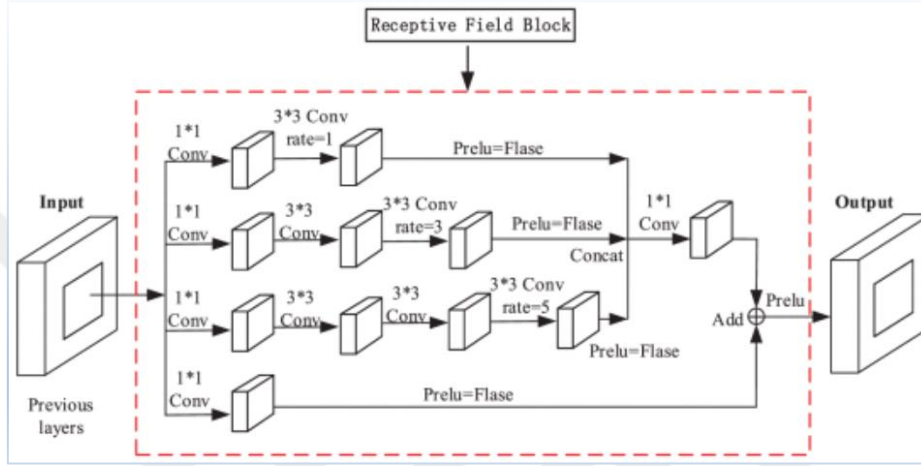
Bu dört algoritmanın YouTube görüntüleri üzerinde doğruluk ve hız performanslarının karşılaştırmalı değerlendirmesi yapılmıştır. Şekil 2.4'te bu değerlendirmenin doğruluk ve hız gibi sonuçları yer almaktadır.

Tablo 2.4. Yüz algılama dedektörlerinin karşılaştırılması (Tariyal ve diğ., 2024)

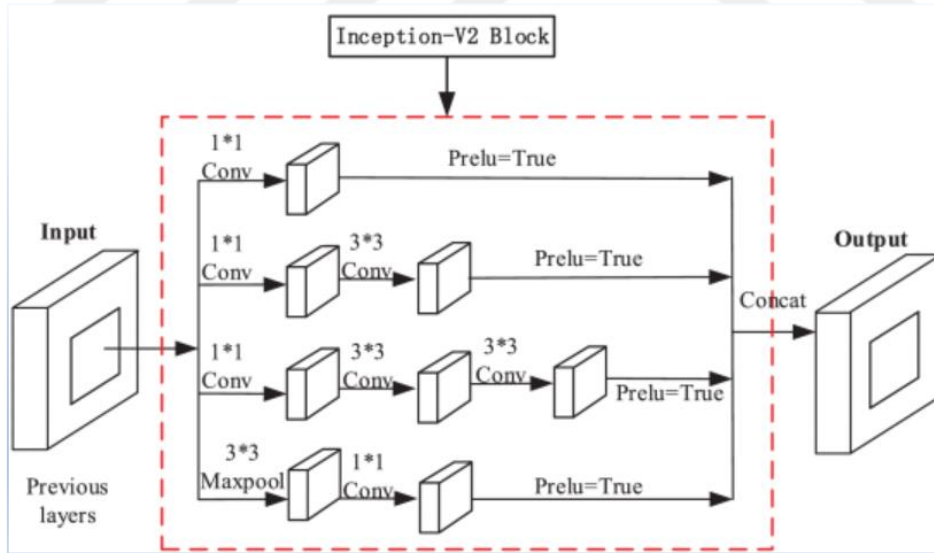
Algoritma	Kesinlik/Duyarlılık/F1 Skoru	Arayüz Zamanı
MTCNN	0.92 / 0.95 / 0.93	13.6 ms
SSD	0.89 / 0.91 / 0.92	12.4 ms
YOLO	0.80 / 0.82 / 0.85	5.3 ms
Viola-Jones	0.75 / 0.77 / 0.79	8.6 ms

Sonuç olarak, değerlendirilen algoritmaların güçlü ve zayıf yönleri ortaya konmuştur. Geleneksel yöntemlere yeni bir yaklaşım getiren Viola-Jones en hızlı, MTCNN ise en doğru performans göstermiştir. Algoritma seçimi, hedeflenen uygulama ihtiyaçlarına göre yapılması gerektiği belirtilmiştir (Tariyal ve diğ., 2024).

Bir başka çalışmada, çok aşamalı ve çok görevli bir yapı olan MTCNN'nin yüz algılama performansını iyileştirmek amacıyla RFE-MTCNN adı verilen yeni bir model tanıtılmıştır. Geleneksel MTCNN, özellikle küçük yüzlerin algılanmasında ölçek değişkenliği gibi zorluklarla karşılaşırken, RFE-MTCNN bu sorunları çözmek için Inception-V2 ve alıcı alan bloğu (Receptive Field Expansion) gibi ileri seviye özellikleri entegre etmiştir. Şekil 2.1. ve Şekil 2.2' de eklenen özelliklerin mimarisi verilmiştir.

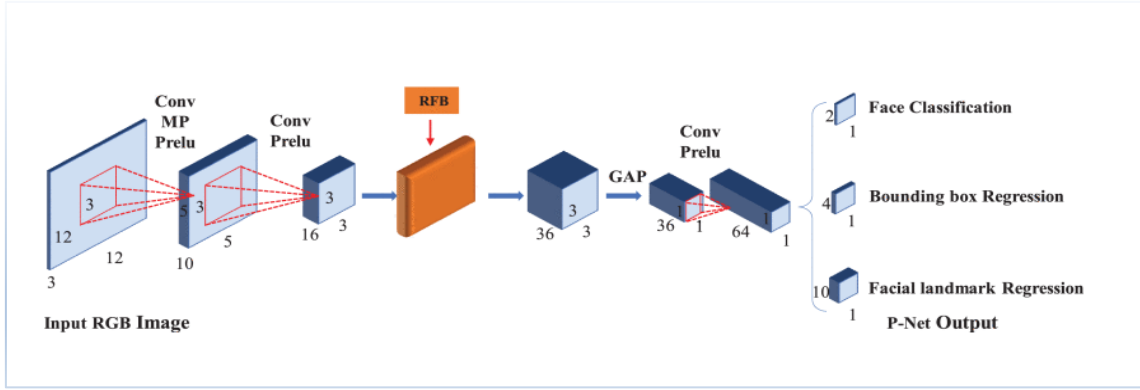


Şekil 2.1. RFB mimarisi (Li ve diğ., 2020)

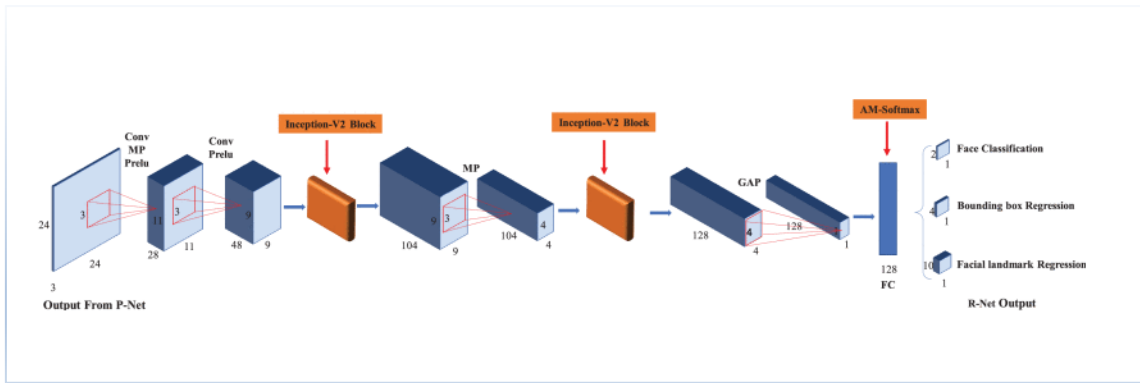


Şekil 2.2. Inception-V2 mimarisi (Li ve diğ., 2020)

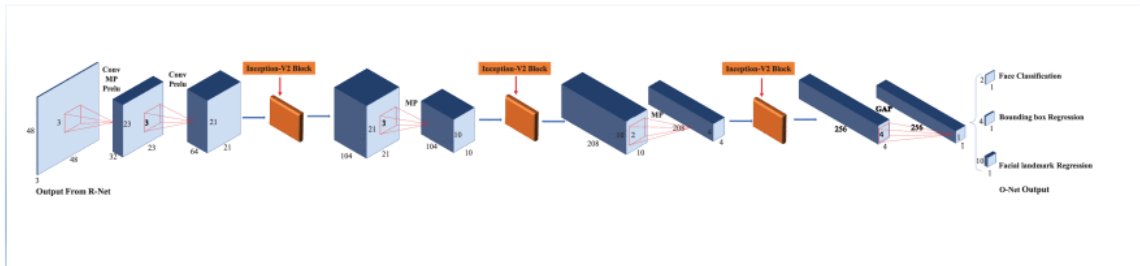
Şekil 2.3, Şekil 2.4 ve Şekil 2.5, RFE-MTCNN modelinin önerdiği üç alt ağı göstermektedir. Ayrıca, AM-Softmax kayıp fonksiyonu kullanılarak özellik temsili ve sınıflandırma performansının artırılması hedeflenmiştir.



Şekil 2.3. MTCNN P-Net mimarisi (Li ve diğ., 2020)



Şekil 2.4. MTCNN R-Net mimarisi (Li ve diğ., 2020)



Şekil 2.5. MTCNN O-Net mimarisi (Li ve diğ., 2020)

Eğitim sürecinde WIDER FACE ve CelebA (Liu ve diğ., 2015) veri setleri kullanılmış olup, toplamda 215.479 görüntü içeren bu veri setlerinde sırasıyla 12.880 ve 202.599 görüntü bulunmaktadır. Eğitim sırasında negatif, pozitif, parça yüzleri ve dönüm noktası yüzleri olmak üzere dört tür veri tanımı kullanılmıştır.

RFE-MTCNN'nin performansı, MTCNN ile karşılaştırıldığında yaklaşık %1,5'lik bir iyileşme göstermiştir. Ayrıca, önerilen modelin parametre boyutunu önemli ölçüde azaltmayı hedeflediği de belirtilmiştir. Tablo 2.5'de görüldüğü MTCNN modelinin

başarısını göstermektedir. Aynı zamanda aynı tabloda önerilen dedektörün diğerlerine göre başarısı gösterilmektedir (Li ve diğ., 2020).

Tablo 2.5. RFE-MTCNN'nin Wider Face veri seti ile tespit performansı karşılaştırması

Metod	Ortalama Hassasiyet		
	Kolay	Orta	Zor
MultiScale Cascade	0.691	0.634	0.345
Faceness	0.713	0.664	0.424
LDCF+	0.790	0.769	0.522
MTCNN	0.851	0.820	0.607
Önerilen Model	0.874	0.841	0.673

Literatür çalışmalarımızın sonucunda, yüz tespiti, nesne tespiti için geliştirilen algoritmaların insan yüzü gibi özel bir nesneye uygulanmasıyla gerçekleşir. Nesne tespitinden miras alınan teknikler, yüz tespiti algoritmalarının performansını artırmış ve yüz doğrulama, yüz ifadesi analizi ve yaş tahmini gibi ileri düzey görevler için sağlam bir temel oluşturmuştur.

2.1.3. Yüz Tespitinde Kullanılan Veri Setleri

2.1.3.1. FERET Veri Seti

FERET veri tabanı, yüz tanıma algoritmalarını geliştirmek ve değerlendirmek amacıyla oluşturulmuş bir veri kümesidir. FERET veri tabanı, yüz tanıma alanında araştırmaların geliştirilmesi ve değerlendirilmesi amacıyla Aralık 1993 ile Ağustos 1996 tarihleri arasında toplanmıştır. Bu süreçte, toplam 1.199 farklı bireye ait 14.126 yüz görüntüsü elde edilmiştir. Ayrıca, farklı tarihlerde çekilmiş 365 tekrarlı görüntü seti de veri tabanına dahil edilmiştir. Görseller, başlangıçta 35mm kameralarla çekilmiş, ardından gri tonlamalı (8-bit) 256x384 çözünürlükte dijitalleştirilmiştir. Görsellerde gözler arası mesafe 40-60 piksel aralığında tutulmaya çalışılmıştır. Her görsele, bireylerin kimlik numaralarını, poz bilgilerini ve çekim tarihlerini içeren benzersiz bir dosya adı atanmıştır. 2003 yılında, Savunma İleri Araştırma Projeleri Ajansı (DARPA), bu görüntülerin yüksek çözünürlüklü ve 24-bit renkli versiyonlarını yayınlamıştır (Phillips ve diğ., 1998). Test süreçlerinde kullanılan veri seti ise 856 farklı bireyi temsil eden 2.413 sabit yüz görüntüsünden oluşmaktadır (Li ve Jain, 2011).

2.1.3.2. FDDB Veri Seti

FDDB (Face Detection Data Set and Benchmark), 2845 resim ve 5171 yüz içerir. Veri seti, zor baş pozları, düşük çözünürlük, bulanık yüzler ve yüz tıkanıklıkları gibi çeşitli zorluklar barındırır. Yüz bölgeleri, elips şeklinde ve farklı boyut, şekil ve yönlendirmelere sahip olarak etiketlenmiştir. Hem renkli hem de gri tonlamalı resimler içerir. FDDB, Yahoo! haber web sitesinden derlenen görüntülerden oluşturulmuştur ve büyük poz, aydınlatma ve arka plan çeşitliliğine sahiptir. Başlangıçta otomatik yüz tanıyıcılarla etiketlenen yüzler, doğruluklarını artırmak için yeniden işlenmiş ve yakın kopya görüntüler çıkarılmıştır. Bu, veri setini, daha zorlayıcı ve çeşitli senaryolar için uygun hale getirmiştir (Jain ve Learned-Miller, 2010).

2.1.3.3. Wider Face Veri Seti

WIDER FACE, en büyük yüz tanıma veri setlerinden biridir ve WIDER veri setinden seçilen 32.203 resim ile 393.703 yüz içerir. Önceki kısımlarda belirtilen 12.880 sayısı versiyon farkından kaynaklanmaktadır. Yüzler, ölçek, poz ve engelleme açısından yüksek varyasyon gösterir. Veri seti, 60 etkinlik kategorisine ayrılmıştır ve her bir etkinlik için veriler, eğitim, doğrulama ve test setlerine rastgele %40/%10/%50 oranında bölünmüştür. İki ana test senaryosu bulunmaktadır: Scenario-Ext (dış veri ile eğitim) ve Scenario-Int (WIDER FACE veri ile eğitim). Yüzler, belirgin ölçek ve poz değişiklikleriyle etiketlenmiştir, engellenmiş yüzler ise görünür alanın tahminiyle işaretlenmiştir. Zorluk seviyesi, yüzlerin ölçeği, engellenme durumu ve poz deforme düzeyine göre değişir. WIDER FACE, çeşitli arka planlar ve büyük çeşitliliğe sahip etkinliklerle eğitim için etkili bir kaynak sunar (Yang ve diğ., 2016).

2.1.3.4. AFW Veri Seti

AFW (Annotated Faces in-the-Wild) veri seti, 2013 yılında yayınlanmış ve Flickr'dan rastgele seçilen görüntülerle oluşturulmuştur. Bu veri seti, her biri en az bir büyük yüz içeren 205 görüntüden oluşur ve toplamda 468 yüz etiketlenmiştir. Görüntüler, karmaşık arka planlarla birlikte, yüz bakış açısı ve görünümünde büyük varyasyonlar gösterir (yaşlanma, güneş gözlüğü, makyaj, cilt rengi, ifade vb.). Her yüz, bir sınırlayıcı kutu, 6 işaret noktası (gözler, burun ucu, ağız köşeleri ve merkezi) ve her iki yönü (pitch ve yaw)

boyunca 15° aralıklarla -90° ile 90° arasında bir bakış açısı ile etiketlenmiştir. Ayrıca, yüzün sağ, ortada ve solda olan rollere göre etiketlendikleri de belirtilmiştir. AFW veri seti, benzer "in-the-wild" koleksiyonlardan farklı olarak bir görüntüde birden fazla, frontal olmayan yüzü etiketler (Zhu ve Ramanan, 2012).

2.2. Duygu Tespiti

Yüz ifadeleri, insan düşünceleri ve duygularının en belirgin göstergelerden biri olduğunu daha önce belirtmiştik. Bu ifadeler, izleyicilere niyet, motivasyon ve duygu gibi çeşitli sosyal ipuçları sunarak güçlü bir sessiz iletişim aracı işlevi görür. Yüz ifadelerinin analizi, insan davranışlarını anlamada daha derin bir bakış açısı sağlar. Son yıllarda, Yapay Zekâ (AI) destekli Yüz duygu Tanıma (FER) teknolojisi, dinamik analiz, örüntü tanıma, kişilerarası etkileşim ve zihinsel sağlık izleme gibi alanlarda büyük bir araştırma konusu haline gelmiştir. COVID-19 pandemisi ile birlikte çevrimiçi platformlara geçişin artması, görsel veri üretiminin yoğunlaşmasıyla birlikte yeni bir FER analiz çerçevesine ve bu alanda yenilikçi yaklaşımlara duyulan ihtiyacı gündeme getirmiştir. Ayrıca, çocuklar, yetişkinler ve yaşlı bireylerin yüz ifadelerinde duygu farklılıkları gözlemlenmekte olup, bu durum da duygu tespiti süreçlerinde dikkate alınmalıdır. FER üzerine yapılan araştırmalar oldukça fazladır; ancak, bu çalışmaların kapsamlı bir şekilde sunulduğu ve gelecekteki araştırma yönlerine dair rehberlik sağlayan bir literatür derlemesi bulunmamaktadır (Dalvi ve diğ, 2021).

Duygu tespiti çıkarımı birçok yöntemle ve girdiyle gerçekleştirilmektedir. Yüz tespiti sonrasında yüzdeki mikro ifadeler, yüz üzerinde kas hareketleri, ses tonu ve yazılan metinlerden duygu tespiti yapılabilmektedir. Çalışmamızda yüz tespiti ve buradan yapılan duygu analizleri yöntemleri üzerinde durulacaktır. Öncelikle duygu tespiti yapabilmek için yüzlerdeki ifadelerin hangi duyguyu ifade ettiği ve nasıl temsil edildiğini anlamak gerekir. Bu konuda Paul Ekman, insan yüzlerindeki duygu ifadelerinin evrensel olduğunu ve yedi temel duyguyu (mutlu, üzgün, kızgın, korkmuş, şaşırılmış, tiksinti ve küçümseme) ifade ettiğini belirten ve bu alanda Yüz Hareketi Kodlama Sistemi (FACS) adı ile 1978'de yaptığı çalışma ile ön planda olan bir psikologdur (Ekman ve Friesen, 1978). Bu çalışma Carl-Herman Hjortsjö adlı İsveçli bir anatomist tarafından geliştirilen bir sistemi temel alan çalışmadır (Hjortsjö, 1970). Paul Ekman bu çalışması 2002'de yayınlanan kitabı ile büyük ölçüde güncellemiştir. FACS, yüz ifadelerini 43 kas hareketine dayandırarak analiz

eden ve görüntü tabanlı duygu tespitinde yaygın olarak kullanılan bir sistemdir. Bu sistemin kas hareketlerine karşılık gelen kodlar aşağıdaki Tablo 2.6’da gösterilmiştir.

Tablo 2.6. Yüz ifadesi tanıma için kullanılan ana kodlar

Au	İsim	Au	İsim	Au	İsim	Au	İsim
1	Nötr yüz	2	İç kaş kaldırıcı	3	Dış kaş kaldırıcı	4	Kaş aşağı
5	Üst kapak yükseltici	6	Yanak yükseltici	7	Kapak sıkılaştırıcı	8	Dudaklar birbirine doğru
9	Burun kırışıklığı	10	Üst dudak yükseltici	11	Nazolabial derinleştirici	12	Dudak köşe çektirmesi
13	Keskin dudak çekici	14	Gamze	15	Dudak köşesi bastırıcı	16	Alt dudak bastırıcı
17	Çene yükseltici	18	Dudak büzücü	19	Dil gösterisi	20	Dudak sedye
21	Boyun gerdirici	22	Dudak hunisi	23	Dudak sıkılaştırıcı	24	Dudak bastırıcı
25	Dudak kısmı	26	Çene düşmesi	27	Ağız germe	28	Dudak emmek

Tablo 2.7. Yüz ifadesi tanıma için kullanılan yardımcı kodlar

Derece	Tanım
A	İz, belirgin olmayan etki
B	Hafif
C	Belirgin şekilde
D	Şiddetli veya aşırı
E	Maksimum

Tablo 2.8. Yüz ifadesi tanıma için kullanılan yardımcı kodlar-2

Derece	Tanım
R	Eylem yalnızca yüzün sağ tarafında gerçekleşiyorsa
L	Eylem yalnızca yüzün sol tarafında gerçekleşiyorsa
U	Eylem yalnızca bir yüzde gerçekleşiyor, belirli bir taraf belirtilmiyorsa
A	Eylem iki taraflı olup bir tarafta daha güçlü gerçekleşiyorsa

Duygu durumlarının yüz ifadelerine göre belirli "Action Unit" (eylem birimi) kombinasyonları ile nasıl kodlandığını yedi temel duygu için Tablo 2.9 ile gösterilmiştir. Aksiyon Birimi (AU), yüz ifadelerindeki kas hareketlerini temsil eden bir koddur. Bu kodlar, farklı duyguların nasıl ifade edildiğini analiz etmek için kullanılır (Clark ve diğ., 2020). Tablo 2.7 ve Tablo 2.8’de gösterilen yardımcı kodlar ile birlikte işte yedi temel duygunun "Action Unit" kodlarına göre kısa açıklaması gösterilmiştir.

Tablo 2.9. Yedi temel duygunun "Action Unit" kodlarına göre açıklaması

Duygu	Eylem	Açıklama
Şaşkınlık	1+2+5B+26	Kaşların kalkması (1, 2), gözlerin genişlemesi (5B), ağız açıklığı (26)
Üzüntü	1+4+15	Kaşların birleşip yukarı kalkması (1, 4), dudak köşelerinin aşağı çekilmesi (15)
Mutluluk	6+12	Göz kırışması (6), gülümseme için dudak köşelerinin yukarı çekilmesi (12)
Korku	1+2+4+5+7+20+26	Kaşların kalkması ve birleşmesi (1, 2, 4), gözlerin genişlemesi (5), dudak kenarlarının çekilmesi (20, 26)
Tiksinme	9+15	Burun kırıştırma (9), dudak köşelerinin aşağı çekilmesi (15)
Küçümseme	R12A+R14A	Yüzün sağ tarafında gülümseme (R12A) ve dudak köşesinin yukarı kalkması (R14A)
Öfke	4+5+7+23	Kaşların aşağıya çekilmesi (4), gözlerin kısılması (5), dudakların sıkılması (23)

2.2.1. Temel Duygu Kategorileri

Duygu tespiti, insanların duygusal hallerini anlamaya yönelik önemli bir araştırma alanıdır. Duyguların doğru bir şekilde sınıflandırılması, duygu tespiti algoritmalarının doğruluğunu ve etkinliğini doğrudan etkiler. Literatürde, duygular genellikle üç ana kategoriye ayrılır: pozitif duygular, negatif duygular ve nötr duygular. Bu temel kategoriler, duygusal deneyimlerin çeşitliliğini ve karmaşıklığını daha iyi anlamak için yaygın bir şekilde kullanılmaktadır.

Pozitif duygular, bireylerin hoşnut, huzurlu ve güvenli hissettikleri, yaşam kalitelerini artıran duygulardır. Bu duygular arasında en yaygın olarak karşımıza çıkanlar mutluluk, neşe, huzur, güven, sevgi ve minnettarlık gibi duygulardır. Pozitif duygular, insanların daha sosyal, empatik ve işbirlikçi davranmalarını teşvik edebilir, bu nedenle duygu tespiti araştırmalarında oldukça önemli bir yer tutar. Pozitif duyguların tespiti, çoğunlukla yüz ifadeleri, ses tonları ve vücut dili gibi çeşitli duygusal ipuçlarının analiz edilmesiyle yapılır.

Negatif duygular, bireylerin hoşnutsuzluk, rahatsızlık veya olumsuz bir durumla karşılaştıkları duygulardır. Bu kategorideki duygular arasında öfke, korku, üzüntü, hayal kırıklığı, endişe, suçluluk ve stres gibi duygular yer alır. Negatif duygular, çoğunlukla kaçınma, savaşıma veya donma gibi savunma mekanizmalarını tetikler. Bu duygular,

bireylerin çevrelerine karşı daha dikkatli olmalarını sağlayabilir, ancak uzun süreli olumsuz duygusal durumlar, psikolojik ve fiziksel sağlık üzerinde olumsuz etkiler yaratabilir.

Nötr duygular, bireylerin ne olumlu ne de olumsuz bir duygusal hal içinde oldukları, genellikle düşük yoğunluklu ve geçici durumları ifade eder. Bu tür duygular, genellikle duygusal bir tepki vermeyen durumlarla ilişkilidir. Nötr duygular, bir kişinin rahat bir durumda, kararsız bir ruh haliyle veya duygusal olarak kayıtsız olduğu anlarda görülür. Nötr ifadeler, bazen "duygusal boşluk" olarak tanımlanır ve duygusal anlam taşımayan bir yüz ifadesi ya da davranış biçimiyle kendini gösterir. Tablo 2.6'da temel duyguların valans sınıflandırılması gösterilmiştir.

Tablo 2.10. Duygu Sınıflandırması

Pozitif		Negatif			Nötr	
Mutluluk	Süpriz	Üzgün	Kızgın	Korku	Sinirli	Doğal

Duygu tespiti literatüründe, bu üç temel duygu kategorisi sıklıkla çeşitli sınıflandırma modellerine entegre edilir. Özellikle derin öğrenme ve yapay zekâ tabanlı algoritmalar, yüz ifadelerinin ve ses tonlarının analizi ile bu duygusal kategorileri sınıflandırmayı hedefler.

Örneğin, yüz ifadeleriyle duygu tespiti yapan modeller, genellikle Ekman'ın temel duygularını kullanırken, bazı gelişmiş modeller, daha geniş bir duygu yelpazesi sunarak karmaşık duygusal durumları tanımaya çalışmaktadır.

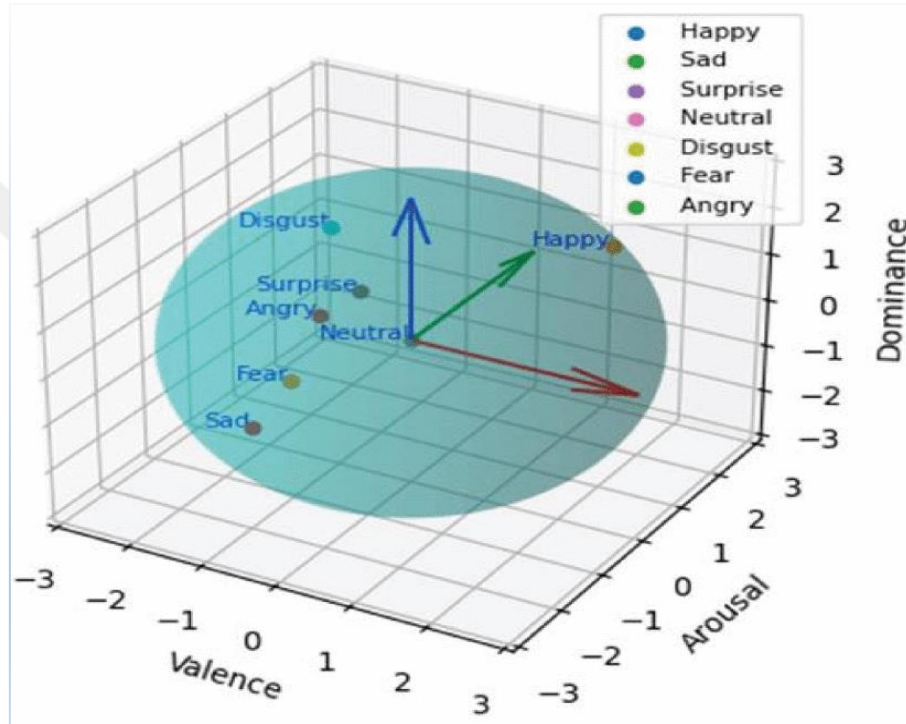
Geleneksel duygu etiketleme, genellikle yedi temel duygu kategorisini içerir: öfke, tikslenme, korku, sevinç, üzüntü, sürpriz ve nötr. Ancak, duygu zekasına olan artan talep ile daha ayrıntılı ve doğru duygu durumlarının elde edilmesi gerekliliği doğmuştur.

Kategorik Model: Duygular, Ekman'ın temel 7 duygusu gibi sınırlı bir kategori setinden seçiliyor. Karmaşık veya yoğunluk farklılıkları bu modelde yeterince ifade edilemiyor.

Boyutsal Model: Duygular, sürekli bir ölçek üzerinde "valans" (olumlu-olumsuz) ve "uyarılma" (heyecanlı-sakin) boyutlarıyla temsil ediliyor. Bu model, duygusal farklılıkları ve yoğunluğu daha iyi ifade edebiliyor.

Bu bağlamda boyutsal modeller, duyguları sürekli bir uzayda temsil eder ve genellikle aşağıdaki üç boyut kullanılır;

- Valans (V): Duygunun pozitif veya negatif oluşunu ifade eder.
- Uyarılma (A): Duygunun yoğunluğunu veya enerjisini belirtir.
- Dominans (D): Duygunun kontrol gücünü veya bireyin çevresi üzerindeki hakimiyetini tanımlar.

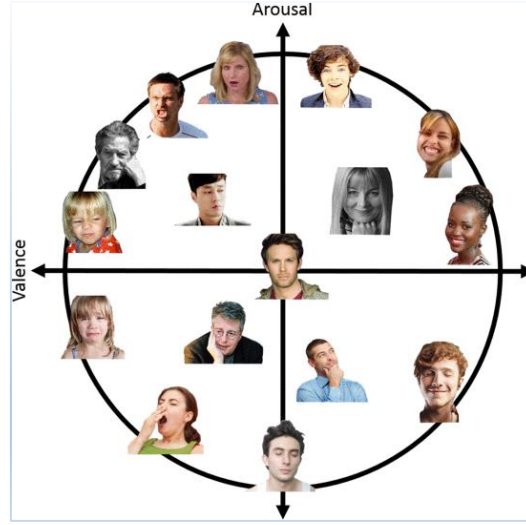


Şekil 2.6. Boyutsal modelleme (Huo ve Ge, 2024)

Şekil 2.6’da temel 7 duygunun yerleştirilmesi gösterilmiştir. Yi Huo ve arkadaşların daha ayrıntılı duygu tespili için, VA (Valans-Arousal) modelinin, duyguları valans ve uyarılma düzeylerine göre sınıflandırmasının yanı sıra, Dominance (D) boyutunun da önemli bir faktör olduğunu vurguladılar.

Bu, sosyal bağlamda duyguları anlamada kritik bir rol oynamaktadır. Çalışma, bu bulgudan yola çıkarak, duygusal kategorilerin üç boyutlu VAD (Değerlik-Uyarılma-Dominans) uzayında noktalar olarak değerlendirilebileceğini öne sürmektedir.

Şekil 2.7’de örnek duyguların valans ve uyarılma ekseninden yerleştirilmesi gösterilmiştir (Huo ve Ge, 2024).



Şekil 2.7. Değerlik uyarılma döngüsündeki örnek görüntüler

2.2.2. Duygu Tespit Modelleri

Duygu tanımlamada kullanılan modeller CNN mimarisi üzerine geliştirilmiş modellerdir. Bu modeller kullanılan ortama ve veri setlerine göre değişiklik gösterebilmektedir.

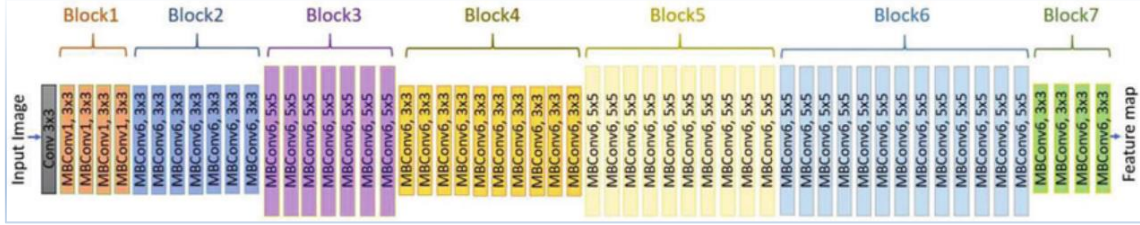
2.2.2.1. EfficientNetB7

EfficientNetB7, EfficientNet ailesinin en büyük ve en güçlü modelidir. Bu model, bileşik ölçeklendirme (compound scaling) yöntemiyle tasarlanarak derinlik, genişlik ve giriş çözünürlüğü açısından optimize edilmiştir. EfficientNetB7, büyük veri setlerinde üstün doğruluk sağlamanın yanı sıra verimlilik açısından da dikkat çekicidir; daha önceki en iyi CNN modellerine kıyasla 8.4 kat daha küçük bir model boyutuna ve 6.1 kat daha hızlı bir işlem hızına sahiptir.

EfficientNetB7'nin mimarisi, mobil tersine çevrilmiş dar boğaz konvolüsyon (MBConv) bloklarına dayanır ve bu bloklar "squeeze-and-excitation" optimizasyonu ile güçlendirilmiştir. Modelin yapısı, filtre boyutları, stride değerleri ve kanal sayısına göre yedi farklı bloğa ayrılır. Bu bloklar, derin öğrenme görevlerinde daha zengin ve detaylı özellik haritaları çıkarılmasını sağlar.

EfficientNet modelleri, diğer modellere kıyasla daha az parametre ile çalışarak daha iyi sonuçlar sağlar. Bu, özellikle büyük veri setlerinde hesaplama maliyetlerini düşürürken doğruluğu artırmak için avantajlıdır. Şekil 2.8'de EfficientNet-B7 derin öğrenme

modelinin farklı aşamalarını ve katmanlarını göstermektedir. Model, giriş görüntüsünü alarak bir dizi MBConv (Mobile Inverted Bottleneck Convolution) bloğundan geçirilerek derin özellik haritaları oluşturur (Khalil ve diğ., 2022).



Şekil 2.8. EfficientNetB7 mimarisi (Baheti ve diğ., 2020)

EfficientNetB7, yüksek çözünürlüklü giriş verileri üzerinde çalışmak için tasarlandığından, görüntü tabanlı görevlerde (örneğin, sınıflandırma ve semantik segmentasyon) hem doğruluk hem de genel performans açısından diğer varyantları geride bırakır. Özellikle araştırmalarda, EfficientNetB7'nin UNet tabanlı bir decoder ile birleştirilmesi, segmentasyon haritalarının oluşturulmasında etkili sonuçlar sağlamıştır. Bu kombinasyon, modelin güçlü temsil kabiliyetini ve yüksek çözünürlükteki analiz kapasitesini başarılı bir şekilde göstermektedir (Baheti ve diğ., 2020).

2.2.2.2. CNN-LSTM

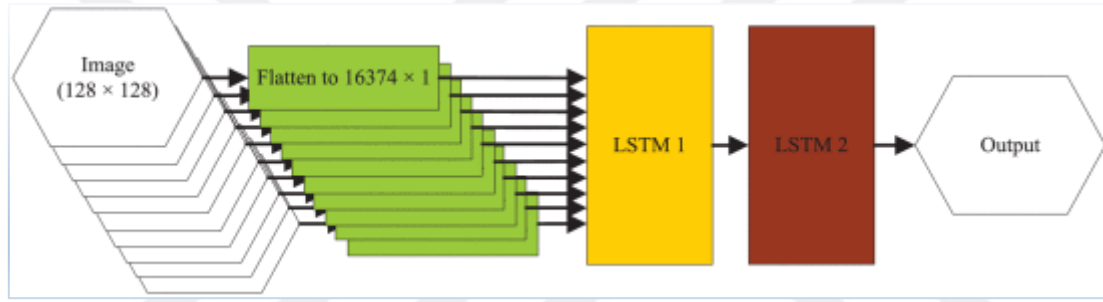
LSTM, özellikle zaman serisi verileri ve ardışık verilerle çalışırken, geleneksel yapay sinir ağlarına göre daha başarılı sonuçlar verebilen bir ağ türüdür. LSTM, geri besleme (feedback) mekanizması sayesinde, önceki zaman adımlarındaki bilgiyi daha uzun süre hatırlayabilir ve bu da onları uzun bağımlılıkları öğrenme konusunda daha etkili kılar.

LSTM hücreleri, üç ana bileşenden oluşur: Yeni bilginin hücreye dahil edilmesini kontrol eder, Hücredeki eski bilgilerin ne kadarının unutulacağına karar verir, Hücreden çıkan bilginin ne kadarının bir sonraki zaman adımına geçeceğine karar verir.

Ancak, LSTM'ler yüksek boyutlu verilerle doğrudan çalışmakta zorlanabilir, çünkü veriler çok büyük olduğunda (örneğin, 128x128 piksel bir görüntü), LSTM'nin öğrenmesi için gereken parametre sayısı çok fazla olur. Bu da modelin verimli çalışmasını engelleyebilir (Li ve diğ., 2019).

Örneğin, LSTM katmanına doğrudan 128x128'lik bir görüntü verildiğinde, bu girişin vektör haline getirilmesiyle oluşan 16384x1'lik bir giriş çok büyük olur ve bu da 17 milyon parametreye kadar çıkabilir. Ancak, CNN ile bu giriş boyutu küçültülüp, 2048x1 gibi bir vektör haline getirildiğinde, LSTM katmanlarındaki parametre sayısı 2.3 milyon'a düşer. Bu, çok daha verimli bir modelin ortaya çıkmasını sağlar.

CNN (Convolutional Neural Network) genellikle görüntüler gibi yüksek boyutlu verilerin işlenmesinde kullanılır. CNN, görüntüdeki önemli özellikleri çıkararak boyutları küçültür ve modelin öğrenmesi için daha yönetilebilir hale getirir. Bu özellik çıkarımı, özellikle yüz tanıma veya duygu analizi gibi görevlerde çok önemlidir Şekil 2.9'da LSTM entegre edilmiş bir model mimarisi gösterilmektedir.



Şekil 2.9. Örnek bir LSTM mimarisi (Li ve diğ., 2019)

2.2.2.3. CNN

Bu alanda en çok tercih edilen yöntemlerden biri, Convolutional Neural Networks (CNN), yani Evrimsel Sinir Ağlarıdır (Shawar ve Atwell, 2005) (Stilgoe, 2017). CNN'ler, ayırıcı özellikleri otomatik olarak çıkarabilme kapasitesine sahip oldukları için, bu tür tanıma görevlerinde oldukça etkilidir (Sharma ve diğ., 2020) (Loussaief ve Abdelkrim, 2016) (Coşkun ve diğ., 2017).

CNN'lerin tarihi, 1989 yılında LeCun'un el yazısı tanıma çalışmasıyla başlamıştır (LeCun ve diğ., 1989) ve sonrasında yapılan büyük yarışmalarla (örneğin ILSVRC-2012 (Krizhevsky ve diğ., 2017), ICML 2013 ve EmotiW 2013 (Kahou ve diğ., 2013) bu alandaki gelişmeler hız kazanmıştır. Ancak, daha derin yapılar içeren Deep CNN (DCNN) mimarileri, daha karmaşık sorunları çözebilmek için daha fazla hesaplama gücü gerektirmektedir. Bu tür yapılar, çok sayıda parametre içerdiği için küçük veri setleriyle eğitildiğinde modelin aşırı uydurulması (overfitting) gibi sorunlarla karşılaşılabilir.

Son yıllarda ise, büyük modellerin daha verimli ve kompakt hale getirilmesine yönelik çalışmalar hız kazanmıştır. Örneğin, MobileNetV1 ve MobileNetV2, VGG16'ya yakın doğruluklarla çalışırken, daha az parametre kullanarak çok daha küçük boyutlara sahiptir. Bunun yanı sıra ShuffleNet ve EfficientNet b0 gibi modeller de düşük parametre sayısı ile benzer performanslar sergileyebilmiştir. (Gursesli ve diğ., 2024)

Bu gelişmeler, özellikle duygu tanıma (FER) gibi alanlarda doğruluğu artırmak için büyük fırsatlar yaratmıştır. Örneğin, Zeng (Zeng ve diğ., 2018) gibi araştırmacılar, el yapımı özellikler ve derin ağ öğrenmesi gibi teknikleri birleştirerek FER-2013 veri seti üzerinde başarılarını artırmışlardır.

2.2.3. Duygu Tespitinde Kullanılan Veri Setleri

Duygu tanıma araştırmalarında kullanılan veri setlerinin kronolojik ve metodolojik sınıflandırması şu şekildedir:

Temel Laboratuvar Temelli Veri Koleksiyonları: JAFFE, Cohn-Kanade, MMI ve MultiPie gibi referans veri setleri, kontrollü laboratuvar koşullarında, deneklerin spesifik emosyonel ekspresyonları simüle etmesiyle oluşturulmuştur. Bu koleksiyonlar optimum görüntü kalitesi ve standardizasyon sunmakla birlikte, spontane duygusal ifadelerin otantik karakteristiğini tam anlamıyla yansıtmamaktadır.

Spontane Ekspresyon Metodolojisi: DISFA, AM-FED ve Belfast veri setleri, bireylerin doğal yaşam koşullarındaki kendiliğinden gelişen duygusal manifestolarını kayıt altına almak amacıyla tasarlanmıştır. Bu koleksiyonlar multimodal veri akışları (akustik, fizyolojik parametreler vb.) içermekle beraber, demografik çeşitlilik açısından kısıtlılıklar göstermektedir.

Naturalistik Ortam Veri Koleksiyonları: AFEW ve SFEW: Sinematografik materyallerden ekstrakte edilmiş duygusal ekspresyonları içermektedir, ancak görsel örneklem ve denek sayısı açısından sınırlılıklar mevcuttur.

FER-2013: Google Görsel Arama platformu üzerinden derlenmiş, 35.887 görüntüden oluşan kapsamlı bir veri setidir. Bununla birlikte, düşük çözünürlük ve duygusal ifadelerin homojen dağılımı gibi metodolojik kısıtlılıklar içermektedir.

FER-Wild: Yüksek çözünürlüklü 24.000 görüntü barındırmakla beraber, spesifik emosyonel kategorilerde (örneğin disgust ve fear) örneklem yetersizliği gözlenmektedir.

Çağdaş Geniş Ölçekli Veri Setleri: EmotionNet, internet ortamından toplanan bir milyon görüntüyü kapsamaktadır. Otomatize anotasyon algoritmaları kullanılarak duygusal sınıflandırma yapılmıştır, ancak dimensional model entegrasyonu eksikliği gibi limitasyonlar mevcuttur.

Aff-Wild: Valens-arousal paradigması çerçevesinde natüralistik ortamda oluşturulmuş en kapsamlı veri setlerinden biridir, fakat katılımcı demografisi açısından kısıtlılıklar içermektedir.

Dimensional Model Veri Setleri: Belfast, RECOLA ve DEAP gibi koleksiyonlar, emosyonel durumları sürekli değişkenler (valens, arousal) üzerinden quantifiye etmektedir. Ancak anotasyon prosedürünün kompleks olması nedeniyle, genellikle sınırlı katılımcı sayısı ile karakterizedir.

Benchmark Araştırmaları ve Kompetisyonlar: AVEC serisi yarışmalar, sürekli emosyonel durum predikasyonuna yönelik veri setleri ve modellerden oluşmaktadır. SEMAINE ve RECOLA gibi veri setleri bu kompetisyonlarda kullanılmıştır.

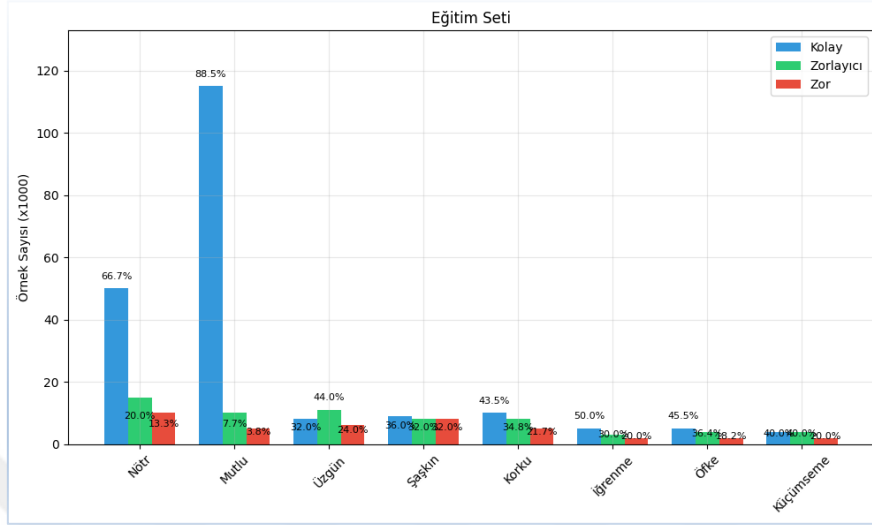
Aff-Wild Veri Seti Spesifikasyonları: YouTube platformundan derlenmiş 500'ü aşkın video sekansı içermektedir. Her video sekansı, valens ve arousal parametreleri açısından frame-by-frame anotasyona tabi tutulmuştur. Bununla birlikte, katılımcı demografik çeşitliliği açısından optimum seviyeye ulaşamamıştır (Mollahosseini ve diğ., 2019).

Bu veri setleri, farklı yaş grupları, cinsiyetler, kültürel geçmişler ve gerçek dünya koşullarındaki çeşitlilikle zenginleştirilmiştir. Her biri, temel duyguları sınıflandırmak veya karmaşık ifadeleri analiz etmek amacıyla farklı özelliklere sahip veriler sunmaktadır.

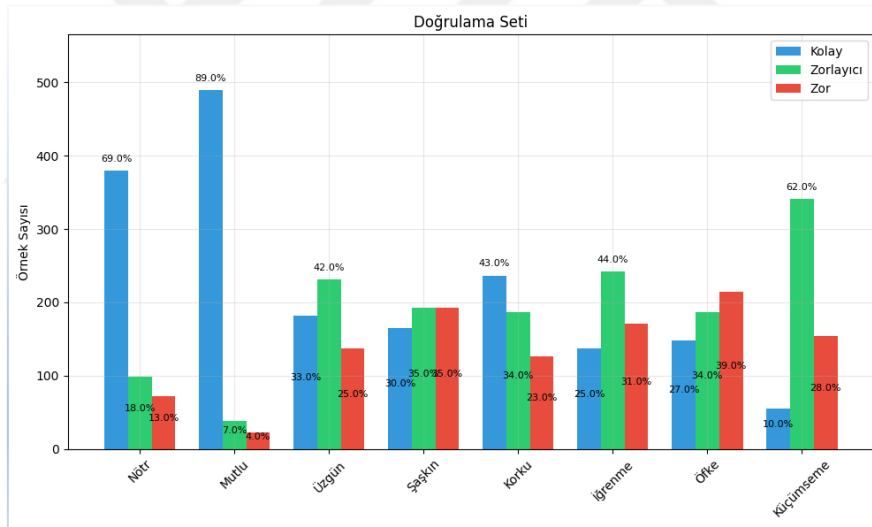
2.2.3.1. AffectNet Veri Seti

Doğal ortamlarda (yani laboratuvar ortamı dışında) kaydedilmiş duygu ifadelerini içeren en büyük veri tabanıdır. İki farklı duygu modeli içerir. Kategorik model, duyguları

mutluluk, üzüntü gibi belirli kategorilere ayıran ve boyutsal model, duyguları valens ve uyarılma gibi sürekli değerler üzerinde ölçen modellerdir.



Şekil 2.10. AffectNet eğitim seti dağılımı



Şekil 2.11. AffectNet doğrulama seti dağılımı

Şekil 2.10 ve 2.11’de AffectNet veri setinin eğitim ve doğrulama dağılımı gösterilmektedir. AffectNet, çevrim içi ortamdan alınan kategorik ve boyutsal duygusal modelleri içeren en büyük yüz ifadeleri veri setidir. Veri seti, duygu ile ilgili anahtar kelimelerin üç arama motorunda sorgulanması ve uzmanlar tarafından etiketlenmesiyle oluşturulmuştur.

Sorgulama Süreci: “Mutlu kız”, “öfkeli adam” gibi yaklaşık 362 İngilizce sorgu, cinsiyet, yaş ve etnik köken gibi terimlerle birleştirilerek oluşturulmuştur. Bu sorgular İspanyolca,

Portekizce, Almanca, Arapça ve Farsça dillerine çevrilmiş; yerel konuşmacılar tarafından kültürel uygunluk açısından gözden geçirilmiştir.

Arama Motorları: Google, Bing ve Yahoo kullanılmış, Baidu ve Yandex ise veri çekim sınırlamaları nedeniyle tercih edilmemiştir.

Filtreleme: Çizim, animasyon ve stok fotoğraflar gibi insan yüzü içermeyen veya yapay görselleri elemek için negatif terimler ve filtreler uygulanmıştır.

Sonuç: Toplamda 1,8 milyon URL'den yüzler tespit edilmiş, 1 milyonun üzerinde yüz içeren görseller veri setine dahil edilmiştir. Ortalama yüz çözünürlüğü 425x425 pikseldir.

Etiketleme Yöntemi: Amazon Mechanical Turk gibi kitlesel etiketleme hizmetleri yerine, Denver Üniversitesi'nden 12 uzman etiketleme uzmanı kullanılmıştır. Toplam 450,000 görsel, kategorik (mutlu, üzgün, şaşkın vb.) ve boyutsal (valans ve uyarılma) olarak etiketlenmiştir.

Araçlar: Etiketleme için bir yazılım geliştirilmiş; etiketleme uzmanları detaylı eğitim ve örneklerle yönlendirilmiştir.

Kategoriler: Mutlu, Üzgün, Şaşkın, Korkmuş, Öfkeli, Tiksinmiş, Hor görme, Nötr, Belirsiz ve Yüz Yok kategorileri kullanılmıştır. Görselin kategoriye uygunluğu, yoğunluk düzeyine bakılmaksızın değerlendirilmiştir.

Sonuçlar: Mutlu ifadeler en yüksek oranla (%48) etiketlenmiş, bunu nötr ifadeler takip etmiştir. Tiksinme, korku ve hor görme ifadeleri en düşük oranlarda etiketlenmiştir (%2-4).

Valans ve Uyarılma: Valans, olayın pozitif veya negatif oluşunu; uyarılma ise ne kadar sakin ya da heyecanlı olduğunu ifade eder. Etiketleme uzmanları, her yüzün valans ve uyarılma değerlerini iki boyutlu bir koordinat sisteminde belirtmiştir.

Denetim: Yazılım, kategorik ifadelerin belirli valans ve uyarılma aralıklarını aşması durumunda uyarı vermiş, ancak etiketleme uzmanları emin olduklarında bu sınırların dışına çıkabilmiştir.

Sonuç olarak AffectNet, yüz ifadelerinin hem kategorik hem de boyutsal özelliklerini içeren geniş kapsamlı bir veri setidir. Bu veri seti, duygusal ifadelerin otomatik algılanması ve analizinde önemli bir kaynak sunmaktadır (Mollahosseini ve diğ., 2019).

2.2.4. Duygu Tespitinde Karşılaşılan Zorluklar

Duygu tanıma sistemlerinin geliştirilmesi, birçok faktör nedeniyle önemli zorluklarla karşılaşmaktadır. Duygu ifadelerinin algılanması sadece temel duygusal tepkilerle sınırlı olmayıp, kültür, ırk, dil ve din gibi değişkenlerden de etkilenmektedir. Bu çeşitlilik, sistemlerin genelleştirilebilir özelliği ve doğruluğu üzerinde olumsuz etkiler yaratmaktadır. Ayrıca, yüksek hesaplama gücü gereksinimi, özellikle mobil cihazlar ve düşük güçlü sistemler için, hafif modeller geliştirme ihtiyacını gündeme getirmektedir. Ancak bu alanda yapılan çalışmalar henüz istenen düzeye ulaşmamıştır (Gursesli ve diğ., 2024).

Duygu ifadeleri, duygu durumu analizi için değerli bilgiler sağlasa da bireylerin ifadeleri bilinçli olarak gizleyebilmesi veya manipüle edebilmesi sistemlerin güvenilirliğini düşürmektedir. Bunun yanı sıra, ifadelerde belirgin değişikliklerin olmadığı durumlarda model performansı önemli ölçüde azalmaktadır (Ji ve Dong, 2024). İnsan yüz ifadelerinin makineler için karmaşık sinyaller olarak kalması, duyguların evrensel tanımı konusundaki tartışmalarla daha da karmaşık hale gelmektedir. Örneğin, Ekman'ın yedi temel duygu modeli, yaş, cinsiyet ve kültürel faktörlerin etkilerini yeterince yansıtmadığı gerekçesiyle eleştirilmektedir. Bu durum, veri kümelerinin homojenliğini ve doğal ortam koşullarında kaydedilen ifadelerin kalitesini de olumsuz etkilemektedir (Dalvi ve diğ., 2021).

COVID-19 pandemisi sürecinde yaygınlaşan maske kullanımı, yüz bölgesindeki kısmi kapanma nedeniyle duygu tanıma sistemlerini daha da zorlaştırmıştır. Bu durum yüz bölgesindeki başka cisimlerle kapatılmış alanlara dikkat çekilmesine sebebiyet verdi. Örnek olarak, pozitif duyguların genellikle alt yüz bölgesinden algılanması, maske kullanımının bu bölgeyi kapatmasıyla performans kaybına yol açmaktadır. Bu durum, maskeli ve maskesiz görüntüler arasında transfer öğrenme yaklaşımlarını geliştirme gerekliliğini artırmıştır (D'incà ve diğ., 2023).

Duygu tanıma sistemlerinin karşılaştığı bir diğer temel zorluk, yüz ifadelerinden özellik çıkarımı ve model eğitimi süreçlerinin yüksek hesaplama maliyetidir. İnsan duygularının fizyolojik ve non-fizyolojik tepkiler şeklinde manifestasyonu, model yapılarını daha karmaşık hale getirmektedir. Geleneksel tek modallı yaklaşımların sınırlılıkları, multimodal sistemlere olan ilgiyi artırmıştır. Ancak bu tür sistemlerde, modlar arası korelasyonun optimizasyonu ve temporal yüz ifadelerinden kritik çerçevelerin seçimi gibi metodolojik zorluklar devam etmektedir (Wang ve diğ., 2023).

2.3. Ortam Duygu Yoğunluğu Analizi

Ortam duygu yoğunluğu analizi, bir ortamın genel duygusal atmosferini tespit etmeyi amaçlayan kapsamlı bir yaklaşımdır. Bu analiz, bireylerin yüz ifadelerinden elde edilen duygusal durumların toplu değerlendirilmesine dayanmaktadır. Sistem, insan yüzlerindeki temel duyguların belirlenmesi ve bu verilerin matematiksel yöntemlerle işlenmesi prensibine göre çalışmaktadır.

Sistemin temel işleyişi, görüntü işleme teknolojilerinin kullanımıyla başlar. Haar Cascade gibi geleneksel yöntemler ve MTCNN gibi günümüz derin öğrenme teknikleri gibi algoritmalarla yüz tespiti gerçekleştirilirken, EfficientNet-B7, Derin CNN ve ConvLSTM gibi modellerle temel duygular sınıflandırılır. Her karede tespit edilen yüz ifadelerinin ortalaması alınarak ortamın genel duygusal durumu belirlenir, anlık değişimler ve duygu değişim süreçleri analiz edilerek zaman içerisindeki eğilimler ortaya konur.

Bu teknolojinin çeşitli uygulama alanlarında önemli katkıları bulunmaktadır. İş toplantıları, dersler ve seminerlerde katılımcıların duygusal tepkileri analiz edilerek verimlilik artırılabilir. Sağlık sektöründe, doktor-danışan görüşmelerinde bireylerin duygu durumlarının izlenmesi, terapi süreçlerinde daha etkili müdahaleler sağlar. Pazarlama ve müşteri ilişkileri alanında, müşteri duygularının analizi ile hizmet veya ürün geliştirme süreçleri iyileştirilebilir. Eğitim alanında ise öğrencilerin derslere olan duygusal tepkileri analiz edilerek öğretim yöntemleri optimize edilebilir.

Ortam duygu yoğunluğu analizi, görüntü işleme ve yapay zekâ tekniklerini birleştirerek, bireylerin ruh hali ve sosyal etkileşimlerine dair değerli bilgiler sunar. Bu bilgiler, farklı sektörlerde karar verme süreçlerini destekleyerek, hizmet kalitesini artırmada önemli rol

oynar. Gerçek zamanlı veri işleme ve analiz yetenekleriyle, anlık müdahale ve optimizasyon imkânı sağlayan bu teknoloji, modern yaşamın çeşitli alanlarında yüksek katma değerli çözümler sunabilir.

2.3.1. Yöntemlerin Karşılaştırılması

Duygu tespiti ve yüz tanıma alanında kullanılan farklı yöntemler, performans, doğruluk ve verimlilik açısından çeşitli avantajlar ve dezavantajlar sunmaktadır. Bu bölümde, bu yöntemlerin karşılaştırılması ele alınmıştır.

Viola-Jones algoritması, geleneksel yöntemler arasında popüler bir yüz tespit yöntemi olarak dikkat çekmektedir. Düşük çözünürlüklü görüntülerde gerçek zamanlı algılama sağlayabilen bu algoritma, yüksek çözünürlüklü veri setleri karşısında performans sınırlamaları yaşamaktadır. Weiyan Wang ve ekibinin GPU tabanlı optimizasyon çalışmaları, algoritmanın performansını artırmış, CPU uygulamalarına kıyasla 13 kata kadar hız artışı sağlamıştır. Ancak, bu optimizasyonlar daha çok donanımsal destekle sınırlıdır ve algoritmanın temel yapısında bir değişiklik yapılmamıştır.

HOG yöntemi, yüz tanıma için dayanıklı ve hızlı bir özellik çıkarma yöntemi olarak öne çıkmaktadır. Gabor dalgacı ve yerel ikili model (LBP) yöntemleriyle karşılaştırıldığında, HOG yöntemi zaman ve ortam değişikliklerine daha dayanıklı sonuçlar vermektedir. Özellikle, işlem süresi açısından Gabor yönteminden daha hızlıdır. FERET veri setinde yapılan testler, HOG'un aydınlatma ve çözünürlük değişimlerine karşı yüksek performans gösterdiğini ortaya koymuştur. Bununla birlikte, LBP yöntemine kıyasla boyut avantajı ve tanıma oranı daha yüksektir.

Şablon eşleştirme ile 2DPCA (İki Boyutlu Ana Bileşen Analizi) algoritmasının birleşimi, yüz tespitinde doğruluk oranını artıran bir yöntem olarak karşımıza çıkmaktadır. Tekli yüz testlerinde %92, çoklu yüz testlerinde ise %85,96 doğruluk oranına ulaşmıştır. Ayrıca, işlem hızında belirgin bir artış sağlanmıştır.

Bu yöntemin özellikle karmaşık arka planlara sahip görüntülerde etkili olduğu gözlemlenmiştir. Ancak, şablon eşleştirme aşamasında kullanılan algoritmaların, veri setine göre optimize edilmesi gerekmektedir.

EfficientNetB7 modeli, bileşik ölçeklendirme prensibiyle derin öğrenme modelleri arasında üstün bir performans sergilemektedir. Hem daha az parametre kullanması hem de yüksek doğruluk oranı sağlaması, bu modeli öne çıkarmaktadır. Görüntü sınıflandırma ve segmentasyon gibi yüksek çözünürlüklü görevlerde diğer yöntemlerden daha başarılıdır. Ancak, modelin büyük veri setleri üzerinde eğitilmesi ve optimize edilmesi gerekmektedir.

CNN'ler, görüntü özelliklerini etkili bir şekilde çıkarma kapasitesiyle duygu analizi ve yüz tanıma alanında sıklıkla kullanılmaktadır. Bununla birlikte, zaman serisi verileriyle çalışılması gerektiğinde, CNN'lere LSTM katmanları eklenerek performans artırılmaktadır. CNN-LSTM kombinasyonu, ardışık görüntülerde hem verimlilik hem de doğruluk açısından avantaj sağlar. Ancak, LSTM katmanlarının doğrudan yüksek boyutlu verilerle çalışamaması nedeniyle CNN ile ön işleme yapılması gerekmektedir.

Bu yöntemler, farklı veri türleri ve uygulama alanlarına göre avantaj ve dezavantajlar sunmaktadır. Geleneksel yöntemler (Viola-Jones, HOG), belirli veri koşullarında etkili olurken, derin öğrenme tabanlı modeller (EfficientNetB7, CNN-LSTM) daha geniş veri kümeleri üzerinde yüksek doğruluk oranları sağlamaktadır. Ancak, derin öğrenme modellerinin yüksek hesaplama maliyetleri ve eğitim süreci, bu yöntemlerin uygulanabilirliğini sınırlayabilmektedir. Yöntem seçiminde, kullanılacak veri setinin yapısı ve uygulama alanının gereklilikleri dikkate alınmalıdır.

2.3.2. Ortam Uygulamaları

Ortam duygu yoğunluğu analizi alanındaki teknolojik gelişmeler ve metodolojik ilerlemeler, bu sistemlerin uygulama alanlarını sürekli genişletmektedir. Bu gelişmeler, insan-bilgisayar etkileşiminin optimizasyonuna katkı sağlarken, çeşitli sektörlerde daha sofistike ve verimli uygulamaların geliştirilmesine olanak tanımaktadır. Bu alandaki araştırmalar ve inovasyonlar, duygu analizi sistemlerinin adaptif yeteneklerini artırarak, daha etkili ve kullanıcı odaklı çözümlerin ortaya çıkmasını sağlamaktadır (Wang ve Gu, 2018).

L Alberto yüz ifadelerinin tanınması için bir robot geliştirmiş ve birleştirilmiş bir algoritma (PCA + ANN + GA) kullanmıştır. Bu algoritmanın en yüksek tanıma oranını

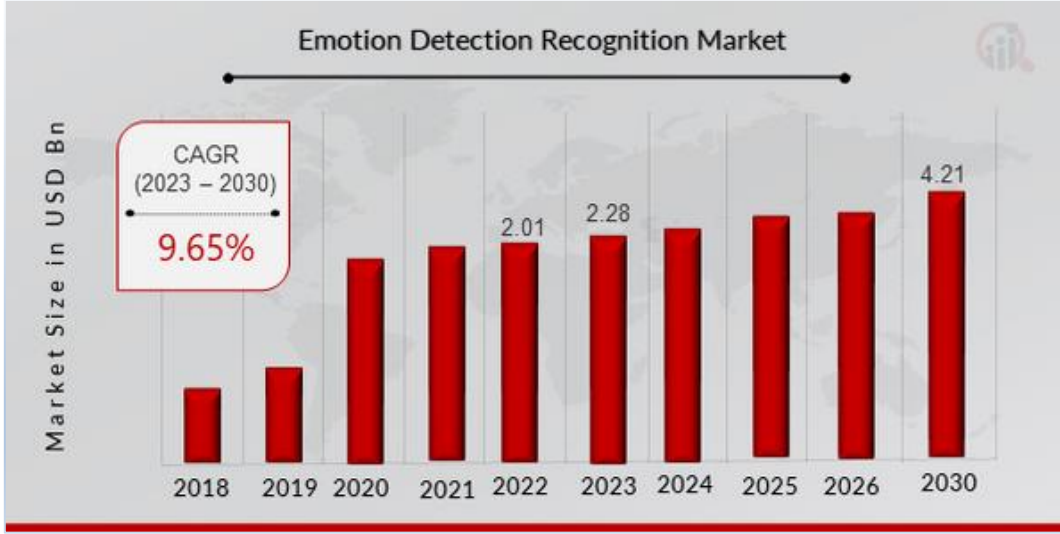
elde ettiđi ve özellikle PCA yönteminin "üzüntü" duygusunu tanımada daha başarılı olduđu belirtilmiştir (Perez-Gaspar ve diğ., 2016).

Zhentaο Liu ve ekibi, FEER-HRI adlı bir sistem geliştirerek, insan-bilgisayar etkileşiminde yeni gereksinimlere uyum sağlamayı amaçlamıştır. Bu sistem, robotların sadece ifadeleri tanınmasını deđil, aynı zamanda insan duygularına uygun grafikler üretmesini sağlamaktadır. Algoritmada 2D-Gabor, LBP ve over-limit learning machine (Aşırı limitli öğrenme) yöntemleri kullanılmıştır. Robotlar, ifadeleri basit birer çizgi film sembolüyle temsil etmiş ve bu semboller LED ekranlarda gösterilmiştir. Sistem, rehberlik, eğlence, ev hizmetleri ve simülasyon gibi dört farklı senaryoda başarılı şekilde çalışmıştır (Liu ve diğ., 2017).

Microsoft'un Azure platformunda, ifadeleri tanımak için API'ler barındıran ticari bir bulut hizmeti olarak tanıtılmaktadır. Azure'un gücü, geliştirme, test etme, uygulama dağıtımı ve yönetimini kolaylaştırarak tasarım verimliliğini artırmasından gelmektedir. Ayrıca, bu hizmetlerin farklı cihazlar (iOS, Android, Windows) ve platformlar arasında kolayca kullanılabilir olduđu belirtilmiştir.

Ramon Rodriguez ve arkadaşları LSTM ile dijital göçmenlerin (dijital çağ öncesinde doğmuş ve teknolojiyi hayatlarının ilerleyen dönemlerinde öğrenmiş bireyler) teknoloji kullanımı sırasında yaşadıkları duygusal durumları analiz etmektir. Özellikle, yüz ifadelerini kullanarak duyguları sınıflandırmak ve bu duygulara etki eden faktörleri incelemek hedeflenmektedir. Araştırma, dijital göçmen olan öğretim üyelerine Microsoft uygulamaları (Office365, Microsoft Teams, Microsoft Forms) kullanmayı öğretmek için bir eğitim müdahalesi gerçekleştirilmiştir. Müdahale sırasında öğretim üyelerinin yüz ifadeleri bir web kamerası ile kaydedilmiştir (Rodriguez ve diğ., 2023).

Ekonomik Pazar karşılığı olarak, uygulama bazında Hukuk, Gözetim ve İzleme, Pazarlama ve Reklamcılık, Medya ve Eğlence gibi segmentlere ayrılır. Hukuk sektörü, 2023-2030 dönemi boyunca en hızlı büyümesi beklenen segmenttir. Bu teknoloji, kamu güvenliğini artırmak için bireylerin duygularını ve niyetlerini değerlendirmede kullanılmaktadır. Aşağıdaki Şekil 2.12 de duygu tanımlama ile ilgili 2023 ve 2030 yılları arasında tahmini pazar büyüme miktarları gösterilmiştir.



Şekil 2.12. FER pazar payı

2022 yılında 2,01 milyar USD olan pazar büyüklüğünün, 2023-2030 yılları arasında %9,65 bileşik yıllık büyüme oranıyla 2030'da 4,21 milyar USD'ye ulaşması beklenmektedir. Pazar, özellikle askeri, savunma ve sağlık gibi sektörlerde yeni uygulama alanlarıyla genişlemeye devam etmektedir.

2.4. Mevcut Çalışmaların Değerlendirilmesi

Ning Zhou ve arkadaşları duygu tanımlamasın konusunda çalışmamıza en yakın olan yöntemleri kullandılar. MTCNN ile yüz tespiti ve geleneksel derin evrişimli sinir ağı modelindeki tam bağlantılı katmanın yerine Küresel Ortalama Havuzlamayı kullanarak parametreleri azaltmış ve hafıza kaynaklarının işgali azalmıştır. Çalışmada belirtildiği gibi gerçek zamanlı senaryolarda test edilmemiş olması ve düşük doğruluk oranı sebebi ile performans olarak düşüktür. Aynı zamanda yoğunluk üzerine odaklanmamıştır (Zhou ve diğ., 2020).

Mustafa Can Gürsesli ve arkadaşlarının çalışmasında yine birçok veri seti üzerinden test edilen evrişimli sinir ağı modeline dayanmaktadır. Bu çalışmanın odak noktası yüksek performans karşılaştırılmasıdır. Bu çalışma ana odak noktamız olan yoğunluk tespitine odaklanmamıştır (Gursesli ve diğ., 2024).

Michal Krumnikl tarafında ise çalışmamızın hedef noktalarından biri üzerine odaklanmıştır. Mobil uygulama üzerin tekli duygu tespiti üzerine çalışılmıştır. Farklı model ve teknolojileri üzerine birçok test yapılmıştır. Çalışmada gecikme oranları gibi

değerler üzerine testler yapılması vurgulanmıştır. Bu durum bizim çalışmamızdaki en önemli parametrelerden biri olan çalışma hızı bakımından yetersizdir (Krumnikl ve Maiwald, 2024).

Sanoar Hossain ve arkadaşları (2024), derin kuantum evrişimli sinir ağı kullanarak hızlı ve verimli bir yüz ifadesi tanıma sistemi geliştirmiştir. Bu yaklaşım, özellikle büyük veri setlerinde hızlı eğitim sağlamaktadır. Ancak, bu çalışma ortalama duygu yoğunluğu yerine bireysel yüz ifadeleri üzerine odaklanmıştır ve kuantum hesaplama gereksinimleri geniş çapta uygulanabilirliği sınırlandırabilir (Hossain ve diğ., 2024).



3. MATERYAL VE YÖNTEMLER

3.1. Kullanılan Veri Setleri

3.1.1. FER-2013 Veri Seti

Bu veri setleri, farklı yaş grupları, cinsiyetler, kültürel geçmişler ve gerçek dünya koşullarındaki çeşitlilikle zenginleştirilmiştir. Her biri, temel duyguları sınıflandırmak veya karmaşık ifadeleri analiz etmek amacıyla farklı özelliklere sahip veriler sunmaktadır.

2013 Yüz İfadesi Tanıma veri seti (FER-2013), Kaggle tarafından sağlanan ve 2013 yılında Uluslararası Makine Öğrenimi Konferansı'nda (ICML) Pierre-Luc Carrier ve Aaron Courville tarafından tanıtılmış bir veri setidir. Bu veri seti her bir kare görüntü üzerinden bir CNN modeli eğitmek için sayı olarak en fazla görüntüye sahip veri setidir.

FER-2013 veri seti, farklı yaş gruplarından insanların yüz ifadelerini içeren statik görüntülerin yanı sıra eğitim ve test verilerini kapsar. Kaggle web sitesinde bulunan bu veri seti, 48x48 piksel çözünürlüklü gri tonlamalı görüntülerden oluşur. Google Görseller aracılığıyla toplanmış, 35.887 görüntü içeren geniş bir veri setidir. Aynı zamanda 7 farklı duygu sınıfına aittir. Ancak düşük çözünürlük ve ifadelerin homojenliği gibi sınırlamaları vardır.

Duygu etiketleri arasında Kızgın, Üzgün, Mutlu, Nötr, İğrenme, Korku ve Sürpriz yer almaktadır. FER-2013 veri seti, duygu sınıflandırması için diğer veri setlerine göre daha geniş bir çeşitlilik sunar. Bu veri setinde farklı yaş gruplarından insanların yüz ifadelerinin yanı sıra düşük kontrastlı görüntüler ve insanların aksesuarlar taktığı bazı örnekler de bulunmaktadır.

Ayrıca yüz içermeyen bazı görüntüler ve bazılarında metin bulunan yüzler de mevcuttur (Kumar ve Tummala, 2023). Şekil 3.1'de bu veri seri içerisinde seçilmiş bazı örnek görüntüler mevcuttur.

Şekil 3.2'de bu veri setinin yüzde 80'e 20 oranında ayrılmış doğrulama ve eğitim veri seti dağılımı gösterilmiştir.



Şekil 3.1. FER-2013 veri setine ait temel duygulara örnek görüntüler (Kınlı, 2018)

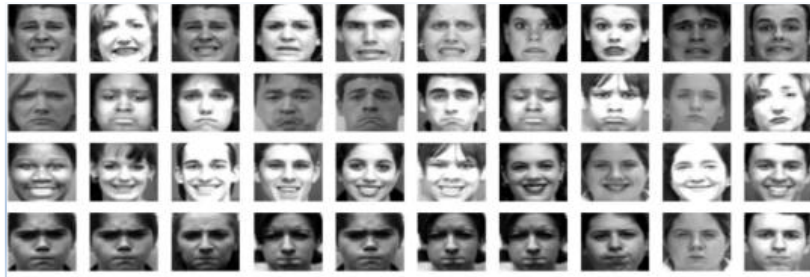


Şekil 3.2. FER-2013 Doğrulama ve eğitim veri seti örnek dağılımı

3.1.2. CK+48 Veri Seti

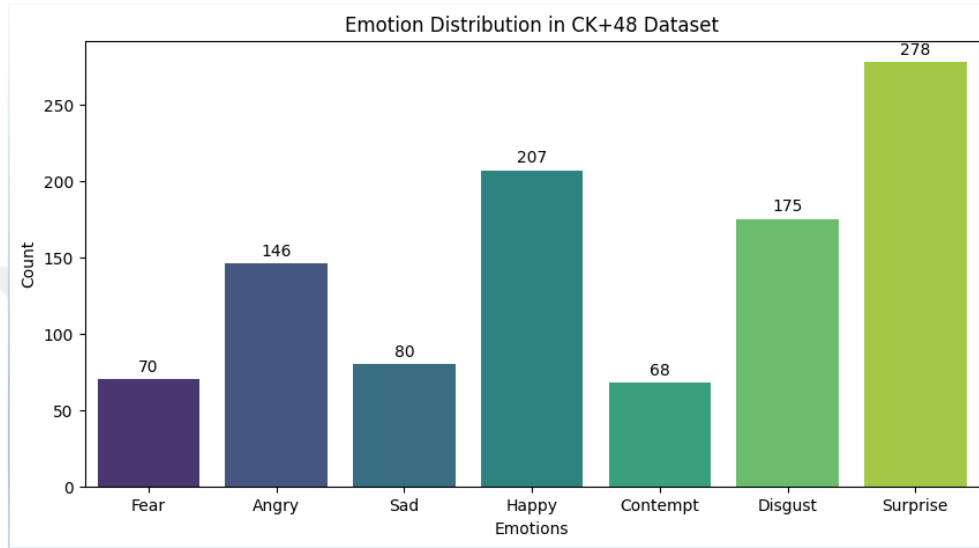
CK+48, 981 görüntüden oluşan küçük bir veri kümesidir ve yedi farklı duygu sınıfını içerir: korku, iğrenme, üzülmeye, mutluluk, nötr, sürpriz ve kızgınlık. Bu görüntüler, 48×48 piksel boyutlarında ve gri tonlamalıdır, yani renk içermemektedir. Görüntüler genellikle video karelerinden alınmıştır, bu yüzden veri kümesindeki varyasyonlar sınırlıdır ve diğer veri kümeleriyle karşılaştırıldığında boyutu oldukça küçüktür.

FER-2013 ile karşılaştırıldığında, CK+48 veri kümesindeki görüntüler yüz ifadelerini daha net ve düzenli bir şekilde temsil etmektedir. Şekil 3.3'de bu veri seti içerisinde seçilmiş bazı örnek görüntüler mevcuttur.



Şekil 3.3. CK+48 veri setine ait temel duygulara örnek görüntüler

CK+48 veri kümesi, 1000'den az görüntü içerir ve FER-2013 gibi eğitim ve doğrulama kümelerine ayrılmamıştır. Şekil 3.4, CK+48 veri kümesindeki duyguların adetleri ve dağılımı mevcuttur. Tiksizlik sınıfına ait örnek bulunmazken, aşağılanma sınıfının örnekleri de daha azdır. Görüntüler PNG formatındadır ve piksel değerleri 0 ile 255 arasında değişmektedir (Chaudhari ve diğ., 2022).



Şekil 3.4. CK+48 Doğrulama ve eğitim veri seti örnek dağılımı

3.2. Yüz Tespiti

Yüz tespiti, duygu yoğunluğu analiz sistemlerinin kritik bir bileşeni olup, görüntülerdeki yüzlerin doğru şekilde tanımlanması, sonraki adımlarda gerçekleştirilecek analizlerin başarısını etkiler.

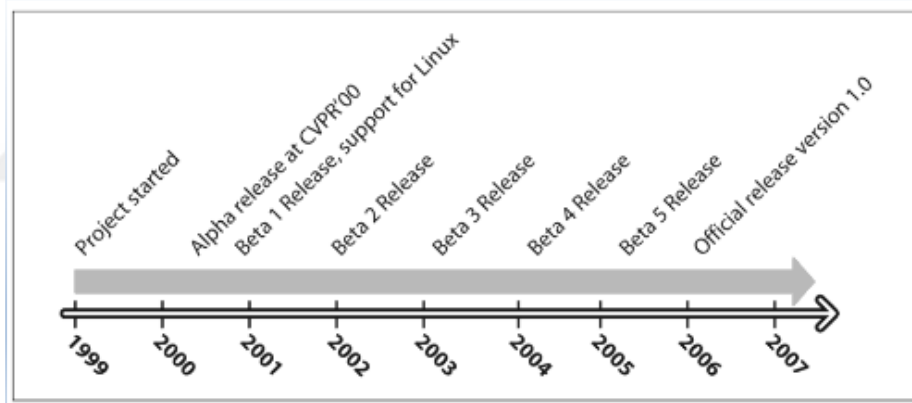
Bu çalışmada, yüz tespiti için iki ana algoritma kullanılmıştır: OpenCV ve MTCNN. OpenCV, görüntü ve video analizi için optimize edilmiş geniş bir algoritma seti sunan, yaygın kullanılan bir kütüphanedir. Çok çekirdekli işlemci desteği ve esnek yapısı sayesinde gerçek zamanlı yüz tespitinde yüksek hız ve verimlilik sağlamıştır.

MTCNN ise, yüz ve yüzdeki özellik noktalarını belirleyen üç aşamalı bir evrişimli sinir ağı yapısına sahip olup, yüksek doğruluk oranıyla duygu analizlerine önemli katkı sağlamaktadır. Bu algoritmalar, yüz tespiti süreçlerinde yüksek performans göstererek duygu yoğunluğu analizi çalışmalarında etkili sonuçlar elde edilmesine yardımcı olmuştur.

3.2.1. OpenCV

OpenCV, 1999 yılında yayınlandığından beri görüntü ve video analizi için 500'den fazla optimize edilmiş algoritma sunan açık kaynaklı bir bilgisayar görüşü kütüphanesidir. Fabrika ürün denetimi, tıbbi görüntüleme, güvenlik, kullanıcı arayüzleri, kamera kalibrasyonu, stereo görüş ve robotik gibi çeşitli alanlarda yaygın olarak kullanılmaktadır.

Kütüphane, optimize edilmiş C dilinde yazılmış olup çok çekirdekli işlemcilerden tam verim almayı hedeflemektedir. İlk sürümü 2006 yılında çıkan OpenCV 1.0, 2009'da önemli değişikliklerle birlikte C++ arayüzü ile OpenCV 2.0 olarak yayımlandı. Yeni arayüz, görüntü işleme işlevselliğini kodlarken gereksiz kod satırlarını azaltmayı ve OpenCV kullanırken yaygın olarak karşılaşılan hafıza sızıntıları gibi programlama hatalarını otomatik veri tahsisi ve serbest bırakma ile azaltmayı amaçlamaktadır (Matuska ve diğ., 2012).



Şekil 3.5. Open CV sürüm zaman çizelgesi (Bradski ve Kaehler, 2008)

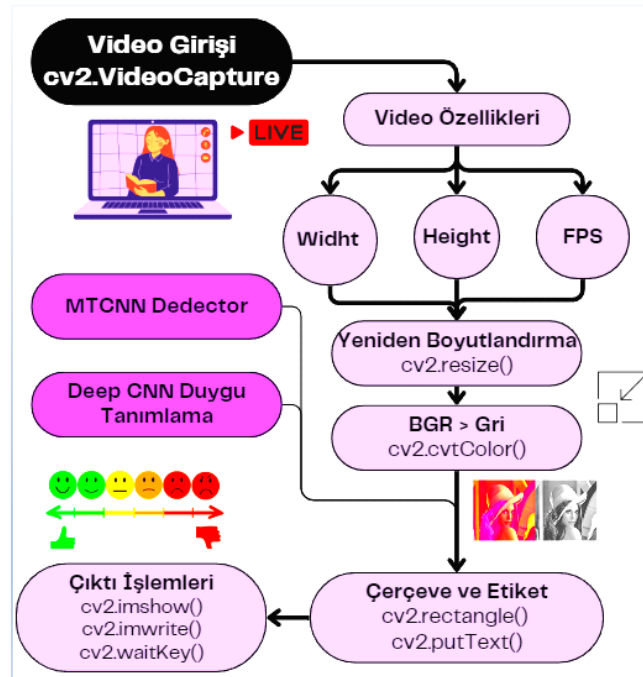
OpenCV, Python ile tam uyumlu çalışarak, NumPy gibi güçlü matematik kütüphaneleriyle entegre bir kullanım sağlar. Python'ın TensorFlow ve Keras gibi derin öğrenme kütüphaneleriyle kolay entegrasyonu, karmaşık görüntü işleme görevlerini basitleştirir. Ayrıca, Python'ın okunabilir ve kullanıcı dostu syntax'ı, algoritmaların uygulanmasını hızlandırır ve geliştirme sürecini kolaylaştırır.

OpenCV, açık kaynaklı ve ücretsiz bir kütüphane olarak endüstride geniş bir kullanım alanına sahiptir. Sahip olduğu büyük geliştirici topluluğu sayesinde bol miktarda dokümantasyon ve örnek kod sunar. Sürekli güncellenen yapısı, OpenCV'nin güncel ihtiyaçlara uyum sağlamasını ve yenilikçi çözümler sunmasını mümkün kılar.

OpenCV, yüz tespiti ve duygu analizi projelerinde birçok avantaj sağlar;

- Yüz tespiti ve duygu analizi için gerekli görüntü ön işleme yetenekleri.
- Video frame'lerini gerçek zamanlı işleyebilme kapasitesi.
- Görselleştirme araçları sayesinde analiz sonuçlarını anlık olarak gösterebilme.
- Frame'ler üzerine metin ve şekil ekleme kolaylığı.

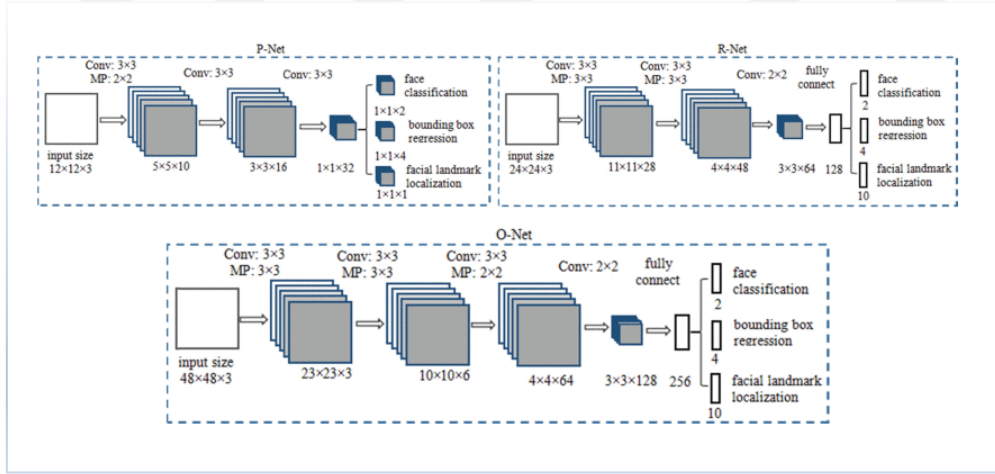
Yukarıdaki elde edilen bilgiler ışığında çalışmamızda Open CV teknolojisi kullanılarak hem gerçek zamanlı duygu tespitinde hem de duygu yoğunluğu analizi işlemlerinde verimli bir performans elde etmek amaçlanmıştır. Öncelikle video verileri cv2.VideoCapture() ile elde edilmiş ve cv2.VideoWriter() aracılığıyla kaydedilmiştir. Görüntü ön işleme aşamasında, hesaplama verimliliğini optimize etmek amacıyla cv2.resize() ile boyutlandırma ve cv2.cvtColor() ile gri tonlama dönüşümü uygulanmıştır. Tespit edilen yüz bölgeleri cv2.rectangle() fonksiyonu ile çerçevelenmiş ve duygu sınıflandırma sonuçları cv2.putText() kullanılarak görselleştirilmiştir. Sistemin çıktıları, cv2.imshow() ve cv2.imwrite() fonksiyonları vasıtasıyla hem gerçek zamanlı görüntüleme hem de kalıcı depolama için kaydedilmiştir. Aşağıda Şekilde 3.6. Open CV kullanım yöntemi şemasında pembe renk ile Open CV'yi sistemimizdeki kullanımı görselleştirilmiştir.



3.2.2. MTCNN Algoritması

MTCNN, yüzleri algılama ve yüzdeki özellik noktalarını (örneğin, gözler, burun, ağız gibi) tespit etme görevlerini bir arada gerçekleştirir. Bu algoritma, üç farklı evrişimli sinir ağı aşamasından oluşur ve her bir aşama sırasıyla yüzleri tespit etmek ve daha sonra yüzdeki özellik noktalarını belirlemek için tasarlanmıştır. İlk aşama, yüzleri tespit etmek için tasarlanmıştır ve genellikle bir resimdeki olası yüz bölgelerini belirlemek için kullanılır. İkinci aşama, belirlenen bu olası yüz bölgeleri üzerinde daha ayrıntılı bir şekilde yüz tespiti yapar. Son olarak, üçüncü aşama, yüzdeki belirli özellik noktalarını (gözler, burun, ağız gibi) tanımlamak için kullanılır.

Detaylı olarak, MTCNN, yüz tespiti ve yüzdeki belirli özellik noktalarını algılama görevlerini aynı anda gerçekleştiren bir yüz algılama algoritmasıdır. Bu algoritma, genellikle üç farklı evrişimli sinir ağı aşamasından oluşur ve her bir aşama sırasıyla yüzleri tespit etmek ve daha sonra yüzdeki özellik noktalarını belirlemek için tasarlanmıştır. Şekil 3.7’de MTCNN’nin barındırdığı üç aşamanın mimari yapısı görülmektedir.



Şekil 3.7. MTCNN Üç aşamalı mimarisi

İlk Aşama (P-Net): İlk aşama, yüzleri tespit etmek için kullanılır ve genellikle bir resimdeki olası yüz bölgelerini belirlemek için uygulanır.

Bu aşama, potansiyel yüz bölgelerini tespit etmek için bir evrişimli sinir ağıdır. Bu bölgeler daha sonra daha ayrıntılı bir inceleme için diğer aşamalara iletilir.

Bu ağ yapısı, aday pencerelerin ve yüz bölgelerinin sınır kutularının regresyon vektörlerini elde eder. Aday pencereler, bu vektörlerle kalibre edilir ve yüksek örtüşen çerçeveler NMS yöntemiyle birleştirilir. Çok ölçekli hedefler için görüntü piramitleri oluşturulur ve farklı boyutlardaki görüntüler Cascade ağlara sırasıyla gönderilir.

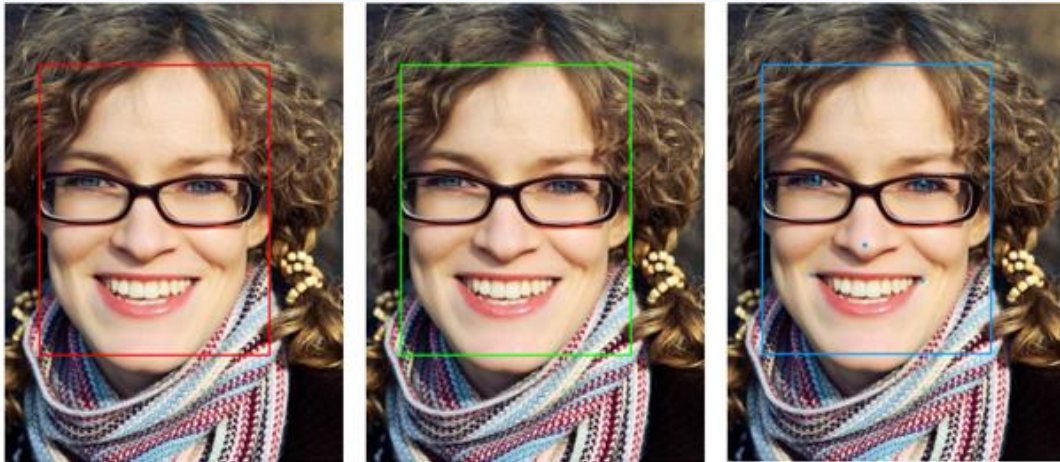
İkinci Aşama (R-Net): İkinci aşama, ilk aşamada belirlenen olası yüz bölgeleri üzerinde daha ayrıntılı bir yüz tespiti yapar. Daha spesifik özellikleri tanımlamak için kullanılır ve yüz bölgelerini daha kesin bir şekilde belirler.

P-Net ağı, aday yüz çerçevesinin boyutunu 24×24 olarak ayarlar ve çıkan sonuçlar yüz skoru ve sınır kutusu regresyonu bilgilerini içerir. P-Net'e kıyasla, bu ağ yapısı, yanlış pozitif örnekleri daha etkili bir şekilde ayıklayabilen bir tam bağlantı katmanına sahiptir.

Üçüncü Aşama (O-Net): Son aşama, belirlenen yüz bölgeleri üzerinde daha detaylı bir analiz yapar ve yüzdeki belirli özellik noktalarını (örneğin, gözler, burun, ağız gibi) tanımlamak için kullanılır. Bu aşama, yüz üzerindeki anahtar noktaları belirleyerek yüz analizini tamamlar.

Bu ağ yapısı, R-Net tarafından işlenmiş yüz aday kutularını daha da inceleyerek filtreler, giriş görüntüsünü 48×48 boyutuna getirir ve her yüz çerçevesindeki özellik noktalarının yerlerini belirler.

Anlatılan bu yapı aşağıdaki Şekil 3.8'de bulunun resim üzerinde P-Net aşaması kırmızı, R-Net aşaması yeşil ile ve O-Net aşaması ise mavi çizgiler ile gösterilmiştir.



Şekil 3.8. MTCNN Yüz bölgesi belirleme

Non-Maximum Suppression (NMS), MTCNN'de kullanılan bir teknik olup, aynı anda birden fazla yüz tespiti yapılabileceği durumlarda yalnızca en doğru tespitleri seçmek için kullanılır.

NMS, aynı bölgede birden fazla bounding box (yüz kutusu) tespit edildiyse, her birinin confidence score değerine göre yalnızca en yüksek skora sahip olan kutuyu alır ve diğerlerini eler (Xiao ve diğ., 2021).

Literatür çalışmamızda da kapsamlı bir şekilde incelediğimiz üzere, yüz tespiti konusunda yapılan çalışmalar yıllar içinde önemli bir evrim geçirmiştir. İlk olarak, Viola ve Jones'un seminal çalışması, yüz tespitinde devrim yaratmış ve gerçek zamanlı performansı ile dikkat çekmiştir. Bu yöntem, el yapımı Haar özelliklerini kullanarak farklı ölçeklerdeki görüntü piramitleri üzerinde çalışmış ve sınıflandırmayı hızlandırmıştır. Ancak, frontal olmayan yüzlerde ve düşük ışık koşullarında yetersiz performans sergilemesi, bu yöntemin sınırlı uygulama alanına sahip olmasına yol açmıştır. Bu eksikliklerin ardından, literatürde birçok iyileştirme ve alternatif yöntem geliştirilmiştir.

Modern derin öğrenme yöntemlerinin yüz tespitinde kullanılmaya başlanmasıyla birlikte, Convolutional Neural Networks (CNN) tabanlı yaklaşımlar önemli bir ilerleme sağlamıştır. CNN'ler, klasik yöntemlerden farklı olarak öğrenilmiş özellikler kullanarak daha genel ve güçlü bir performans sunmaktadır. Örneğin, SSD (Single Shot Detector) gibi tek aşamalı dedektörler, yüz tespiti sürecini basitleştirerek yüksek doğruluk oranlarına ulaşmıştır. Ancak, SSD ve benzeri yöntemler, karmaşık sahnelerde doğruluk kaybı yaşayabilmekte ve bazı durumlarda belirli nesneler arasında karışıklık yaratabilmektedir.

Çok aşamalı dedektörlere baktığımızda, R-CNN gibi yöntemler bölge önerisi ve iyileştirme aşamalarıyla daha kesin sonuçlar sunmuştur. Ancak bu yöntemler, işlem süresi açısından yoğun kaynak tüketimi gerektirmekte ve bu durum, gerçek zamanlı uygulamalar için bir engel oluşturmaktadır (Kaziakhmedov ve diğ., 2019).

Çalışmamızda, yukarıda bahsedilen avantaj ve dezavantajları değerlendirerek, yüz tespiti için MTCNN'yi tercih ettik. MTCNN, çok aşamalı yapısı sayesinde hem yüksek doğruluk

oranlarına ulaşmakta hem de işlem süresi açısından verimlilik sunmaktadır. Literatürde de yaygın olarak kabul gördüğü üzere, MTCNN'nin P-Net, R-Net ve O-Net olmak üzere üç alt ağdan oluşan mimarisi hem yüz tespiti hem de yüz landmark çıkarımı için entegre bir çözüm sağlamaktadır. İlk aşamada kullanılan P-Net, 12x12 boyutundaki reseptif alanı sayesinde küçük bozulmalara dayanıklıdır ve Non-Maximum Suppression (NMS) algoritması, örtüşen kutuları azaltarak daha kesin bölgeler elde etmektedir. R-Net ve O-Net ise sırasıyla bu bölgeleri iyileştirip landmark çıkarımı ile tespiti sonlandırmaktadır.



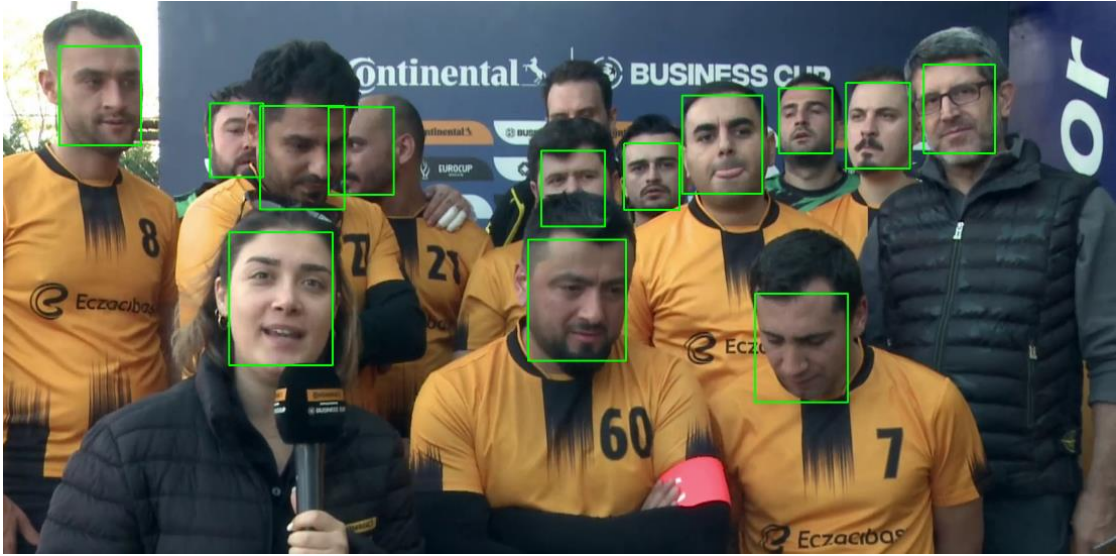
Şekil 3.9. MTCNN ile yüz tespiti süreci

Ayrıca, literatürdeki başarıları ve FaceNet gibi tanınmış yüz tanıma sistemlerinde kullanılıyor olması, MTCNN'nin güvenilirliğini ve pratik uygulanabilirliğini desteklemektedir. SSD ve R-CNN gibi yöntemlere kıyasla, MTCNN'nin hem doğruluk hem de hız açısından sağladığı denge, bizim çalışmamız için ideal bir çözüm sunmaktadır. Bu bağlamda, MTCNN'yi çalışmamızda kullanarak yüksek doğruluk ve dayanıklılık gerektiren uygulamalar için güçlü bir temel oluşturduk.

Şekil 3.9'da MTCNN dedektörünün OpenCV den elde edile kareler ile birlikte yüz tespiti iş akışı şematize edilmiştir. Aşağıda internet ortamından bulunun videolar ile MTCNN dedektörü test edilmiştir. Gözle ayırt edilebilecek tüm yüzlerin tespitinde başarılı bir performans sergilendiği görülmektedir. Çalışmamızda kullanılmasına bu performansının

yanı sıra gelişmiş yüz tanıma uygulamalarında kullanılıyor olması tercih etmemizde etken olmuştur.

Şekil 3.10’da bir futbol maçından sonra yapılan takım röportajı üzerinden MTCNN dedektörünün yüz tespiti çalışması yapılmıştır. Şekil 3.11’de bir mezuniyet töreninde daha küçük yüzler içeren video üzerinde yüz tespiti yapılmıştır. Şekil 3.12’de ise bir mezuniyet töreni tanıtım resminde yüz tespiti çalışması yapılmıştır.



Şekil 3.10. MTCNN ile kalabalık ortam yüz tespiti test çalışması



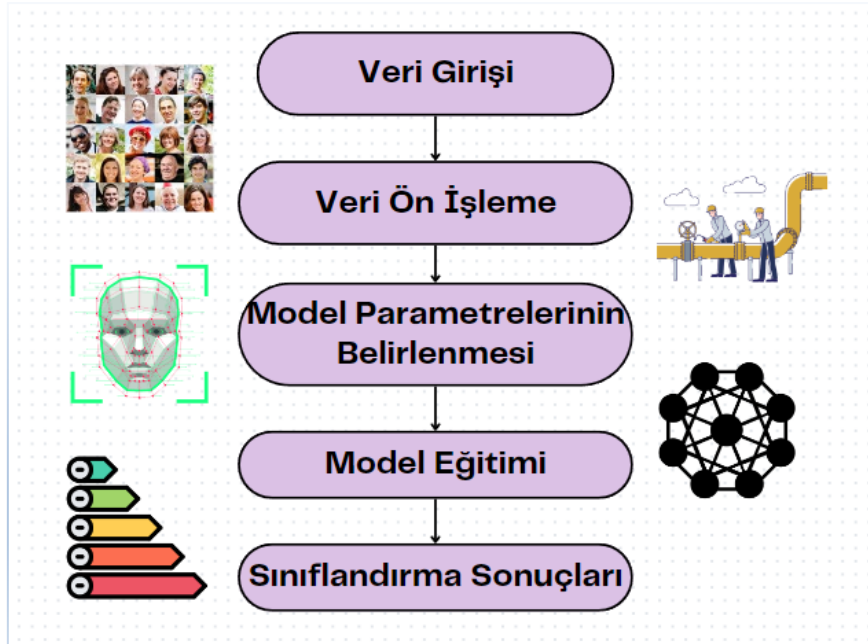
Şekil 3.11. MTCNN ile kalabalık ortam yüz tespiti test çalışması-2



Şekil 3.12. MTCNN ile kalabalık ortam yüz tespiti test çalışması-3

3.3. Duygu Sınıflandırma Model Eğitimi

Çalışmamız kapsamında sınıflandırma görevine yönelik olarak hem güncel teknolojiler hem de sık kullanılan geleneksel özellik çıkarıcılar incelenmiş ve uygulanmıştır. Sınıflandırma sürecinde izlenen genel iş akışı, veri setinin hazırlanmasıyla başlamakta ve şu aşamalardan oluşmaktadır: veri ön işleme, model mimarisi ve parametrelerinin belirlenmesi, modelin eğitilmesi ve sınıflandırma sonuçlarının detaylı bir şekilde analiz edilmesi. Bu yapılandırılmış süreç, sınıflandırma modellerinin etkinliğini ve genel başarımını değerlendirmek için sistematik bir temel sağlamaktadır. Şekil 3.13'te model eğitim işlem hattına ait genel şema gösterilmektedir.

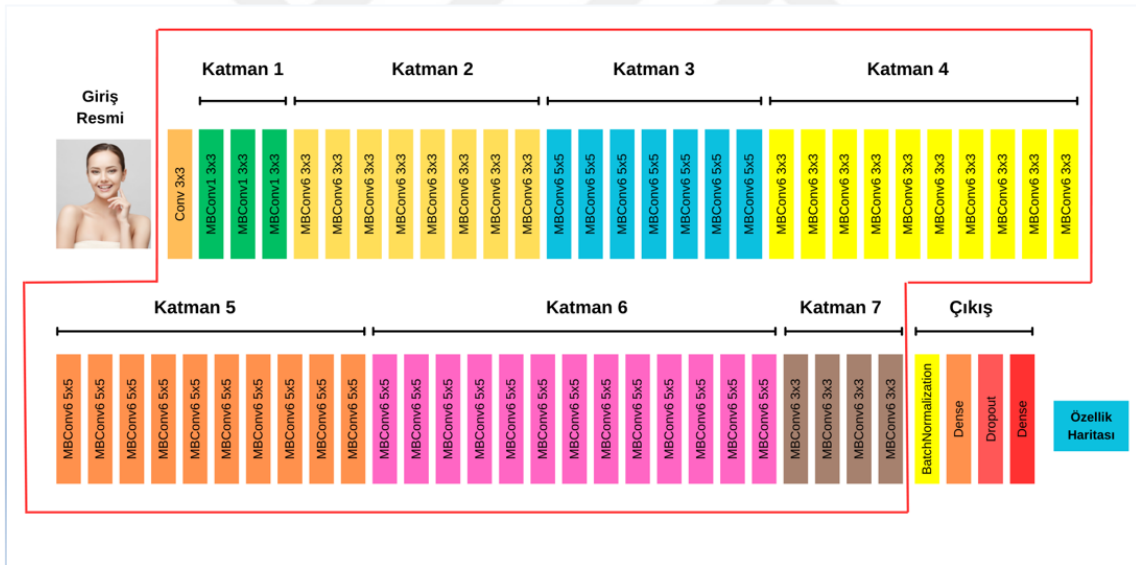


Şekil 3.13. Modellerin genel eğitim aşaması

3.3.1. EfficientNet-B7

EfficientNet, derin öğrenme alanında yüksek doğruluk ve verimlilik sağlamak için geliştirilmiş bir konvolüsyonel sinir ağı (CNN) ailesidir. EfficientNet'in temel yeniliği, modelin ölçeklenmesi sürecini optimize eden bir bileşen ölçekleme yöntemidir. Bu yöntem, modelin derinliğini (katman sayısı), genişliğini (kanal sayısı) ve çözünürlüğünü (giriş görüntü çözünürlüğü) orantılı olarak artırır.

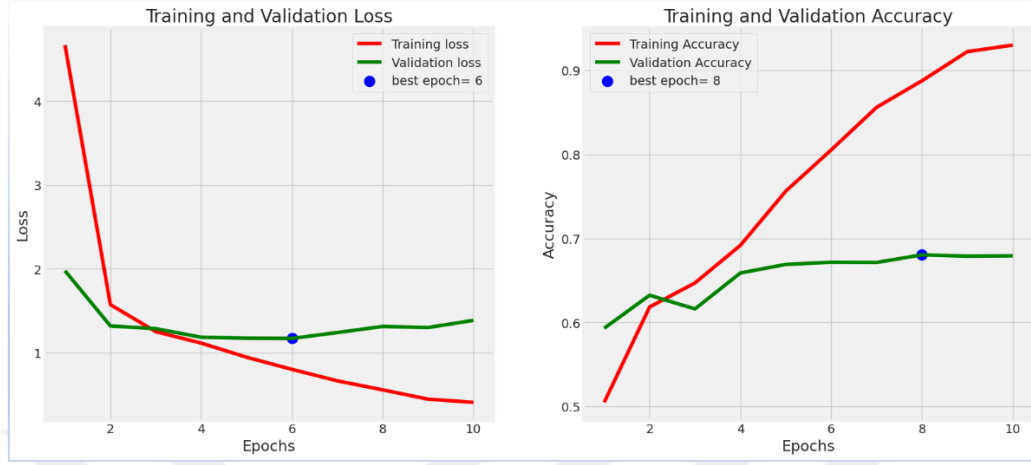
EfficientNet-B7, yüksek doğruluk ve verimlilik sağlayarak görüntü sınıflandırma görevlerinde üstün performans gösterir. Verimlilik: Hesaplama maliyeti ve doğruluk arasında etkili bir denge sağlar. Ölçeklenebilirlik: Bileşen ölçekleme yönteminin etkinliğini gösterir ve artan hesaplama kaynakları ile yüksek doğruluk elde eder (Tan ve Le, 2019). Şekil 3.14'te kullanılan EfficientNet-B7 ve çıkış katmanı mimarisi görülmektedir. Kullanılan modelin teknik detayları EK-A'da açıklanmıştır.



Şekil 3.14. EfficientNet-B7 Katman mimarisi

EfficientNet-B7 aslında mobil cihazlarla uyumlu çalışabilecek bir model olarak geliştirildi fakat eğitim süresi en uzun süren modellerden biri oldu. Resim boyutlarını ilk olarak 224*224 olarak kullandığımız model için 10 epochs ile eğitim gerçekleştirdik. En iyi doğrulama kaybı için 6. epoch olduğunu gördük. En iyi doğrulama doğruluğu için ise 8. Epoch olarak belirlendi. Tüm duygu durumları için tahminleme ve skorlar hesaplanırken modelin başarılı olduğu görülmüştür.

Duygu tespiti modellimiz için tercih ettiğimiz modellerden biri olan EfficientNet-B7 modelini eğitmek için FER-2013 veri setini kullandık.



Şekil 3.15. EfficientNet-B7 FER-2013 sonuç grafiği

Şekil 3.15’de modelimizin her epochta gösterdiği kayıp ve doğruluk oranları gösterilmiştir.

FER-2013’te elde edilen sonuçlar, veri setinin zorluğunu yansıtır. Modelin derin yapısı, özellikle belirgin duygu sınıflarını tanımada başarılı olmuştur. Ancak, veri setindeki dengesizlikler ve zorluklar, daha az belirgin sınıflarda (örneğin, korku, üzüntü) modelin zorlanmasına neden olmuştur. Bu nedenle, modelin genel performansını artırmak için veri artırma (data augmentation) ve sınıf ağırlıkları gibi stratejiler düşünülebilir.

Tablo 3.1. EfficientNet-B7 FER-2013 veri seti ile sınıflandırma sonuçları

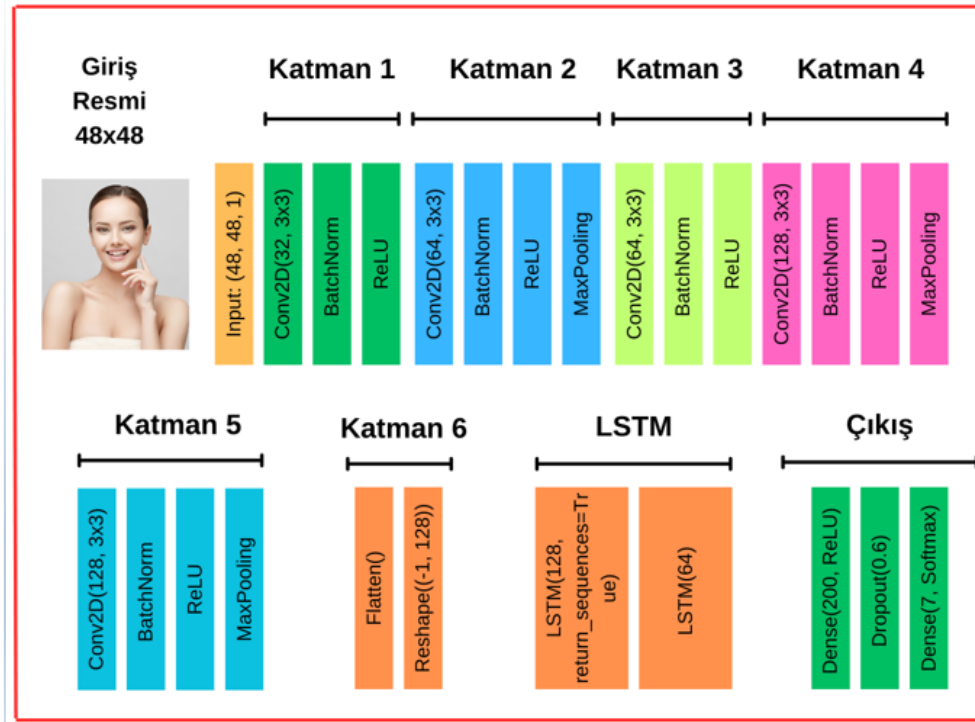
Duygu	Kesinlik	Duyarlılık	F1-Skoru	Destek Sayısı
Öfke	0.61	0.65	0.63	400
Tiksinti	0.82	0.74	0.78	42
Korku	0.52	0.55	0.53	420
Mutluluk	0.92	0.85	0.88	709
Doğal	0.61	0.71	0.65	505
Üzüntü	0.58	0.45	0.51	476
Şaşkınlık	0.77	0.84	0.80	320
Genel			0.68	2872
Doğruluk				
Ortalama	0.69	0.68	0.68	2872
Ağırlıklı	0.69	0.68	0.68	2872
Ortalama				

Tablo 3.1’de eğitilen modelin test veri seti ile modelin sınıflandırma sonucu performansı her duygu sınıfı için gösterilmiştir. Sonuçlar modelimizin eğitim sırasında ortalama bir performans gösterdiğini belirtmiştir.

3.3.2. ConvLSTM

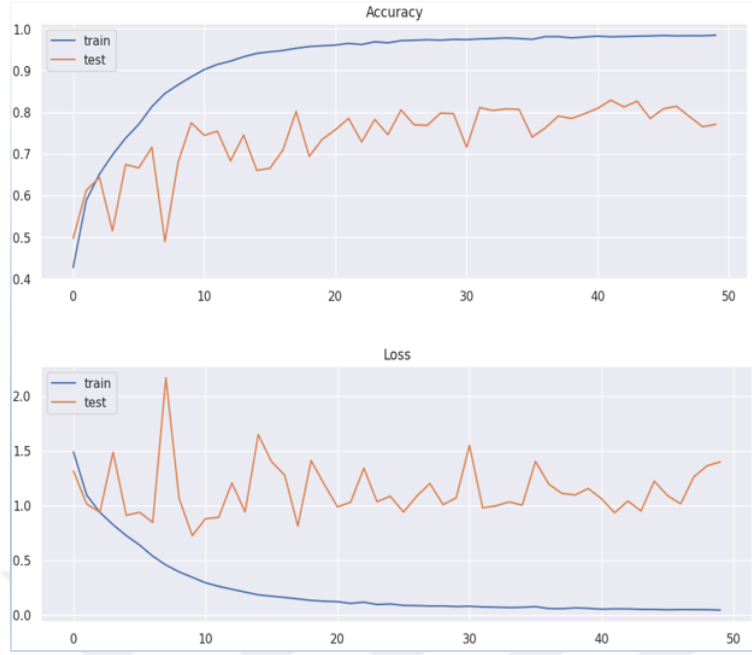
Yüz ifadelerinin sınıflandırılmasında, ConvLSTM (Convolutional Long Short-Term Memory) tabanlı bir sinir ağı modeli geliştirilmiştir. Modelin yapısında birden fazla ConvLSTM katmanı, Batch Normalization, MaxPooling ve Dropout katmanları kullanılmıştır. Son katmanlarda, TimeDistributed bir Flatten ve Dense katmanı eklenerek duygu sınıfları arasında sınıflandırma yapılmıştır.

Eğitim sürecinde, kategorik çapraz entropi kaybı fonksiyonu ve Adam optimizasyon algoritması kullanılarak modelin öğrenme kapasitesi optimize edilmiştir. Şekil 3.16’da kullanılan model mimarisi şekillendirilmiştir.



Şekil 3.16. ConvLSTM Katman mimarisi

Modelin performansı, eğitim ve doğrulama setleri üzerindeki kayıp ve doğruluk değerleri izlenerek değerlendirilmiştir. Kullanılan modelin teknik detayları EK-A’da açıklanmıştır.



Şekil 3.17. ConvLSTM FER-2013 sonuç grafiği

Şekil 3.17’de eğitilen modelimizin doğruluk ve kayıp değerleri gösterilmektedir. Modelin çıktılarını göre genelleme sorunları yaşanmasına rağmen gerçek zamanlı testlerinde başarılı sonuçlar verdiği gözlemlendi.

Tablo 3.2. ConvLSTM FER-2013 veri seti ile sınıflandırma sonuçları

Duygu	Kesinlik	Duyarlılık	F1-Skoru	Destek Sayısı
Öfke	0.89	0.64	0.74	1770
Tiksinti	0.85	0.75	0.79	1872
Korku	0.86	0.64	0.74	1775
Mutluluk	0.88	0.95	0.91	1755
Doğal	0.81	0.52	0.63	1825
Üzüntü	0.97	0.99	0.98	1778
Şaşkınlık	0.48	0.92	0.63	1810
Genel			0.77	12585
Doğruluk				
Ortalama	0.82	0.77	0.78	12585
Ağırlıklı	0.82	0.77	0.77	12585
Ortalama				

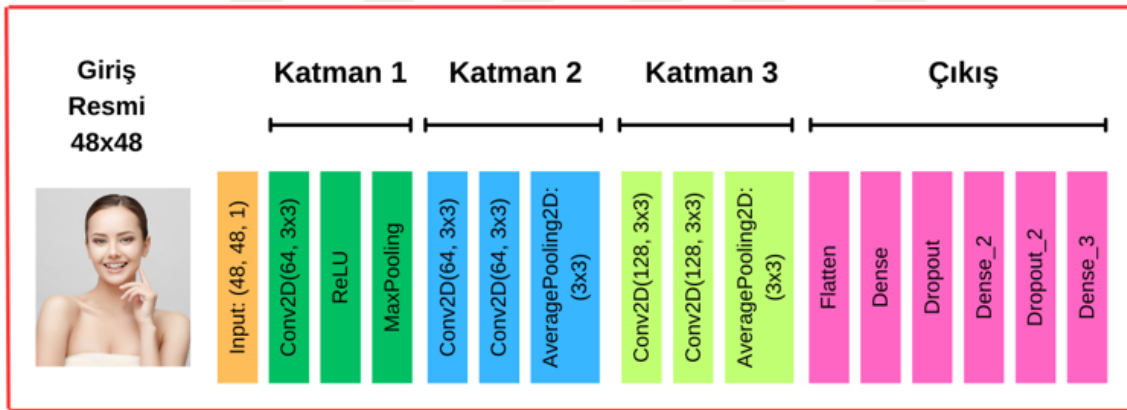
ConvLSTM modeli, gerçek dünya verilerini iyi öğrenmeye çalışsa da bazı sınıflarda zorluk yaşıyor. Tablo 3.2’de de görüldüğü üzere modelimizin özellikle üzüntü ve mutluluk sınıflarında yüksek başarı oranı göstermiştir. FER-2013’te modelin %77

doğruluk elde etmesi, modelin karmaşık yapılar içeren verilerde umut verici bir potansiyele sahip olduğunu gösteriyor.

3.3.3. CNN

Duygu tespitinde kullanılan modellerin doğruluğu ve genelleme kabiliyeti, doğrudan modelin mimarisine ve öğrenme kapasitesine bağlıdır. Bu bağlamda, derin katmanlı bir Convolutional Neural Network (CNN) modeli geliştirmek, karmaşık duygu ifadelerini anlamlandırmada önemli avantajlar sunar. Modelimin tasarımında, duygu tespitinin doğasında bulunan çok boyutlu ve incelikli özelliklerin yakalanması hedeflenmiştir.

CNN yapısı, özellikle yüz ifadelerinde mikro ifadeler ve ince farklılıkları öğrenme kapasitesi ile öne çıkar. Daha fazla katman ekleyerek, modelimin daha soyut ve yüksek seviyeli özellikleri temsil edebilmesini sağladık. Bu derin yapı, girişteki piksel düzeyindeki bilgileri alıp giderek daha karmaşık duygu temsillerine dönüştürebilir. Aşağıda Şekil 3.18’de geliştirilen modelin katmanları gösterilmiştir.



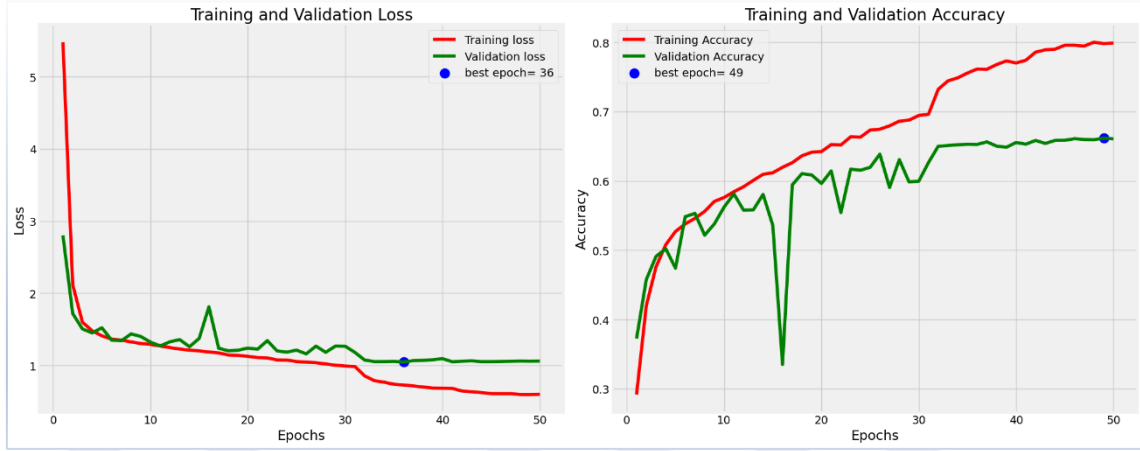
Şekil 3.18. CNN Katman mimarisi

Modelimiz, sırasıyla evrişimli katmanlar, havuzlama katmanları, düzleştirme katmanı ve tam bağlantılı katmanlardan oluşur. Model, girdi olarak 48x48 boyutunda gri tonlamalı yüz görüntüleri alır ve 7 farklı yüz ifadesini tanımak üzere yapılandırılmıştır.

Conv2D katmanları, görüntü özelliklerini çıkarır ve MaxPooling2D/AveragePooling2D katmanları, özellik haritalarının boyutunu azaltır.

Dense katmanlar, sınıflandırma için kullanılır ve Dropout katmanları, aşırı uyumu önlemeye yardımcı olur. Modelin amacı, girdiyi (yüz görüntüsü) işleyerek çıktı olarak 7

farklı yüz ifadesini (Sinirli, İğrenme, Korku, Mutlu, Üzgün, Şaşkın, Doğal) sınıflandırmaktadır.



Şekil 3.19. CNN Modeli FER-2013 sonuç grafiği

Şekil 3.19’da eğitilen modelimizin doğruluk ve kayıp değerleri gösterilmektedir. Yapılan doğrulama sınıflandırması sonucunda yüzde 85 doğruluklara ulaşan modelimiz gerçek zamanlı uygulamalar içinde uygun olduğunu göstermiştir.

Tablo 3.3’de de görülen modelimizin test veri seti ile yapılan sınıflandırma sonucunda mutluluk ve diğer sınıflarda da gayet yüksek oranlar gözlemlenmiştir.

Tablo 3.3. CNN FER-2013 veri seti ile sınıflandırma sonuçları

Duygu	Kesinlik	Duyarlılık	F1-Skoru	Destek Sayısı
Öfke	0.84	0.82	0.83	1770
Tiksinti	0.88	0.85	0.86	1872
Korku	0.82	0.80	0.81	1775
Mutluluk	0.92	0.94	0.93	1755
Doğal	0.83	0.85	0.84	1825
Üzüntü	0.80	0.84	0.82	1778
Şaşkınlık	0.90	0.89	0.84	1810
Genel			0.85	12585
Doğruluk				
Ortalama	0.86	0.86	0.86	12585
Ağırlıklı	0.85	0.85	0.85	12585
Ortalama				

Aşırı öğrenme veya eğitilememe gibi bir sorun yaşamadan ağırlıkları kaydedilmiştir. Kaydedilen model daha sonra gerçek zamanlı testlerimizde kullanılmıştır.

3.4. Model Parametrelerinin ve Sonuçlarının Değerlendirilmesi

EfficientNetB7, ConvLSTM ve CNN gibi üç farklı derin öğrenme modelinin performansını ve hesaplama gereksinimlerini karşılaştırılmıştır. Bu hesaplamaların elde edilmesi Tensorflow çatısı altında bulunan fonksiyonlar kullanılmıştır.

Çalışmada kullanılan modeller arasında, CNN modeli %85 doğruluk oranıyla en yüksek performansı göstermiştir. ConvLSTM modeli ise %82 doğruluk oranıyla CNN'den biraz daha düşük bir performans sergilemiştir. Öte yandan, EfficientNetB7 modeli %69 doğruluk oranıyla en düşük performansa sahip model olmuştur. Bu sonuçlar, CNN ve ConvLSTM modellerinin veri kümesi üzerinde daha iyi genelleme yapabildiğini göstermektedir. Buna karşılık, EfficientNetB7'nin bu veri kümesi veya problem için yeterince etkili çalışmadığı düşünülmektedir. Bu durum, modelin mimarisinin veya eğitim sürecinin veri sayısının uygun olmadığını düşündürmektedir.

Model karmaşıklığını ve eğitim sürecini değerlendirmek için parametre sayıları ve güncellenen parametreler incelenmiştir. EfficientNetB7, 64,4 milyon parametre ile en büyük model olmasına rağmen, yalnızca 15 bin 120 parametre güncellenmiştir. Bu durum, modelin büyük bir kısmının eğitim sürecinde aktif olarak kullanılmadığını ve dolayısıyla eğitim verimliliğinin düşük olabileceğini göstermektedir. ConvLSTM modeli, 474 bin parametre ile daha küçük bir model olmasına rağmen, 39 bin 300 parametre güncellemesiyle daha etkin bir öğrenme süreci sergilemiştir. CNN modeli ise 1,48 milyon parametre ile EfficientNetB7'den daha küçük, ancak ConvLSTM'den daha büyük bir modeldir ve 39 bin 300 parametre güncellemesiyle etkin bir öğrenme performansı göstermiştir. Bu bulgular, ConvLSTM ve CNN modellerinin daha verimli bir şekilde öğrenebildiğini, EfficientNetB7'nin ise büyük boyutuna rağmen eğitim sürecinde daha az etkili olduğunu ortaya koymaktadır.

Flip-Flop sayısı, modelin içindeki bellek öğelerinin değişim sıklığını gösterir ve donanım üzerindeki iş yükünü etkileyebilir. EfficientNetB7, 974,1 milyar Flip-Flop sayısı ile en yüksek değere sahipken, ConvLSTM 18 milyar ve CNN 58.5 milyar Flip-Flop sayısı ile daha düşük değerlere sahiptir. Bu sonuçlar, CNN'nin donanım açısından daha hafif çalıştığını, EfficientNetB7'nin ise en yüksek donanım yüküne sahip olduğunu göstermektedir.

FLOPs (Floating Point Operations Per Second- Kayan Nokta İşlem Sayısı), modellerin hesaplama karmaşıklığını ifade eder. EfficientNetB7, 403 milyon FLOPs ile açık ara en yüksek hesaplama maliyetine sahiptir. ConvLSTM, 197 milyon FLOPs ile en düşük hesaplama maliyetine sahip modeldir. CNN ise 58,5 milyar FLOPs ile makul bir hesaplama yüküne sahiptir. Bu bulgular, CNN'nin yüksek doğruluk oranını korurken, hesaplama açısından da verimli bir performans sergilediğini göstermektedir. Öte yandan, EfficientNetB7'nin yüksek hesaplama maliyetine rağmen düşük doğruluk oranı, modelin bu problem için uygun olmadığını düşündürmektedir.

Bu çalışmada, Tablo 3.4'te de görüldüğü üzere CNN modeli hem yüksek doğruluk oranı hem de düşük hesaplama maliyeti ile öne çıkmıştır. ConvLSTM, daha düşük doğruluk oranına rağmen, düşük parametre sayısı ve hesaplama maliyetiyle verimli bir alternatif olarak değerlendirilebilir. EfficientNetB7 ise büyük boyutu ve yüksek hesaplama maliyetine rağmen, bu veri kümesi üzerinde yeterli performansı gösterememiştir. Bu sonuçlar, model seçiminde doğruluk, parametre verimliliği ve hesaplama maliyeti arasında bir denge kurulması gerektiğini vurgulamaktadır.

Tablo 3.4. Farklı derin öğrenme modellerinin performans ve hesaplama karmaşıklığı karşılaştırması

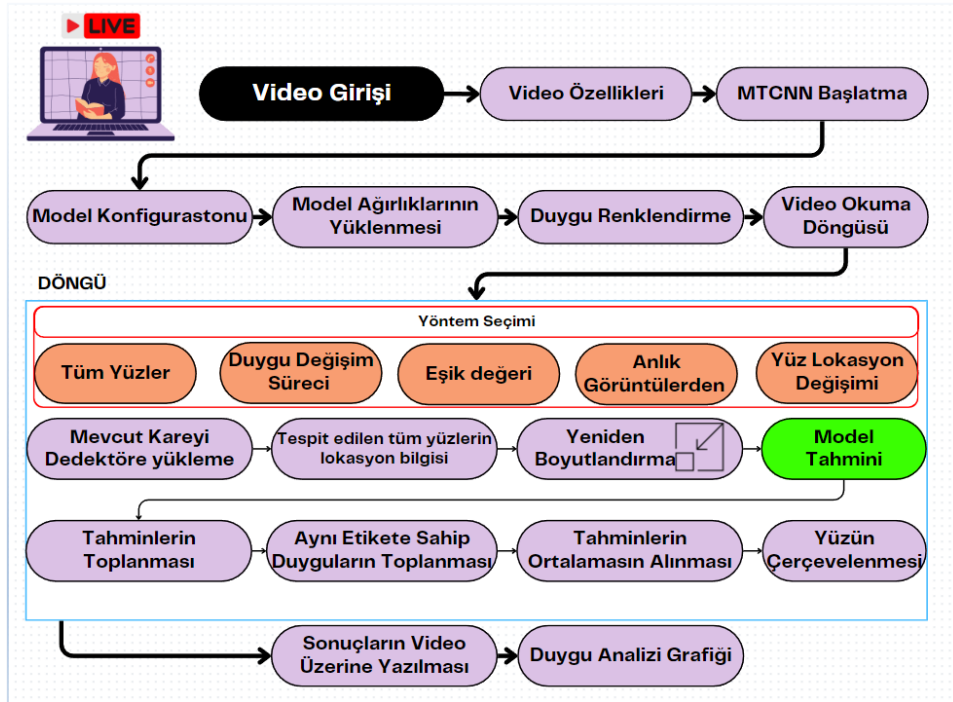
Model	Doğruluk	Parametre Sayısı	Güncellenen Parametre	Flip-Flop Sayısı	Flops Sayısı
EfficientNetB7	69	64.426.398	15.120	974.127.137.760	403.723.773
ConvLSTM	82	474.247	39.300	18.637.907.100	197.116.258
CNN	85	1.485.831	39.300	58.393.158.300	58.508.522

3.5. Ortam Duygu Yoğunluğu Analizi

Ortam duygu yoğunluğu, bir ortamda bulunan bireylerin yüz ifadelerinden elde edilen duygusal durumların toplu bir değerlendirilmesiyle ortaya çıkan genel duygusal atmosferi ifade eder. Bu kavramın analizi, bireylerin yüz ifadeleri, ses tonları veya diğer davranışsal göstergelerden toplanan verilerin işlenmesiyle gerçekleştirilir.

Çalışmamızda, bu analiz süreci, görüntü işleme ve yapay zekâ teknikleri kullanılarak bireylerin yüz ifadelerinin gerçek zamanlı incelenmesi yoluyla gerçekleştirilmiştir. Analiz sırasında, ortamda bulunan her bireyin yüz ifadeleri değerlendirilerek belirli bir zaman diliminde baskın hale gelen duyguların yoğunluğu hesaplanmış ve bu sayede

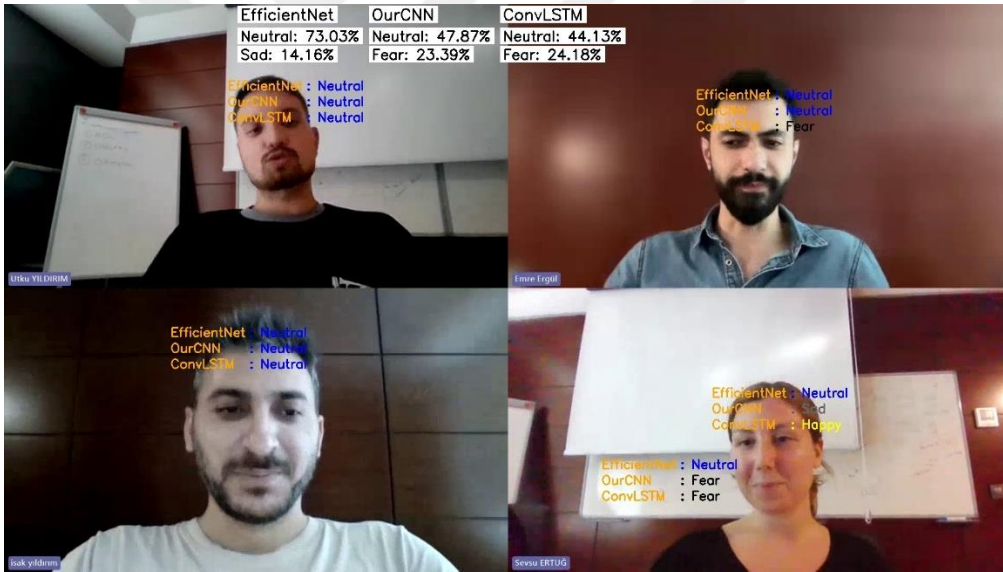
ortamın genel duygusal durumu ortaya konmuştur. Model eğitim sürecinin tamamlanmasının ardından geliştirilen tüm modeller, Şekil 3.20'de gösterilen ortam duygu analizi sisteminde çalıştırılmıştır. Aynı zamanda, bu üç model eşzamanlı olarak gerçek zamanlı ortam duygu yoğunluğu analizi testine tabi tutulmuştur. Şekil 3.20'de, önceki bölümlerde açıklanan OpenCV, MTCNN yüz tespit algoritması ve eğitilen modellerin entegre edildiği bir sistem geliştirilmiştir. Bu sistemde, ortam duygu yoğunluğunu analiz etmek amacıyla video girişinden sonra OpenCV kullanılarak her bir video karesinin işlenmesini sağlayan bir döngü oluşturulmuştur. Bu döngü kapsamında her video karesi üzerinde yüz tespiti gerçekleştirilmektedir. Yüz tespit işleminin ardından, belirlenen yüzler üzerinde duygu tanımlama işlemi yapılmakta ve elde edilen veriler belirli kriterler doğrultusunda dizilerde saklanmaktadır. Duygu tanımlama süreci tamamlandıktan sonra, her bir kare için ortam duygu yoğunluğu hesaplanarak video işleme aşamasına geçilmektedir. Tüm karelerin işlenmesi tamamlandıktan sonra, kareler yeniden birleştirilerek video formatına dönüştürülmektedir. Şekil 3.20'de turuncu ile gösterilen kareler, ortam duygu yoğunluğu hesaplanırken kullanılan farklı yöntemleri temsil etmektedir. Bu aşamada, model performanslarını değerlendirmek amacıyla tüm yüzlerin ve tüm karelerin işlendiği bir yöntem uygulanarak videolar üzerinde üç modelin performansı test edilmiştir.



Şekil 3.20. Ortam duygu yoğunluğu analizi işlem hattı

Modellerimiz ağırlıkları ile kaydederek gerçek zamanlı çalıştırabilir hale geldikten sonra her video karesinden tespit edilen tüm yüzlerin kaydedilen EfficientNet-B7, ConvLSTM ve CNN modellerimize girdi olarak verilerek üç modelinde farklı farklı çalışma performansları gözlemlenmiştir.

Şekil 3.21’de iş toplantı videosunun üzerinde üç modelimizin performansı izlenmiştir. Anlık alına bir görüntüde tüm modellerin ortamda ki genel duygu durumu için 7 temel duygu arasındaki en yüksek ilk iki tahminleri video üzerinde bir süre izlenmiştir. Bu toplantının resim üzerindeki anında görüleceği üzere ortamda doğal yüz ifadeleri daha etkindir. Üç modelimizde bu durumu ~%73, ~%45 ve ~%44 gibi oranlar ile ortam için doğru tahmin etmiştir. Tabi ki Ortam genel duygu durumu yüzde yüz olarak bir tek duyguyu temsil edemeyeceği için belirli kişiler üzerinde de etkin olan duygular modellerimiz tarafından ikinci sırada tahmin edilmiştir.



Şekil 3.21. MetropolCard toplantısı tüm model test görüntüsü

Şekil 3.22’de bir halı saha maçını kaybeden takımın maç sonun röportajı üzerinde üç modelimizin performansı izlenmiştir. Anlık alına bir görüntüde tüm modellerin ortamda ki genel duygu durumu için 7 temel duygu arasındaki en yüksek ilk iki tahminleri video üzerinde bir süre izlenmiştir. Bu toplantının resim üzerindeki anında görüleceği üzere ortamda doğal ve üzgün yüz ifadeleri daha etkindir. Üç modelimizde bu durumu ~%41 doğal, ~%38 üzgün ve ~%39 doğal gibi oranlar ile ortam için doğru tahmin etmiştir.



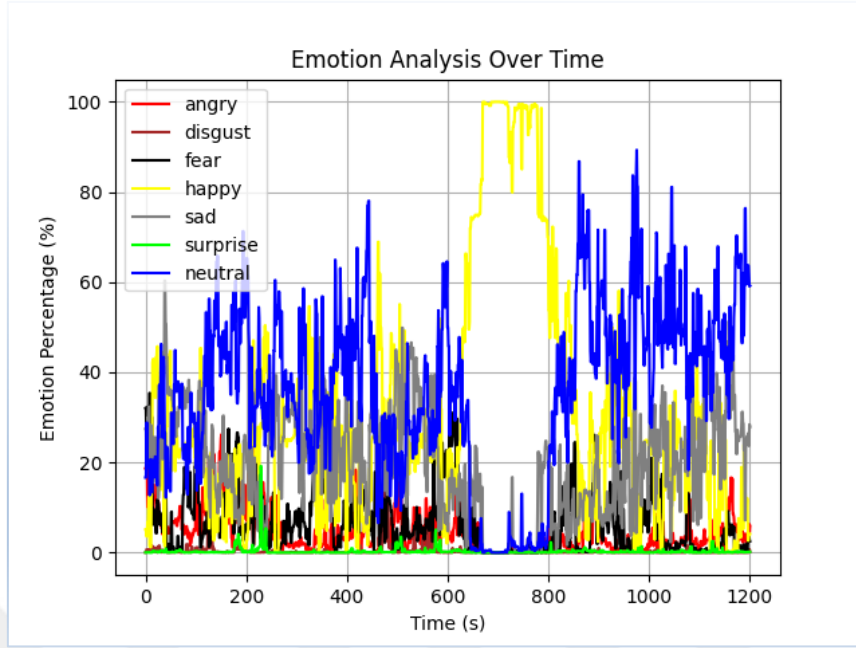
Şekil 3.22. Halı saha röportajı tüm model test görüntüsü

Şekil 3.23'te ise CNN modelinin 7 temel duygu bir oramdaki analizinin yapıldığı test çalışmasıdır. Resimden de görüleceği üzere dört kişinin de aynı anda mutlu olduğu durumda modelimiz ~%99-100 gibi bir ortamın tahminini yapmıştır.



Şekil 3.23. Metropolcard iş toplantısı CNN tüm duygular için ortam tahmini

Şekil 3.24'te bu toplantının 40 saniyelik bölümünde 1200 kare üzerinde ortam duygu analizinin tüm temel duyguları içeren grafiği verilmiştir.



Şekil 3.24. Metropolcard toplantısının ortam duygu analizi grafiği

Grafikten de görüleceği üzere bu toplantının normal duygu durumunda devam ederken belirli bir süre mutluluk yani gülüşmelerin arttığını görebiliriz.

3.6. Ortam Duygu Yoğunluğu Analiz Yöntemleri

Ortam duygu analizi yaparken bazı sorular ve performans sorunları ile karşılaştık. Görüntü işleme ve saniyede 30 video karesi olması sebebi ile işlem ve performans sorunları ile karşılaşıldı. Bu sebeple bu sorunlar aşmak için çeşitli yöntemler denenmiştir. Ortam duygu yoğunluğunu tespit etmek amacıyla beş farklı yöntem geliştirilmiştir.

Tüm Yüzler ve Tahminler Üzerinden Ortalama Yoğunluk Analizi: Bu yöntemde, belirli bir zaman aralığında kaydedilen tüm karelerdeki yüz ifadeleri analiz edilerek ortalama duygu yoğunluğu hesaplanmıştır.

Duygu Değişim Süreçlerine Dayalı Ortalama Yoğunluk Analizi: Bu yöntem, duygu durumundaki değişimlerin zaman ekseninde izlenmesi ve bu süreçlere dayalı olarak duygu yoğunluğunun hesaplanmasını hedeflemektedir.

Eşik Değer Tabanlı Ortalama Yoğunluk Analizi: Belirli bir eşik değerin üzerindeki tahminlerin dikkate alınmasıyla, ortamın genel duygu yoğunluğunun daha hassas bir şekilde belirlenmesi amaçlanmıştır.

Anlık Görüntülerden Ortalama Yoğunluk Analizi: Bu yöntemde, seçilen anlık görüntüler üzerinden duygu yoğunluğu değerlendirilmiştir.

Yüz Bölgesindeki Değişimlerin Takibine Dayalı Ortalama Yoğunluk Analizi: Bu yöntem, yüz bölgesindeki mikro değişimlerin izlenerek bu değişimlerin duygusal durum üzerindeki etkilerinin incelenmesine odaklanmıştır.

Bu yöntemler, ortam duygu yoğunluğunun daha doğru ve kapsamlı bir şekilde belirlenmesini sağlamanın yanı sıra, duygu analiz sisteminin verimliliğini artırmak için stratejik bir çerçeve sunmaktadır.

3.6.1. Tüm Yüzler ve Ortalama Yoğunluk Analizi

Bu yöntem, video girişindeki her bir karenin analiz edilmesini içerir. Tespit edilen tüm yüzler, herhangi bir katsayı veya önceliklendirme olmaksızın modele girdi olarak verilmiştir.

Modelin çıktıları, tahminlenen duyguların kare başına analiz edilmesiyle değerlendirilmiştir. Tahminler, her kare için bir diziye kaydedilerek, önceki hesaplamalar için saklanmıştır.

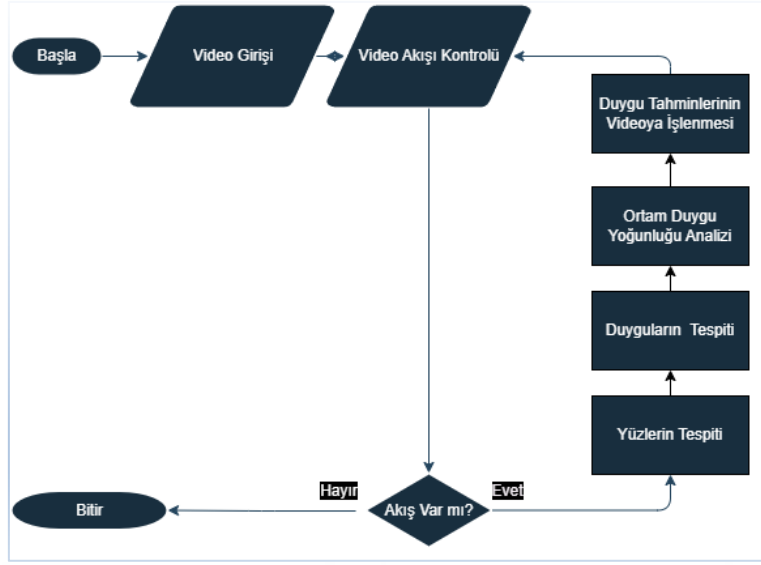
Denklem (3.1)'de gösterildiği gibi, her bir tahminin değerleri toplanmıştır;

$$\text{OutPred}[i] = \sum_{j=1}^N \text{exp_predictions}[j][i] \quad (3.1)$$

Denklem (3.1)'de “N” toplam tahmin sayısıdır ve ‘i’ değeri ise yapılan tahmin sonuçlarıdır. Burada yapılan tüm tahminler 7 boyutlu bir dizi olarak tutulmaktadır. Ardından, ortalama duygu yoğunluğu, Denklem (3.2)'de belirtildiği şekilde, elde edilen tahminlerin toplamının maksimum değeri kullanılarak yüz sayısına oranlanmış ve yüzde olarak ifade edilmiştir;

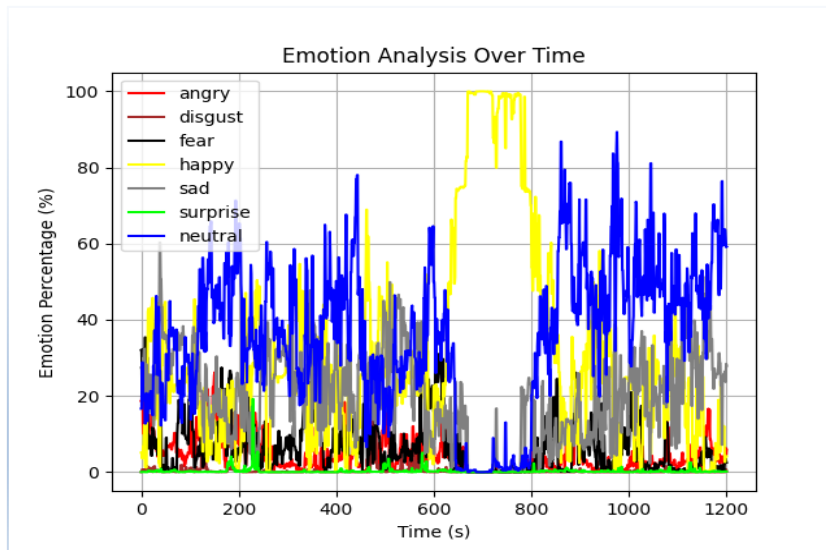
$$\text{Tahmin Yüzdesi} = \left(\frac{\text{OutPred}[\text{max_index}]}{\text{len}(\text{all_face_locations})} \right) \times 100 \quad (3.2)$$

Burada ortamın tahmin edilen en yüksek değeri elde edilmektedir. Aynı zamanda diğer tahminlerde oranlanarak elde edilmektedir.



Şekil 3.25. Tüm yüzler ve ortalama yoğunluk analizi akış diyagramı

Şekil 3.25’de tüm karelerin ve tüm yüzlerin analizinin yapılmasına ait akış diyagramı gösterilmiştir. Görüldüğü üzere video akışı devam ettiği sürece analiz yapılmaya devam ediyor. Bu yöntemin bilgisayar sistemindeki kaynak tüketimi ve çalışma hızı üzerindeki etkileri de değerlendirilmiştir. CPU kullanımı yaklaşık %15 artarken, bellek kullanımı %5 oranında artış göstermiştir. Harici GPU kullanımı, normal koşullarda 0 iken, önceden kaydedilmiş videolar üzerinde işlem yapıldığında %5 seviyelerine çıkmıştır. Bu analiz, yöntemin sistem performansı üzerindeki etkilerini değerlendirerek verimliliğini belirlemek açısından önemlidir.



Şekil 3.26. Tüm yüzler ve tahminler için ortalama yoğunluk yöntemi grafiği

Şekil 3.26’da ortam duygu analizini için kullandığımız en kapsamlı yöntemin zaman-uygu grafiği gösterilmektedir.

Grafikten görüleceği üzere 1200 karelik bir işlem süreci gerçekleşmiştir. Toplantının 20 saniyelik bölümünün normal bir duygu durumu ile seyrettiğini küçük gülüşmelerin olduğunu görebiliriz. Daha sonra ortamdaki herkesin güldüğünü ve ardından normal duygu durumunun devam ettiğini okunmaktadır.

3.6.2. Duygu Değişim Süreci ile Ortam Duygu Analizi

Bu yöntemde, bir önceki yöntemde karşılaşılan her karenin işlenmesi sırasında oluşan zaman kaybını ve aynı karelerin neredeyse aynı sonuçlarla tekrar işleme sokulmasını engellemek amacıyla yeni bir yaklaşım geliştirilmiştir.

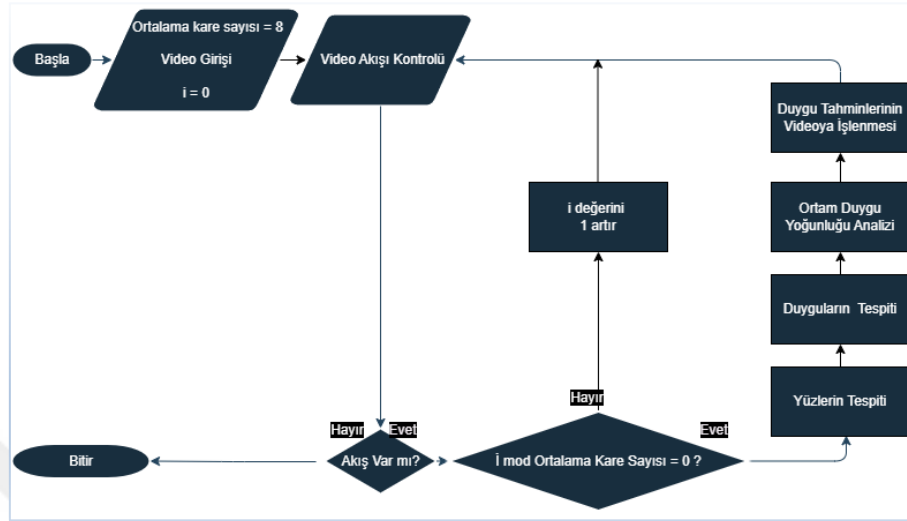
Bu çalışmada, saniyede 30 kareyi analiz etmek yerine, grafiksel sonuçlar incelenerek yaklaşık duygu değişimini belirlemek için ortalama süre hesaplanmıştır. Bu, analiz sürecinin verimliliğini artırarak gereksiz işlem yükünü azaltmayı hedeflemektedir. Bu süre saniye cinsinden değil kaç karede bir değişim yaşandığını hesaplanarak belirlenecektir. Ortalama duygu yoğunluğu değişim süreci, Denklem (3.3)’de belirtildiği gibi hesaplanmaktadır;

$$\text{Ortalama Kare Sayısı} = \left(\frac{\text{Toplam kare sayısı}}{\text{len(all_Predict_Emotion)}} \right) \times T \quad (3.3)$$

Denklem (3.3)’deki “Ortalama Kare Sayısı” ifadesi, ortalama duygu yoğunluğu değişim kare sayısını, “Toplam kare sayısı” ifadesi, videoda kaç kare (frame) olduğunu göstermektedir, “len(all_Predict_Emotion)” ifadesi duygu değişimlerinin toplam sayısını yani kaç kere maksimum değerdeki sınıf değiştiğini, T ifadesi ise tolerans katsayısını ifade eder ve duygu değişim sürecindeki ince ayarlamalar için kullanılır.

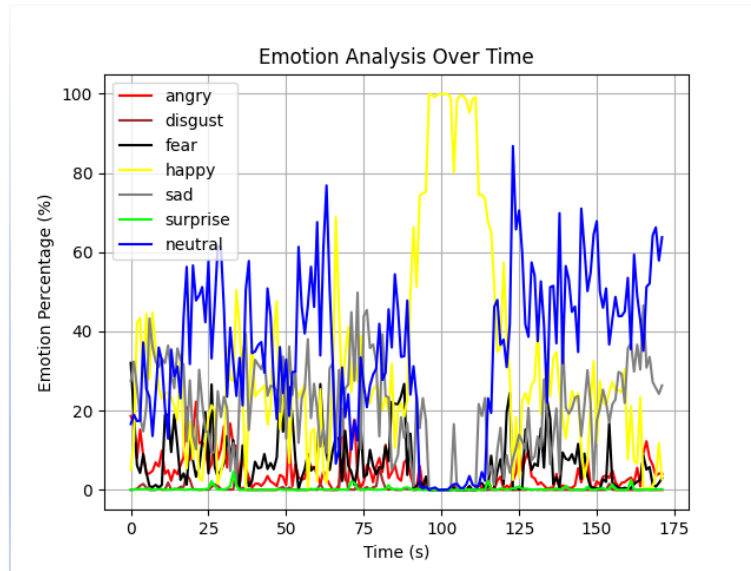
Bu hesaplama sonucunda, duygu yoğunluğunun değişmesi için gerekli olan minimum, maksimum ve ortalama kare sayıları belirli bir tolerans ile tespit edilmiştir. Örneğin, 0.8 tolerans katsayısı kullanılarak ortalama 10 karede bir değişen duygu yoğunluğu her kare yerine 8 karede bir hesaplanacak şekilde düzenleme yapılmış ve bu ayarlarla aynı video üzerinde test edilmiştir.

Tolerans katsayısının belirlenmesindeki amaç, mikro ifadeler ve ani beklenmedik değişimleri kaçırmamak için güvenli bir aralık sağlamak olmuştur.



Şekil 3.27. Duygu değişim süreci ile ortam duygu analizi akış diyagramı

Şekil 3.27’de hesaplanan duygu değişim süresinin sisteme dahil edilmesi ile değişen akış diyagramı gösterilmiştir. Video akışı döngüsü devam etmekle birlikte yüz tespiti ve duygu analizi değişim süresini baz alarak eğer şart sağlanırsa çalıştırılmaktadır.



Şekil 3.28. Duygu değişim süreci ile ortam duygu analizi yöntemi grafiği

Şekil 3.28’de duygu değişim süreçleri belirlenerek yapılan analiz sonuçları görülmektedir. Yöntem 1’e kıyasla işlem süresinde kayda değer bir iyileşme sağlanmış

ve kaynak tüketimi optimize edilmiştir. Ancak, duygu tahminlerinde küçük sapmalar gözlenmiştir; bu da yöntemin doğruluğu bir miktar düşürmektedir.

3.6.3. Eşik Değeri Belirlenmesi ile Ortam Duygu Analizi

Eşik değeri belirleme yöntemi, performans iyileştirmeleri sağlamak ve modelin düşük tahminlerini ortalama hesaplamalarına dâhil etmeyerek daha net ve güvenilir bir duygu yoğunluğu değeri elde etmek amacıyla geliştirilmiştir. Bu yöntemde, bir yüz üzerindeki tahminlerin oranları sıralanarak üçüncü en yüksek tahminlerin ortalaması alınmış ve bu ortalama eşik değeri olarak belirlenmiştir. Bu sayede, model tahminlerinden sonraki hesaplamaların doğruluğu ve tutarlılığı artırılmıştır.

Eşik değeri hesaplaması Denklem (3.4)’te gösterilmektedir;

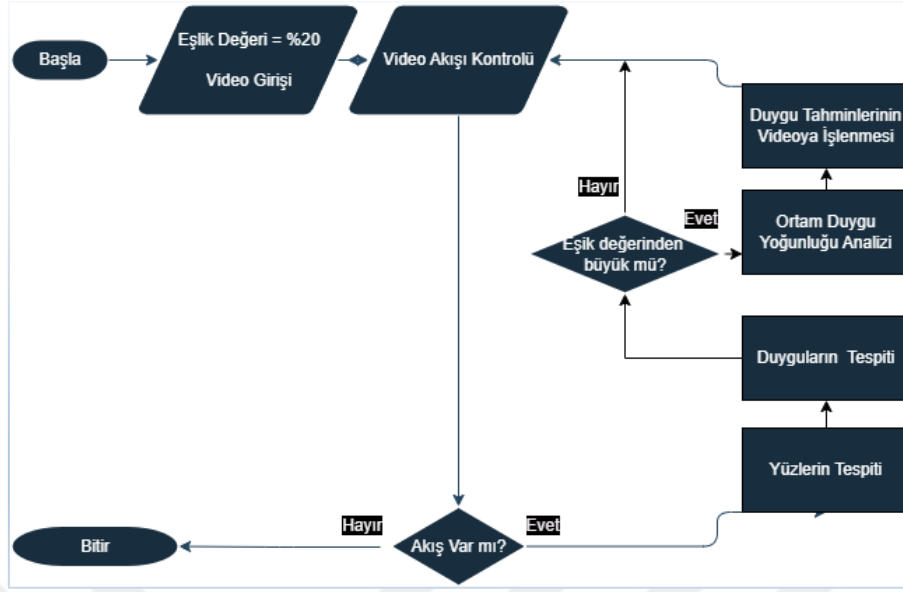
$$\text{Eşik değeri} = \left(\frac{\text{OutPred}[2]_1 + \text{OutPred}[2]_2 + \dots + \text{OutPred}[2]_N}{\text{frame_count}} \right) \quad (3.4)$$

Denklem (3.4)’teki “Eşik değeri” ifadesi, düşük tahminlerin filtrelenmesi için kullanılan sınır değerini, “OutPred[2]” üçüncü en yüksek tahmin değerini, “frame_count” ifadesi ise hesaplama yapılan kare sayısını ifade etmektedir. Bu şekilde en yüksek üçüncü duygu tahminlerinin toplamını kare sayısına bölerek ortalama bir 3. tahmin değeri elde etmiş olacağız.

Ardından belirlenen eşik değeri kullanılarak, model tahminlerinden yalnızca bu eşik değerin üzerindeki değerler hesaplamaya dahil edilmiştir. Böylece, duygu yoğunluğunu etkileyen düşük ve gürültülü tahminlerin etkisi ortadan kaldırılmıştır.

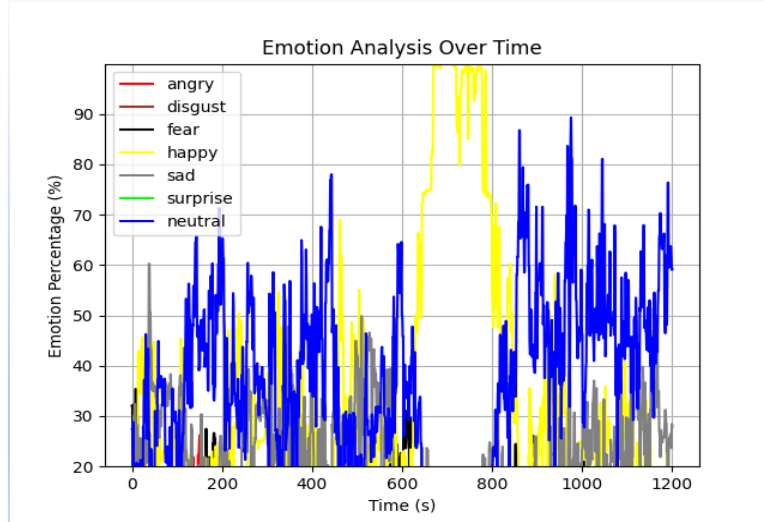
Bu yaklaşım, özellikle anlık biçimsiz ifadelerin ve modelin yanlış algılayabileceği mikro duygu geçişlerinin etkisini azaltarak daha okunabilir ve doğru bir grafik elde edilmesini sağlamıştır.

Şekil 3.29’da eşik değeri yönteminin sisteme uygulandığının akış diyagramı gösterilmektedir. Video akış döngüsünün içinde yüz tespitinden sonra duygu tespiti yapılmasında sonra belirli bir yüzdelik değerin üzerinde ki değerler hesaplamaya dahil edilmiştir.



Şekil 3.29. Eşik değeri belirlenmesi ile ortam duygu analizi akış diyagramı

Eşik değerinin uygulanması, özellikle normal şartlarda gülme gibi pozitif duygulara sahip bireylerin ani ve beklenmedik duygu geçişlerinde modelin hatalı algılamalarını önlemekte etkili olmuştur. Ancak, mikro ifadelerin kaybolmaması için veri setinin genişletilmesi ve bu tür ifadelerin daha yüksek doğrulukla tahmin edilmesinin önem arz ettiği anlaşılmıştır.



Şekil 3.30. Eşik değeri belirlenmesi ile ortam duygu analizi yöntemi grafiği

Yöntem 3 olarak Şekil 3.30'de eşik değeri sisteme entegre edilmiş ve grafiği yukarıdaki gibi 1200 karelik bir işlem gerçekleşmiştir. Görüldüğü üzere gürültüler engellenerek hesaplama maliyetleri azaltılmıştır.

3.6.4. Anlık Görüntülerden Ortam Duygu Analizi

Bu yöntemde, her saniye içerisinde rastgele anlık görüntüler alınarak duygu yoğunluğu analizi yapılması hedeflenmiştir. Bu yaklaşımın amacı, işlem süresini kısaltarak her kare üzerinde işlem yapmaya gerek kalmadan benzer doğrulukta sonuçlar elde etmektir. Bu sayede, hem işlem süresi önemli ölçüde azaltılmakta hem de duygu yoğunluğu analizinin doğruluğu korunmaktadır.

Anlık görüntülerin seçimi aşağıda Denklem (3.5) ve Denklem (3.6) ifade edilmektedir;

$$N = T \times F \quad (3.5)$$

Burada;

- T = Toplam süre (saniye cinsinden), örnek 40 saniye
- F = Kare sayısı (frame rate), örnek 30 fps
- N = Toplam kare sayısı, yani $40 \times 30 = 1200$

Şimdi, her saniyeden rastgele bir kare seçmemiz gerekiyor. Bunu şu şekilde ifade edebiliriz;

Her saniyenin $F=30$ karesi arasından birini seçmek için;

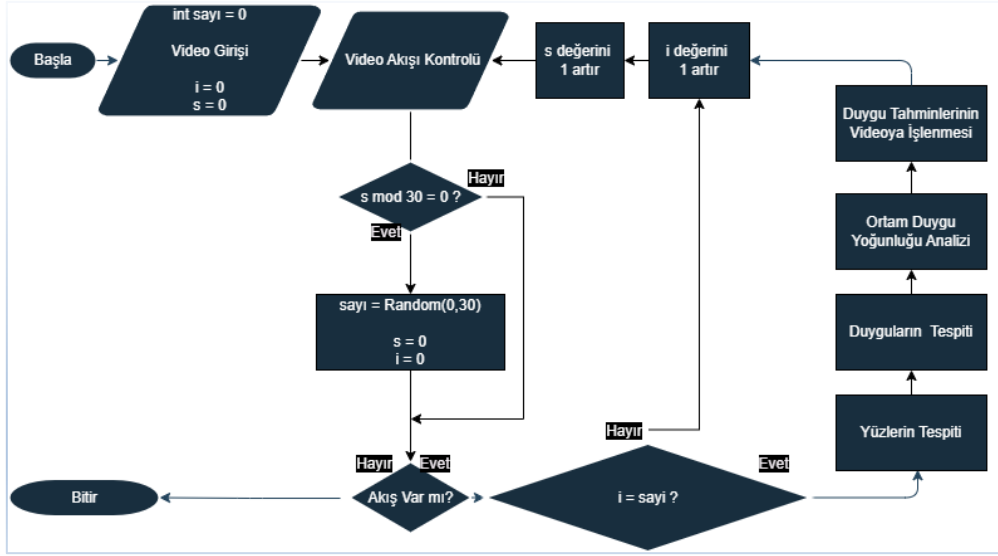
$$S_i = R_i, R_i \sim U(1,30) \quad \forall i \in \{1,2,3, \dots, T\} \quad (3.6)$$

Böylece, seçilen karelerin kümesi şu şekilde olur;

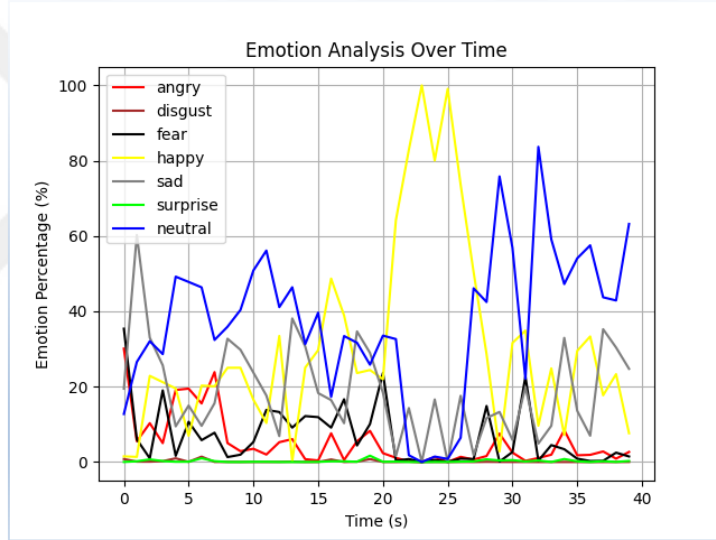
$$S = \{S_1, S_2, S_3, \dots, S_{40}\} \quad (3.7)$$

Denklem (3.7) 40 tane rastgele seçilmiş kareyi temsil eder. Bu her kare döngü sayacında sırası geldiğinde tahmin işlemi yapmayı sağlar.

Şekil 3.31’de anlık görüntülerden ortam duygu yoğunluğunun hesaplanması için yeni akış diyagramı gösterilmektedir. Bu yöntemde 30 kare/saniye’lik bir video girişinden her saniye yani her 30 kare için tek bir kare üzerinden ortam duygu yoğunluğu hesaplaması yapılmıştır. Rastgele seçilen bir sayı için o an ki periyotta yüz tespiti ve duygu analizi yapmadan önce döngünün rasgele sayı ile aynı değere gelmesi beklenmektedir.



Şekil 3.31. Anlık görüntülerden ortam duygu analizi akış diyagramı



Şekil 3.32. Anlık görüntü alınarak ortalama yoğunluk analizi yöntemi grafiği

Şekil 3.32’de anlık görüntülerden çalıştığımız yöntemin grafiği incelenmektedir. Bu yöntem ile yaklaşık 40 saniyelik bir video üzerinde her saniyede 1 kare işlenerek çalışılmıştır. Sonuçlar okuma açısından bir öncekilere yakın ama sert değişikliklerin olduğu kısımlar mevcuttur.

3.6.5. Yüz Lokasyonundaki Değişiklerin Takibi ile Ortam Duygu Analizi

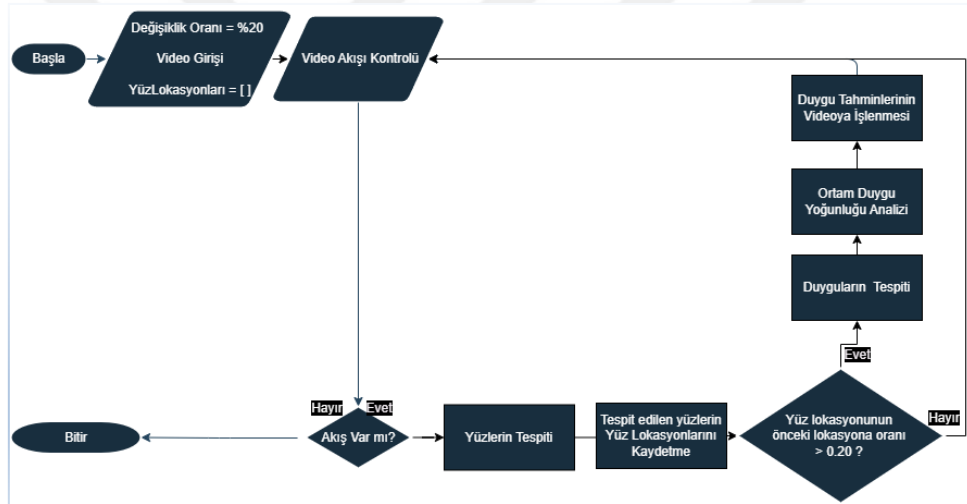
Bu yöntemde, yüz ifadelerindeki değişikliklere bağlı olarak yüz lokasyonlarında meydana gelen değişiklikler tespit edilerek, duygu değişimleri gözlemlenmiş ve ortam duygu yoğunluğu analiz edilmiştir. Video girişinden alınan karelerde, MTCNN algoritması ile

yüz tespiti yapılarak her bir yüzün lokasyonu belirlenmiştir. Bu lokasyonlar hafızada tutulmuş ve bir önceki karedeki lokasyonlarla karşılaştırılmıştır. Eğer lokasyon değişikliği %20'nin (0,2 oranında) üzerinde gerçekleşirse, o kare duygu analizi hesaplamalarına dahil edilmiştir.

Bu yöntemin matematiksel ifadesi Denklem (3.8)'deki gibi ifade edilmektedir;

$$\text{Değişim oranı} = \frac{\text{Anlık lokasyon} - \text{önceki lokasyon}}{\text{önceki lokasyon}} \quad (3.8)$$

Denklem (3.8)'deki “Değişim oranı” ifadesi, iki kare arasındaki yüz lokasyonundaki değişiklik oranını, “Anlık lokasyon” ifadesi, mevcut karedeki yüz lokasyonunu, “önceki lokasyon” ifadesi, bir önceki karedeki yüz lokasyonunu ifade etmektedir.

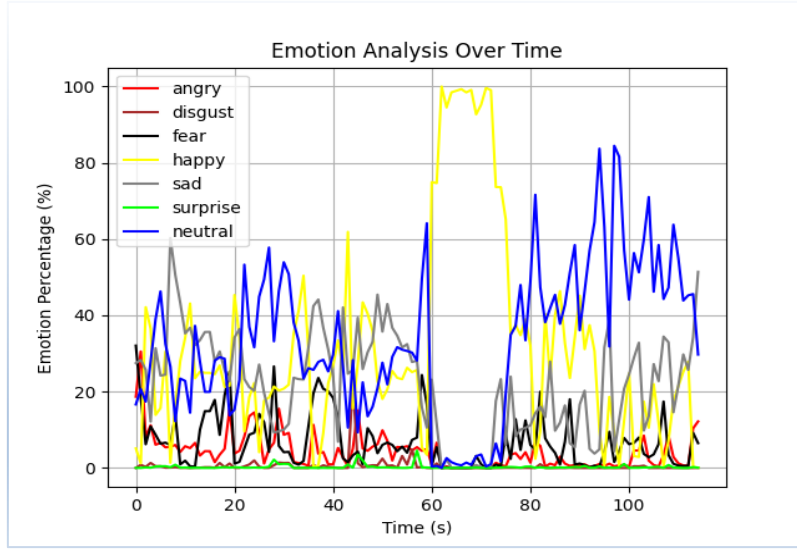


Şekil 3.33. Yüz lokasyonundaki değişikliklerin takibi ile ortam duygu analizi akış diyagramı

Lokasyon değişikliklerinin duygu değişimi için bir sinyal olabileceği ön görülmüştür. Şekil 3.33'de son yöntemimiz olan yüz lokasyonlarındaki değişiklikler incelenmesi ve analiz akış diyagramı gösterilmektedir.

Bu yaklaşım ile her kareyi incelemek yerine, yalnızca yüz lokasyonlarında anlamlı değişiklikler olduğunda duygu analizi işlemleri yapılarak işlem yükü azaltılmış ve verimli sonuçlar elde edilmiştir.

Bu yöntem, duygu değişikliklerini yakalamada etkili olup diğer yöntemlerle birleştirildiğinde verimliliğin daha da artacağı öngörülmektedir.



Şekil 3.34. Yüz bölgesindeki değişikliklerin takibi ile ortalama yoğunluk analizi yöntemi grafiği

Şekil 3.34'te ise son yöntemimiz olan yüz lokasyonlarında yaşanan değişiklikler izlenmiştir. Grafiğe bakıldığında ana hatlarda bir değişiklik görülmemiştir. Yaklaşık 120 kez lokasyon değişikliği gerçekleşmiş ve analiz yapılmıştır.

4. BULGULAR

Bu bölümde, geliştirilen modellerin performans değerlendirmelerine ve elde edilen sonuçlara yer verilmiştir. Çalışma kapsamında farklı model mimarileri kullanılmış, her bir modelin eğitim ve test süreçleri ayrı ayrı incelenmiştir. Bu bağlamda, EfficientNet-B7, ConvLSTM ve CNN modeli üzerinde durulmuş; bu modellerin doğruluk oranları, gerçek zamanlı test sonuçları ve diğer performans metrikleri karşılaştırılmıştır.

Modellerin başarımını değerlendirmek için öncelikle eğitim ve test verileri üzerindeki doğruluk ve kayıp değerleri analiz edilmiştir. Bunun yanı sıra, gerçek zamanlı testler gerçekleştirilerek modellerin pratik kullanıma uygunluğu incelenmiştir. Özellikle, EfficientNet-B7, ConvLSTM ve CNN gibi daha karmaşık modellerin performansları kıyaslanarak, ortam duygu analizinde kullanılacak model seçiminde etkili olan faktörler belirlenmiştir.

Bu bağlamda, aşağıdaki alt başlıklarda;

- Doğruluk oranları, eğitim ve test süreçlerinde modellerin başarımının detaylı analiziyle sunulmuş,
- Gerçek zamanlı test sonuçları, modellerin dinamik veri üzerinde uygulama performansı açıklanmış ve bu sonuçların genel değerlendirmesi yapılmıştır.

Sonuçlar hem modellerin teorik başarımını hem de gerçek dünya uygulamalarındaki pratik performanslarını anlamaya yönelik kapsamlı bir analiz sağlamaktadır.

4.1. Model Performans Sonuçları

Her modelin genel doğruluk oranları ve sınıflandırma raporu incelenmiştir. Çıkan sonuçlar ile model .h5 formatında kaydedilerek gerçek zamanlı testler üzerinde performansları değerlendirilmiştir.

4.1.1. EfficientNet-B7

EfficientNet-B7 modeli ile FER-2013 veri seti üzerinde yapılan eğitim sonucunda elde edilen performans metrikleri, veri setinin farklı karakteristiklerini ve modele etkilerini açıkça ortaya koymaktadır.

FER-2013 veri seti, daha büyük ve çeşitli bir veri kümesi olmasıyla, modelin genelleme yeteneğini test etmek için uygundur. Sonuçlar şunları göstermektedir:

- Genel Doğruluk: %68
- En Yüksek Performans: "Mutluluk" sınıfında F1 skoru 0.88 ile dikkat çekmektedir.
- En Düşük Performans: "Üzgünlük" ve "Korku" sınıflarında sırasıyla 0.51 ve 0.53 F1 skorları görülmektedir. Bu, bu sınıfların daha zor ayırt edilebilir olduğunu veya veri setinde bu etiketlere ilişkin örneklerin dengesiz dağıldığını gösterebilir.
- Makro Ortalama: 0.68, sınıflar arasında genel bir denge sağlandığını ancak bazı sınıfların düşük performans gösterdiğini göstermektedir.
- Ağırlıklı Ortalama: 0.68, veri setindeki sınıf dengesizliğinin performansa doğrudan etkisini vurgulamaktadır.

FER-2013 sonuçlarına göre, model çeşitli sınıflarda kabul edilebilir bir performans sergilemiş, ancak bazı sınıflar için başarıyı sınırlı kalmıştır.

4.1.2. ConvLSTM

Yüz ifadelerini tanımak için diğer bir model olarak ConvLSTM (Convolutional Long Short-Term Memory) katmanları içeren bir sinir ağı modeli tanımlandı. Video gibi sıralı ve zaman indeksli veriler için geliştirilen model gerçek zamanlı uygulamalarda kullanılabilirlik açısından seçilmiştir.

Model, birkaç ConvLSTM ve Batch Normalization katmanı, ardından MaxPooling ve Dropout katmanları kullanılarak inşa edildi. Son olarak, zaman içinde düzeltilmiş (TimeDistributed) bir Flatten ve Dense katmanı eklenerek çıktı sınıflandırıldı. Model, kategorik çapraz entropi kaybı fonksiyonu ve Adam optimizasyon algoritması kullanılarak eğitildi.

Eğitim süreci boyunca, modelin doğruluk ve kayıp değerleri izlenerek performansı değerlendirildi.

FER-2013 veri seti, gerçek dünya senaryolarını yansıtan, büyük ve çeşitli bir veri kümesi olmasıyla, modelin genelleme yeteneğini test etmek için uygundur. Sonuçlar şunları göstermektedir:

- Genel Doğruluk: %77
- En Yüksek Performans: "Mutluluk" sınıfında F1 skoru 0.91 ile dikkat çekmektedir. Bu, mutluluk sınıfının net özelliklere sahip olduğunu ve modelin bu ifadeyi tanımakta başarılı olduğunu göstermektedir.
- En Düşük Performans: "Korku" ve "Doğal" sınıflarında sırasıyla 0.74 ve 0.63 F1 skorları görülmektedir. Bu durum, korkunun diğer duygularla karışabilir olduğunu, doğal ifadelerin ise nötr sınıf ayrımının daha zor olduğunu işaret eder.
- Makro Ortalama: 0.77, sınıflar arasında genel bir denge olduğunu ancak düşük performans gösteren sınıfların bulunduğunu göstermektedir.
- Ağırlıklı Ortalama: 0.77, veri setindeki sınıf dengesizliğinin genel performansı sınırladığını ortaya koymaktadır.

FER-2013 sonuçlarına göre, model geniş veri kümesi üzerinde kabul edilebilir bir performans sergilemiş, ancak sınıf dengesizliği ve bazı sınıfların karmaşıklığı performansı etkilemiştir.

4.1.3. CNN

FER-2013 veri seti, gerçek dünya yüz ifadelerini yansıtan geniş ve çeşitli bir veri kümesi olması nedeniyle modelin genelleme yeteneğini test etmek için kullanılmıştır. Sonuçları göstermektedir:

- Genel Doğruluk: %85
- En Yüksek Performans: "Mutluluk" sınıfında F1 skoru 0.93 ile modelin en iyi performansı gözlemlenmiştir. Bu, mutluluk sınıfının belirgin özellikleri nedeniyle diğer sınıflardan ayrışmasının daha kolay olduğunu göstermektedir.
- En Düşük Performans: "Üzüntü" ve "Korku" sınıflarında F1 skorları sırasıyla 0.82 ve 0.81 olarak ölçülmüştür. Bu durum, bu sınıfların daha karmaşık yapılarından kaynaklanabilir.
- Makro Ortalama (Macro Avg): 0.86, tüm sınıfların performanslarının genel olarak dengeli olduğunu ancak bazı sınıfların daha düşük performans sergilediğini göstermektedir.
- Ağırlıklı Ortalama (Weighted Avg): 0.85, sınıf dengesizliğinin genel doğruluk üzerindeki etkisini ortaya koymaktadır.

FER-2013 sonuçları, modelin genel kabul edilebilir bir doğruluğa sahip olduğunu, ancak özellikle zorlu veya karışabilir sınıflarda performansının sınırlı olduğunu ortaya koymaktadır.

4.1.4. Gerçek Zamanlı Test Sonuçları

Model eğitim süreçleri tamamlandıktan sonra, eğitilen modeller .h5 formatında kaydedilmiştir. Modeller, ağırlıkları ile birlikte kaydedilmiş olup, daha sonra doğru bir şekilde yüklenip yüklenmediklerinin kontrolü gerçekleştirilmiştir. Yüz tespiti yapılan çalışmalarda kullanılan çeşitli videolar aracılığıyla modellerin performans testleri yapılmıştır. Bu testlerde, FER-2013 veri seti ile eğitilmiş modeller tercih edilmiştir. Bunun temel sebebi, CK+48 veri setinin sınırlı sayıda fotoğraf içermesi nedeniyle, bu veri seti ile eğitilen modellerde aşırı öğrenme sorunu yaşadığının düşünülmesidir. Ayrıca, yapılan manuel testler sonucunda, FER-2013 veri seti ile eğitilen modellerin ürettiği ağırlıkların daha doğru sonuçlar verdiği gözlemlenmiştir.

Duygu-yoğunluğu analizi için geliştirilen ve test edilen üç modelin performansı çeşitli açılardan incelenmiştir. Bu modeller; EfficientNet B7, evrişimli LSTM ve CNN modelidir. Aşağıda her bir modelin detaylı açıklaması ve analiz sonucunda elde edilen çıkarımlar yer almaktadır.

EfficientNet ailesine ait B7 modeli, çok sayıda katmandan oluşan derin bir mimariye sahiptir. Bu model, bazı optimizasyonlarla yeniden eğitilmiş ve duygu analizi için uyarlanmıştır. Ancak eğitim süresi beklenenden uzun sürmüş, bu durum modelin karmaşık yapısından kaynaklanmıştır. Özellikle çok sayıda katmanın bulunması hem eğitim hem de uygulama aşamasında çok yüksek donanım gereksinimlerine neden olmuştur. Modelin uygulama boyutunun büyük olması, gerçek zamanlı testlerde ek yük oluşturmıştır. Performans sonuçları, yaklaşık %70 doğruluk oranı ile sınırlı kalmıştır.

Evrişimli LSTM modeli, sıralı veriler için geliştirilmiş bir yapay sinir ağı modeli olarak tasarlanmıştır. Bu model, zamana bağlı verilerin analizinde genellikle başarılı olmakla birlikte, FER-2013 gibi zamansal olmayan statik veri setleriyle eğitildiğinde beklenen performansı sergileyememiştir. Bu durum, veri setinin doğası ile modelin tasarım felsefesinin uyuşmamasından kaynaklanmıştır. FER-2013 veri seti, zamansal bağlantı

gerektirmeyen, bireysel yüz ifadelerini içeren statik bir yapıya sahiptir. Bu nedenle, LSTM'nin zamansal bağıllık üzerine kurulu mimarisi bu tip bir veri setinde verimli bir şekilde kullanılamamıştır.

4.2. Ortam Duygu Yoğunluğu Analizi Sonuçları

Duygu analizi çalışmalarında doğruluk, işlem süresi ve kaynak kullanımı gibi kriterler, yöntem seçimini belirleyen temel unsurlardır. Bu çalışmada beş farklı yöntem değerlendirilmiş ve her birinin avantajları, dezavantajları ile kaynak tüketimine olan etkileri analiz edilmiştir. Tüm yüzler ve tahminler için ortalama yoğunluk hesaplanması yönteminin, en yüksek doğruluğu sağlarken, işlem süresi ve kaynak tüketimi açısından maliyetli bir yapı sergilemiştir. Duygu değişim süresi ve eşik değeri eklenmesi yöntemlerinin, kaynak tüketimini optimize ederek makul doğruluk oranları sunmuş, anlık görüntülerden analiz yöntemi ise işlem süresini büyük ölçüde azaltarak hızlı sonuçlar elde edilmesini sağlamıştır. Son olarak, lokasyonlardaki değişim yöntemi, bellek tüketimini optimize etmesi ve diğer yöntemlerle birleştirildiğinde sağlayabileceği potansiyel verimlilikle dikkat çekmiştir. Her bir yöntem, farklı ihtiyaçlara yönelik çözümler sunarak geniş bir uygulama yelpazesi için değerlendirilmiştir.

4.2.1. Farklı Yöntemlerin Karşılaştırılması

Yöntem 1 (Tüm Yüzler ve Tahminler için Ortalama Yoğunluk Hesaplanması);

Bu yöntem, her bir karedeki tüm yüzlerin duygu tahminlerinin yapılmasını ve ortalama yoğunluğun belirlenmesini kapsamaktadır. Sonuçlar, yüksek doğruluk oranları sağlamakla birlikte, CPU kullanımı %15, bellek kullanımı %5 ve GPU kullanımı %5 oranında artmıştır. Bu yöntem, en yüksek doğruluğu sağlamakla birlikte, işlem süresi ve kaynak tüketimi açısından en maliyetli yöntemdir. Bu sebeple, yüksek doğruluk gerektiğinde tercih edilebilir ancak maliyetli kaynak tüketimi nedeniyle sınırlı kullanım alanına sahiptir.

Yöntem 2 (Eşik değeri belirlenmesi ile ortam duygu analizi);

Bu yöntemde, belirlenen eşik değerinin altındaki tahminler dikkate alınmayarak düşük tahminlerin etkisi ortadan kaldırılmıştır. Bu sayede, duygu yoğunluğu tahminlerinde daha

net ve güvenilir sonuçlar elde edilmiştir. Kaynak tüketiminde önemli bir değişiklik gözlenmemekle birlikte, duygu yoğunluğunun doğru tahmin edilmesi açısından önemli bir katkı sağlamıştır. Bu yöntem, gürültü ve düşük tahminleri filtreleyerek analiz kalitesini artırmaktadır.

Kaynak tüketimi ve performans açısından yapılan değerlendirmede, bu yöntem bir önceki yönteme göre CPU kullanımını %10 oranında artırmış, bellek kullanımı ise aynı kalmıştır. GPU kullanımı ise %3'e düşmüştür. Sonuçlarda tahminlerde küçük değişiklikler gözlenirken, genel grafiksel seyir büyük ölçüde korunmuştur. Bu sonuçlar, yöntemin performansını artırırken duygu değişimlerini güvenli bir şekilde yakalamaya devam ettiğini göstermektedir.

Yöntem 3 (Duygu değişim süreci ile ortam duygu analizi);

Bu yöntemde, her kareyi analiz etmek yerine belirli aralıklarla duygu değişimleri incelenmiştir. CPU kullanımını %10 oranında artmış, bellek kullanımı sabit kalmış ve GPU kullanımı %3'e gerilemiştir.

Sistemde eşik değeri uygulanırken kaynak tüketimi açısından bir değişim gözlenmemiştir, bu da yöntemin performans açısından verimli olduğunu göstermektedir. Sonuç olarak, bu yöntemin diğer analiz yöntemleriyle birlikte kullanılması sistemin genel doğruluğunu artıracak bir yaklaşım olarak öngörülmektedir.

Yöntem 4 (Anlık Görüntü Alınarak Ortalama Yoğunluk Analizi);

Her saniye içerisinde rastgele anlık görüntüler alınarak duygu yoğunluğu analizi yapılmıştır. Bu yöntem, işlem süresini %90 oranında azaltmış, CPU ve bellek kullanımı Yöntem 2 ile aynı seviyede kalmış ancak GPU kullanımı artış göstermiştir. GPU kullanımındaki artışın nedeni, her karede anlık görüntülerin incelenmesidir. Bu yöntem, işlem süresini kısaltmak isteyen uygulamalar için ideal bir çözümdür, ancak GPU kullanımı yüksek olan sistemlerde dikkatli değerlendirilmelidir.

Yöntemin en büyük katkısı, işlem süresini ortalama %90 oranında azaltmasıdır, böylece her kare üzerinde analiz yapma ihtiyacını ortadan kaldırmaktadır.

Yöntem 5 (Yüz Bölgesindeki Değişikliklerin Takibi ile Ortalama Yoğunluk Analizi);

Bu yöntemde, yüz ifadelerindeki değişikliklere bağlı olarak yüz lokasyonlarındaki değişimler izlenmiş ve işlem yükü azaltılmıştır. İşlem süresinde belirgin bir iyileşme sağlanmasa da bellek tüketimi optimize edilmiştir. Yöntem, diğer yöntemlerle birleştirildiğinde daha yüksek verimlilik sağlayabileceği öngörülmektedir. Kaynak tüketimi açısından CPU kullanımı %10-20 arasında değişiklik gösterirken, GPU kullanımı %5-6 oranında artmıştır.

Tablo 4.1’de de görüldüğü üzere, her yöntem kendi içinde farklı avantajlar ve dezavantajlar sunmaktadır.

İlk yöntem en yüksek doğruluğu sağlarken, işlem süresi ve kaynak tüketimi açısından en maliyetli yöntemdir.

İkinci ve üçüncü yöntemler, daha düşük kaynak tüketimi ile kabul edilebilir doğruluk oranları sunarken, dördüncü yöntem işlem süresini büyük ölçüde azaltmasıyla dikkat çekmektedir. Beşinci yöntem ise bellek tüketimini optimize ederek öne çıkmaktadır.

Tablo 4.1. Yöntemlerin performans karşılaştırmaları

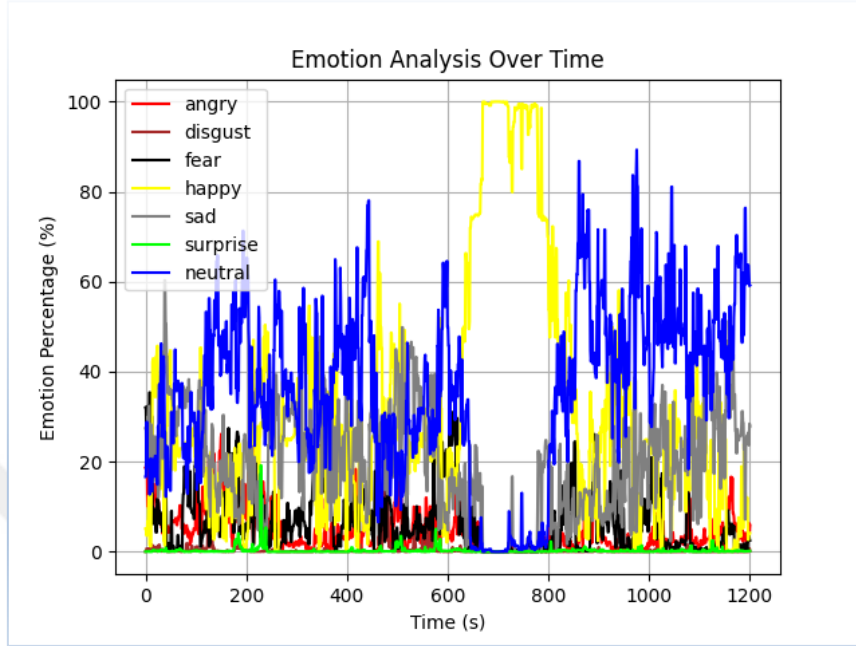
Yöntem	CPU	GPU	Bellek	Süre (dk)
Tüm yüzler	+ %15	+ %0-5	%12	30
Eşik değeri	+ %10	+ %0-3	%11,6	10
Değişim süreci	+ %11	+ %0-2	%12,1	30
Anlık görüntü	+ %10	+ %3-5	%11,6	3
Lokasyon	+ %10-20	+ %5-6	%12,2	30

4.2.2. Uygulama Alanlarına Göre Bulgular

Toplantı anında aldığımız video kaydı ile daha sonrasında çalışmalarımız test etme imkânı elde ettik. Toplantı anında teknik imkanlardan dolayı gerçek zamanlı olarak test edemediğimiz sistemimizi daha sonrasında geliştirmemizde kullanarak görsel ve grafiksel sonuçlar ile toplantı anındaki kendi deneyimlerimiz karşılaştırılmıştır.

Toplantımızın büyük bir bölümünü izleme şansı elde ettiğimizde sistemde elde ettiğimiz sonuçlar ile örtüştüğünü gözlemledik. Şekil 4.1’de gösterilen grafik çoğunlukla mavi ve sarı renkler baskındır. Bu da toplantımız verimli geçtiğini ve moral seviyesinin yüksek olduğunu gösterdi.

Eğer gerçek zamanlı test edebilseydik belki toplantıda ki arkadaşlarımız duygu grafiğine ağırlıklı yansıyan ifadelerini göre toplantı akışını değiştirebilirdik.



Şekil 4.1. Toplantı sonrası oluşan duygu grafiği

5. TARTIŞMA

5.1. Elde Edilen Sonuçların Değerlendirilmesi

Bu çalışma gerçek zamanlı olarak evrişimli sinir ağları kullanılarak ortam duygu yoğunluğu analizi yapılabileceğini önermektedir. Son zamanlardaki duygu tanımlama yöntemleri incelenmiş ve en yaygın kullanılan veri setleri ve yöntemler tercih edilmiştir. FER-2013 veri seti ile çalışma yapılması ile model yüksek doğruluk oranlarında çalışması sağlanmıştır. Geliştirdiğimiz sistemin diğer duygu tespiti video çıktılarının üzerinden test edildiğinde çok yakın doğruluk oraları ile çalıştığı gözlemlenmiştir. Genel olarak çoklu yüzlerin tespiti ve duygu analizinde verimli bir sonuç alınmıştır.

FER-2013 veri setinin geniş bir görüntü çeşitliliği sunduğunu ve farklı yaş gruplarından, düşük kontrastlı, aksesuar takılı ve yüz içermeyen bazı görüntüleri içerdiğini belirtmiştiniz. Bu çeşitlilik, modelin çeşitli yüz ifadelerini tanıma yeteneğini artırmak için kritiktir. Bu geniş veri yelpazesi, modelin genelleme yeteneğini güçlendirir.

Etiket Sayısı ve Çeşitliliği: FER-2013 veri setinin 7 farklı duygu sınıfını içerdiğini belirtmiştiniz. Bu, modelin duygu sınıflandırmasında daha ince ayrımlar yapabilmesi için önemli bir kriterdir. CK+48 veri seti de 7 duygu sınıfı içeriyor ancak daha az sayıda örneğe sahip. Bu, daha derinlemesine analiz ve modelin hassasiyetini test etmek için bir kriter olabilir.

Görüntü Boyutu ve Formatı: 48x48 piksel gri tonlamalı görüntüler, modelin girdi boyutunu optimize etmek ve eğitim süresini kısaltmak için uygun bir seçimdir. Ayrıca, düşük çözünürlük, modelin daha gerçekçi koşullarda çalışmasını sağlar, çünkü birçok uygulamada yüksek çözünürlüklü görüntüler her zaman mevcut olmayabilir.

MTCNN yüz tespiti dedektörü kullanarak kaskad sınıflandırıcı özelliğini çoklu görev yapısına sahip olarak çalışması sağlanmıştır. Bu da projemizin amaçlarından bir olan gerçek zamanlı uygulamalar için performans avantajı sağlamıştır.

EfficientNet-B7 ve ConvLSTM modelleri, farklı alanlarda önemli avantajlar sunarken, bazı durumlarda bu modellerin karmaşıklığı ve hesaplama maliyetleri kullanım kısıtlamalarına neden olabilir. Bunun yerine, derin bir Convolutional Neural Network

(CNN), hem etkinlik hem de verimlilik açısından daha dengeli bir seçenek olarak öne çıkmaktadır.

EfficientNet-B7, model verimliliğini artırmak için ölçeklendirme yöntemlerini optimize eder; ancak, yüksek parametre sayısı ve büyük model boyutu nedeniyle hesaplama gücü gereksinimleri oldukça yüksektir. Derin CNN yapıları, bu zorlukları çözmek için daha hafif ve özelleştirilebilir bir yapı sunar. Özellikle, derin CNN'ler farklı veri setlerine göre kolayca özelleştirilebilir. EfficientNet-B7'ye kıyasla daha az GPU belleği tüketir, bu da daha hızlı eğitime olanak tanır.

ConvLSTM modelleri, özellikle zaman serisi ve sekans verileri için güçlü bir araçtır; ancak, yüz ifadeleri gibi zamansal bağlam gerektirmeyen durumlarda fazla karmaşık olabilir. Derin CNN'ler, zaman serisi analizine gerek duyulmayan görevlerde daha doğrudan çözümler sunar. ConvLSTM'ye göre daha düşük hesaplama süresi ve daha hızlı çıkarım yapma kapasitesine sahiptir.

Görüntü işleme ve yapay zekâ teknikleri, ortamlardaki duygu yoğunluğunu analiz etmeye yönelik farklı yöntemlerin geliştirilmesinde ve değerlendirilmesinde etkin bir şekilde kullanılmıştır. Bu çalışmada, MetropolCard'ın haftalık iş toplantısı kaydı üzerinde uygulanan beş farklı analiz yöntemi, performans, kaynak kullanımı ve işlem süresi açısından kapsamlı bir şekilde karşılaştırılmıştır. Çalışmadan elde edilen bulgular aşağıda özetlenmiştir:

Yöntem Performanslarının Değerlendirilmesi; Tüm yüzlerin ve tahminlerin dahil edildiği birinci yöntem, doğruluk oranı açısından en yüksek performansı sağlamış; ancak işlem süresi ve kaynak kullanımı bakımından en maliyetli yöntem olarak belirlenmiştir. Yüksek doğruluk gerektiren analizlerde tercih edilebilse de, bu yöntemin kaynak yoğun yapısı pratik uygulamalarda sınırlı bir kullanım alanına sahiptir. Buna karşın, diğer yöntemler, işlem süresini azaltma ve kaynak kullanımını optimize etme yönündeki çözümleriyle gerçek zamanlı analizler için daha uygun bir yapı sunmuştur.

Kaynak Kullanımı ve İşlem Süresi Optimizasyonu; Yöntemler arasında kaynak kullanımı ve işlem süresi açısından dikkate değer farklılıklar gözlemlenmiştir. Özellikle ikinci ve dördüncü yöntemler, işlem süresinde kayda değer iyileştirmeler sağlarken, CPU ve bellek

kullanımı açısından optimize edilmiştir. Dördüncü yöntemin, işlem süresini %90 oranında azaltarak önemli bir verimlilik sağladığı görülmüştür. Bununla birlikte, bu yöntemin GPU kullanımı üzerindeki etkisi artmıştır. Bu durum, işlem yükünün hafifletilmesi açısından avantaj sağlarken, kaynak kısıtlaması bulunan sistemlerde dikkatle değerlendirilmesi gereken bir unsur olarak öne çıkmaktadır.

Doğruluk ve Güvenilirlik Boyutu; Düşük doğruluğa sahip tahminlerin analizden çıkarılmasına dayanan eşik değeri belirleme yöntemi, daha güvenilir duygu yoğunluğu tahminleri elde edilmesine imkân tanımıştır. Gürültü ve düşük tahminlerin filtrelenmesiyle analiz kalitesini artıran bu yaklaşım, mikro ifadelerin tespitinde kritik bir rol oynamıştır. Bu yöntemin, diğer analiz yöntemleriyle bir arada kullanılması, genel doğruluğun artırılması için etkili bir strateji olarak değerlendirilmektedir.

Yüz Lokasyonlarındaki Değişimlerin Takibi; Beşinci yöntem, yüz ifadelerindeki değişimlere bağlı olarak yüz lokasyonlarında meydana gelen hareketleri analiz etmeye odaklanmıştır. Bu yöntem, her kareyi analiz etmek yerine yalnızca anlamlı değişimlerin bulunduğu kareleri değerlendirerek işlem yükünü önemli ölçüde azaltmıştır. Kaynak tüketimi üzerindeki olumlu etkileriyle dikkat çeken bu yaklaşım, diğer yöntemlerle entegre edildiğinde genel sistem verimliliğini artıracak bir potansiyel taşımaktadır.

Gerçek Zamanlı Uygulamalar için Uygulanabilirlik; Geliştirilen yöntemler, duygu yoğunluğu analizinin çeşitli ortam ve senaryolarda uygulanabilirliğini göstermiştir. Seminerler, dersler ve doktor-danışan görüşmeleri gibi farklı bağlamlarda kullanılabilecek bir çerçeve sunan bu yöntemler, yüz ifadelerinden elde edilen duygu yoğunluğu ile ortamın genel duygusal atmosferinin izlenmesine olanak tanımaktadır.

Sonuç olarak, bu çalışma, görüntü işleme ve yapay zekâ tekniklerinin bir arada kullanılmasıyla ortam duygu yoğunluğu analizinde yüksek doğruluk ve verimlilik sağlanabileceğini ortaya koymuştur. Geliştirilen her bir yöntem, farklı kullanım senaryolarına uygun stratejiler sunarak, duygu analizi sistemlerinin performansını optimize etmek için önemli katkılar sağlamaktadır. Bu bulgular, duygu yoğunluğu analizine yönelik daha kapsamlı ve kullanıcı dostu sistemlerin geliştirilmesi için gelecek çalışmalara rehberlik edebilecek niteliktedir.

5.2. Araştırma Kapsamındaki Sınırlılıklar

Duygu analizi çalışmalarında, model eğitimi ve test süreçlerinde karşılaşılan bazı sınırlılıklar, araştırmanın kapsamını ve elde edilen sonuçların genelleştirilebilirliğini etkilemiştir. Bu sınırlılıklar aşağıda detaylandırılmıştır:

Teknik Ekipman Yetersizliği; Model eğitimi sırasında yüksek işlem gücü gerektiren derin öğrenme algoritmaları ve görüntü işleme teknikleri, güçlü donanım gereksinimlerini beraberinde getirmektedir. Ancak bu araştırmada kullanılan teknik ekipmanların sınırlı işlem kapasitesi, model eğitimi sırasında işlem yükünün artırılmasına ve eğitim sürecinin uzamasına neden olmuştur. Özellikle GPU tabanlı işlemcilerin eksikliği, daha karmaşık modellerin denenmesini ve optimize edilmesini sınırlandırmıştır.

Sunucu Tabanlı Çalışma Ortamı Yetersizliği; Araştırma sürecinde sürekli erişim sağlanabilen bir sunucu altyapısının eksikliği, model eğitim ve test süreçlerinde önemli sınırlamalara yol açmıştır. Bulut tabanlı platformlar, özellikle Google Colab gibi çevrimiçi çalışma ortamları, işlem gücü gereksinimlerini karşılamak açısından avantaj sağlasa da bu platformların bazı kısıtlamaları da bulunmaktadır.

Colab ortamında çalışırken, yerel ekipmanlara (örneğin, kameralar, mikrofonlar veya diğer donanımlar) doğrudan erişim sağlanamaması, gerçek zamanlı testlerin gerçekleştirilmesini zorlaştırmıştır. Bu sorun, özellikle duygu analizi modellerinin kamera verileriyle çalıştığı senaryolarda öne çıkmıştır. Verilerin platforma yüklenmesi ve işlenmesi sırasında ortaya çıkan gecikmeler, gerçek zamanlı analizlerin etkinliğini azaltmış ve süreci karmaşıklştırmıştır.

Ayrıca, Colab gibi platformların çalışma süresi sınırlamaları ve bağlantı kesilme sorunları, sürekli eğitim ve test süreçlerini kesintiye uğratmıştır. Bu sınırlamalar, bulut tabanlı altyapıların kullanımını sınırlandırmış ve modelin performansını daha kapsamlı bir şekilde değerlendirme imkanını kısıtlamıştır.

Bu nedenle, sunucu tabanlı çalışma ortamlarının yetersizliği hem donanım erişimi hem de süreklilik açısından önemli bir dezavantaj oluşturmuş ve araştırmanın kapsamını doğrudan etkilemiştir. Gelecekteki çalışmalarda hem güçlü sunucu altyapılarının hem de

yerel donanımlara erişim imkânı sunan hibrit çözümlerin kullanımı, bu tür sınırlamaları aşmak için önemli bir adım olarak değerlendirilmektedir.

Gerçek Zamanlı Test İçin Ortam Yetersizliği; Gerçek zamanlı duygu analizi testleri için ideal koşulların oluşturulması, araştırma kapsamında zorluk teşkil eden bir diğer unsur olmuştur. Test senaryolarının, doğal ortamları temsil eden verilerle gerçekleştirilmesi hedeflenmiş olsa da toplantı, seminer veya ders ortamları gibi gerçek dünyadaki kullanım alanlarına erişimde yaşanan sınırlamalar, gerçek zamanlı analizlerin kapsamını daraltmıştır. Bu durum, modelin gerçek dünyadaki performansını değerlendirme sürecinde sınırlamalara yol açmıştır.

Gerçek Zamanlı Uygulamalarda Entegrasyon Zorlukları; Duygu analizi modellerinin gerçek zamanlı uygulamalara entegrasyonu, özellikle toplantı veya video konferans uygulamaları gibi platformlarla uyumluluk açısından zorluklar barındırmıştır. Uygulama programlama arayüzü (API) kısıtlamaları ve entegrasyon süreçlerindeki teknik problemler, modellerin bu tür uygulamalarda etkili bir şekilde çalışmasını sınırlandırmıştır. Ayrıca, entegrasyon süreci için gereken zaman ve kaynak eksikliği, çalışmanın bu alandaki potansiyelini tam anlamıyla ortaya koymasını engellemiştir.

Zaman Problemleri ve Çalışma Koşulları; Model geliştirme ve test süreçlerinde zaman yönetimi, sınırlayıcı bir faktör olarak öne çıkmıştır. Sürekli çalışmaya olanak sağlamayan altyapı ve sınırlı kaynaklar, araştırmacıların model eğitimi ve testleri arasında uzun süre beklemelerine yol açmıştır.

Sonuç olarak, bu sınırlamalar, araştırma bulgularının daha kapsamlı bir şekilde ele alınmasını engellemiş olsa da bu zorlukların üstesinden gelinmesi, gelecekteki çalışmalar için değerli bir çerçeve sağlayabilir. Daha güçlü donanım, sürekli çalışma imkânı sunan altyapılar ve entegrasyon süreçlerini kolaylaştıran platformlar, bu tür çalışmaların etkinliğini artıracak temel unsurlar olarak değerlendirilmektedir.

5.3. Gelecek Çalışmalar İçin Öneriler

Gelecek çalışmalarda, duygu analizi alanında hem model geliştirme hem de uygulama çeşitliliğini artırmaya yönelik birçok fırsat bulunmaktadır. Öncelikle, mevcut duygu analizi modellerinin gerçek zamanlı ve çevrimdışı senaryolarda daha verimli hale

getirilmesi için optimize edilmesi planlanmaktadır. Bunun yanı sıra, duygu analizi yöntemlerinin, farklı uygulama alanlarında genişletilmesi için çalışmalar yürütülebilir.

Toplantılar sırasında kaydedilen seslerden konuları otomatik olarak tanımlayarak, hangi konu konuşulurken ortamın genel duygu durumunun nasıl değiştiği analiz edilebilir. Bu yaklaşım, yalnızca yüz ifadelerinden değil, aynı zamanda konuşma içeriğinden de anlam çıkararak daha kapsamlı bir duygu analizi yapılmasına olanak tanır. Örneğin, bir toplantıda tartışılan ana konulara (örneğin, proje ilerleme durumu, sorunlar veya başarılar) göre katılımcıların duygusal tepkilerinin haritalanması, liderler ve ekip üyeleri için anlamlı içgörüler sağlayabilir.

Duygu analizinin gerçek zamanlı olarak uygulanabilirliğini artırmak için, daha hızlı ve daha az kaynak tüketen modellerin geliştirilmesi önemlidir. Bunun yanında, gerçek zamanlı duygu analizi ile entegre çalışan raporlama araçları geliştirilerek, toplantılar veya eğitim ortamlarında anlık duygu durumlarının görselleştirilmesi sağlanabilir. Bu raporlar, katılımcıların duygu değişimlerini izlemek ve belirli konuşma anlarının duygusal etkilerini ölçmek için kullanılabilir.

Duygu analizi araçlarının daha geniş bir kullanıcı kitlesine ulaşması için, kolayca uygulanabilir ve otomatikleştirilmiş çözümler geliştirilmesi önemlidir. Bu bağlamda, düşük maliyetli cihazlar ve bulut tabanlı hizmetlerle entegre çalışan uygulamalar, küçük ve orta ölçekli işletmelerde de yaygın bir şekilde kullanılabilir hale gelebilir.

6. SONUÇLAR VE ÖNERİLER

6.1. Araştırmanın Genel Sonuçları

Çalışmamızda gerçekleştirilen araştırmalar, geliştirmeler ve deneyler sonucunda belirli bir doğrulukla tespit edilebilen yüzler üzerinde duygu analizi yapılabildiği görülmüştür. Kullanılan yöntemler değişiklik gösterse de sonuçlar üzerindeki etkisini kullanılan yöntemlerin uygulanması ve kaynakların yeterliliğini değiştirdiği görülmüştür. Bu alanda faaliyet gösteren firmalar ve teknoloji ekipleri tarafından geliştirilen modellerin performanslarına ulaşmak için geliştirdiğimiz modelimizin benzer başarılar verdiği görüşmüştür.

Model performanslarından sonra ortam duygu yoğunluğu analizi için elde edilen verilerin nasıl bir matematiksel işlemler dizininden geçeceği önemli bir sorundur. Bu sorunun çözümü performans, verimlilik, kaynakların doğru kullanımı ve doğruluk açısından önemlidir.

Geliştirdiğimiz ortam duygu analizi hesaplama yöntemlerinin gösterdiği üzerin her yöntemin kendi üzerinde avantajları ve dezavantajları var. Bu durum bize yöntemlerin nasıl kullanılacağı hakkında bilgi birikimi kazandırmıştır. Aynı zamanda çalışmamız bize yeni fikirler ve yapay zekâ alanında yeni deneyimler kazandırmıştır.

6.2. Öneriler ve Uygulamalar

Duygu analizini daha güçlü hale getirmek için farklı yöntemleri entegre ederek duygu analizi sonuçlarını daha güvenilir hale getirebiliriz. Bunu için aşağıdaki yöntemler ile entegre edilerek her açıdan bir ölçüm sağlanabilir.

Metin Analiziyle Entegrasyon: İnsanların yüz ifadeleri kadar yazılı ifadeleri de duygularını yansıtır. Metin tabanlı duygu analizinde NLP (Natural Language Processing) teknikleri kullanılarak sosyal medya gönderileri, yorumlar veya mesajlar analiz edilebilir. Yüz ifadelerinden elde edilen duygu tahmini ile metin analizi birleştirilebilir.

Uygulama örneği: Bir video konferans uygulamasında konuşma sırasında yüz ifadeleriyle tespit edilen duygu analizi, kullanıcıların yazdığı mesajlarla desteklenebilir.

Ses Analiziyle Entegrasyon: Ses tonu ve konuşma hızı, duyguların belirlenmesinde kritik rol oynar. Sen sinyalleri üzerinden duygu analizi yapmak için MFCC (Mel-Frequency Cepstral Coefficients) veya spectrogram tabanlı CNN modelleri kullanılabilir. Çağrı merkezi analizi; konuşmacının yüz ifadeleri ile ses tonu entegre edilerek müşteri memnuniyeti tahmin edilebilir.

Fizyolojik Veriyle Entegrasyon: Kalp atışı, cilt direnci gibi biyometrik veriler, duyguları anlamada yardımcı olabilir. Akıllı saat veya giyilebilir cihazlardan alınan biyometrik veriler yüz ifadeleriyle birleştirilebilir.

Uygulama örneği: Sağlık sektöründe, stres yönetimi veya duygu durum bozukluklarının teşhisi için kullanabilirsiniz.

Yukarıdaki yöntemleri bir araya getirerek daha güvenilir tahminler yapmak için multimodal analiz yaklaşımı kullanılabilir.

KAYNAKLAR

- Baheti, B., Innani, S., Gajre, S., Talbar, S. (2020). Eff-UNet: A Novel Architecture for Semantic Segmentation in Unstructured Environment. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops CVPRW*. Seattle, WA, USA, 14-19 Haziran 2020.
- Bradski, G., Kaehler, A. (2008). *Learning OpenCV: Computer Vision with the OpenCV Library* (1st ed.). Sebastopol: O'Reilly Media.
- Chan, A. A. S., Abdullah, M.F.L., Mustam, S., Poad, F. A., Joret, A. (2024). Face Detection with YOLOv7: A Comparative Study of YOLO-Based Face Detection Models. *International Conference on Green Energy, Computing and Sustainable Technology GECOST*, Miri Sarawak, Malaysia, 17-19 January 2024.
- Chaudhari, A., Bhatt, C., Krishna, A., Mazzeo, P. L. (2022). ViTFER: Facial Emotion Recognition with Vision Transformers. *Multidisciplinary Digital Publishing Institute. Appl. Syst. Innov.*, 5(4), 80. DOI: 10.3390/asi5040080.
- Clark, E. A., Kessinger, J., Duncan, S.E., Bell, M.A., Lahne, J., Gallagher, D.L., O'Keefe, S.F. (2020). The Facial Action Coding System for Characterization of Human Affective Response to Consumer Product-Based Stimuli: A Systematic Review. *Frontiers in Psychology*, 11(920). DOI: 10.3389/fpsyg.2020.00920.
- Coşkun, M., Uçar, A., Yıldırım, Ö., Demir, Y. (2017). Face recognition based on convolutional neural network. *International Conference on Modern Electrical and Energy Systems MEES*, Kremenchuk, Ukraine, 15-17 Kasım 2017.
- Dalal, N., Triggs, B. (2005). Histograms of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR'05*, San Diego, CA, USA 20-25 Haziran 2005.
- Dalvi, C., Rathod, M., Patil, S., Gite, S., Kotecha, K. (2021). A Survey of AI-Based Facial Emotion Recognition: Features, ML & DL Techniques, Age-Wise Datasets and Future Directions. *IEEE Access*, 9(11), 165806-165840. DOI: 10.1109/ACCESS.2021.3131733.
- D'incà, M., Beyan, C., Niewiadomski, R., Barattin, S., Sebe, N. (2023). Unleashing the Transferability Power of Unsupervised Pre-Training for Emotion Recognition in Masked and Unmasked Facial Images. *IEEE Access*, 11(8), 90876-90890. DOI: 10.1109/ACCESS.2023.3308047.
- Durand, K., Gallay, M., Seigneure, A., Robichon, F., Baudouin, J.Y. (2007). The development of facial emotion recognition: The role of configural information. *Journal of Experimental Child Psychology*, 97(1), 14-27. DOI: 10.1016/j.jecp.2006.12.001.

- Ekman, P. and Friesen, W.V. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto: Consulting Psychologists Press.
- Ekman, P., Friesen, W. V., & Hager, J. C. (2002). Facial Action Coding System. *Manual and Investigator's Guide*, Salt Lake City, UT: Research Nexus.
- Ezerceli, Ö., Eskil, M.T. (2022). Convolutional Neural Network (CNN) Algorithm Based Facial Emotion Recognition (FER) System for FER-2013 Dataset. *International Conference on Electrical, Computer, Communications and Mechatronics Engineering ICECCME*. Maldives, Maldives 16-18 Kasım 2022.
- Guo, G., Zhang, N. (2018). What Is the Challenge for Deep Learning in Unconstrained Face Recognition? *IEEE International Conference on Automatic Face & Gesture Recognition*, Xi'an, China, 15-19 Mayıs 2018.
- Gupta, A. (2024). *Global Emotion Detection and Recognition Market Overview*. Marketresearchfuture, <https://www.marketresearchfuture.com/reports/emotion-detection-recognition-market-3193>. (Ziyaret tarihi: 14.12.2023).
- Gursesli, M. C., Lombardi, S., Duradoni, M., Bocchi, L., Guazzini, A., Lanata, A. (2024). Facial Emotion Recognition (FER) Through Custom Lightweight CNN Model: Performance Evaluation in Public Datasets. *IEEE Access*, 12, 45543 – 45559. DOI: 10.1109/ACCESS.2024.3380847.
- Hasan, K., Ahsan, S., Al-Mamun, A., Newaz, S. H. S., Lee, G.M. (2021). Human Face Detection Techniques: A Comprehensive Review and Future Research Directions. *Electronics*, 10(19), 2354. DOI: 10.3390/electronics10192354.
- Hjortsjo, C.-H. (1970). *Man's face and mimic language*, (2nd ed.). Malmö: Nordens Boktryckeri.
- Hossain, S., Umer, S., Rout, R. K., Marzouqi, H. A. (2024). A Deep Quantum Convolutional Neural Network Based Facial Expression Recognition For Mental Health Analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32, 1556-1565. DOI: 10.1109/TNSRE.2024.3385336.
- Huang, J., Kang, H. (2024). Automatic Defect Detection in Sewer Pipe Closed-Circuit Television Images via Improved You Only Look Once Version 5 Object Detection Network. *IEEE Access*, 12, 92797 – 92825. DOI: 10.1109/ACCESS.2024.3422275.
- Huo, Y., Ge, Y. (2024). Multi-Task Emotion Recognition Based on Dimensional Model and Category Label. *IEEE Access*, 12, 75169- 75179. DOI: 10.1109/ACCESS.2024.3404990.
- Jain, D.K., Shamsolmoali, P., Sehdev, P. (2019). Extended deep neural network for facial emotion recognition. *Pattern Recognition Letters*, 120, 69-74. DOI: 10.1016/j.patrec.2019.01.008.

- Jain, V., Learned-Miller, E. (2010). Fddb: A Benchmark for Face Detection in Unconstrained Settings, *Dept. of Computer Science, University of Massachusetts, Amherst*, UM-CS-2010-009.
- Ji, Y., Dong, S. Y. (2024). Multi-Task Learning by Leveraging Non-Contact Heart Rate for Robust Facial Emotion Recognition. *IEEE Access*, 12, 92175 – 92180. DOI: 10.1109/ACCESS.2024.3422403.
- Kahou, S. E., Pal, C., Bouthillier, X., Froumenty, P., Gülçehre, Ç., Memisevic, R., Vincent, P., Courville, A., Bengio, Y., Ferrari, R. C., Mirza, M., Jean, S., Carrier, P.-L., Dauphin, Y., Lewandowski, N. B., Aggarwal, A., Zumer, J., Lamblin, P., Raymond, J.-P., Desjardins, G., Pascanu, R., Farley, D. W., Torabi, A., Sharma, A., Bengio, E., Côté, M., Konda, K. R., Wu, Z. (2013). Combining modality specific deep neural networks for emotion recognition in video. *15th ACM on International conference on multimodal interaction*, Sydney, Australia, 9 – 13 Aralık 2013.
- Khalil, M. I., Tehsin, S., Humayun, M., Jhanjhi, N. Z., AlZain, M. A. (2022), Multi-Scale Network for Thoracic Organs Segmentation, *Computers, Materials & Continua*, 70(2), 3251-3265. DOI: 10.32604/cmc.2022.020561.
- Kiniz, F. (2016). Deep Learning Lab Episode-3: Fer2013, https://medium.com/@birdortyedi_23820 (Ziyaret tarihi: 14.12.2023).
- Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90. <https://doi.org/10.1145/3065386>.
- Kaziakhmedov, E., Kireev, K., Melnikov, G., Pautov, M., Petiushko, A. (2019). Real-world Attack on MTCNN Face Detection System. *International Multi-Conference on Engineering, Computer and Information Sciences SIBIRCON*, Novosibirsk, Russia, 21-27 Ekim 2019.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation* 1(4), 541-551. DOI: 10.1162/neco.1989.1.4.541.
- Li, S.Z., Jain, A.K. (2011). Handbook of Face Recognition. (2nd ed.). Springer, London.
- Li, T. H. S., Kuo, P. H., Tsai, T. N., Luan, P. C. (2019). CNN and LSTM Based Facial Expression Analysis Model for a Humanoid Robot. *IEEE Access*, 7, 93998 – 94011. DOI: 10.1109/ACCESS.2019.2928364.
- Li, X., Lai, S., Qian, X. (2021). DBCFace: Towards Pure Convolutional Neural Network Face Detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(4), 1792-1804. DOI: 10.1109/TCSVT.2021.3082635

- Li, X., Yang, Z., Wu, H. (2020). Face Detection Based on Receptive Field Enhanced Multi-Task Cascaded Convolutional Neural Networks. *IEEE Access*, 8, 174922 – 174930. DOI: 10.1109/ACCESS.2020.3023782.
- Lioussaief, S., Abdelkrim, A. (2016). Machine learning framework for image classification. *7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications SETIT*, Hammamet, Tunisia, 18-20 Aralık 2016.
- Liu, Z., Wu, M., Cao, W., Chen, L., Xu, J., Zhang, R. (2017). A facial expression emotion recognition based human-robot interaction system. *IEEE/CAA Journal of Automatica Sinica*, 4(4), 668 – 676. DOI: 10.1109/JAS.2017.7510622.
- Liu, Z., Luo, P., Wang, X., Tang, X. (2015). Deep Learning Face Attributes in the Wild. *IEEE International Conference on Computer Vision ICCV*, Santiago, Chile, 07-13 Aralık 2015.
- Matuska, S., Hudec, R., Benco, M. (2012). The comparison of CPU time consumption for image processing algorithm in Matlab and OpenCV. *2012 ELEKTRO*, Rajecke Teplice, Slovakia, 21-22 Mayıs 2012.
- Mollahosseini, A., Hasani, B., Mahoor, M. H. (2019). AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. *IEEE Transactions on Affective Computing*, 10(1), 18 – 31. DOI: 10.1109/TAFFC.2017.2740923.
- Perez-Gaspar, L.-A., Caballero-Morales, S.-O., Trujillo-Romero, F. (2016). Multimodal emotion recognition with evolutionary computation for human-robot interaction. *Expert Systems with Applications*, 66, 42-61. DOI: 10.1016/j.eswa.2016.08.047.
- Phillips, P.J., Wechsler, H., Huang, J., Rauss, P. J. (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5), 295-306. DOI: 10.1016/S0262-8856(97)00070-X.
- Rodriguez, R., Abisado, M., Montefalcon, M. D., Padilla, J. R., Serrano, E., Dasmariñas, N. U. (2023). Classifying Emotions of Digital Immigrants while using Software Application based on Facial Features. *11th International Conference on Information and Education Technology ICIET*, Fujisawa, Japan, 18-20 Mart 2023.
- Shawar, B. A., Atwell, E. S. (2005). Using corpora in machine-learning chatbot systems. *International Journal of Corpus Linguistics*, 10(4), 489 – 516. DOI: 10.1075/ijcl.10.4.06sha.
- Sharma, S., Bhatt, M., Sharma, P. (2020). Face Recognition System Using Machine Learning Algorithm. *5th International Conference on Communication and Electronics Systems ICCES*, Coimbatore, India, 10-12 Haziran 2020.

- Sirovich, L., Kirby, M. (1987). Low-dimensional Procedure for the Characterization of Human Faces. *Journal of the Optical Society of America A*, 4(3), 519-524. DOI: 10.1364/JOSAA.4.000519.
- Souaidi, M., El Ansari, M. (2022). A New Automated Polyp Detection Network MP-FSSD in WCE and Colonoscopy Images Based Fusion Single Shot Multibox Detector and Transfer Learning. *IEEE Access*, 10, 47124 – 47140. DOI: 10.1109/ACCESS.2022.3171238.
- Stilgoe, J. (2017). Machine Learning, Social Learning and the Governance of Self-Driving Cars. *Studies of Science*, 48(1), 25 – 56. DOI: 10.1177/0306312717741687.
- Tan, M., Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *International Conference on Machine Learning (ICML 2019)*, Long Beach, California, USA. 9-15 Haziran 2019.
- Tariyal, S., Chauhan, R., Bijalwan, Y., Rawat, R., & Gupta, R. (2024). A Comparative Study of MTCNN, Viola-Jones, SSD and YOLO Face Detection Algorithms. *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, Bangalore, India, 24-25 Ocak 2024.
- Viola, P., Jones, M. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR*, Kauai, HI, USA, 8-14 Aralık 2001.
- Wang, H.-H., Gu, J.-W. (2018). The Applications of Facial Expression Recognition in Human-computer Interaction. *IEEE International Conference on Advanced Manufacturing ICAM*, Yunlin, Taiwan, 16-18 Kasım 2018.
- Wang, J., Yang, H. (2008). Face Detection Based on Template Matching and 2DPCA Algorithm. *Congress on Image and Signal Processing, Sanya, China*, 27-30 Mayıs 2008.
- Wang, S., Qu, J., Zhang, Y., Zhang, Y. (2023). Multimodal Emotion Recognition From EEG Signals and Facial Expressions. *IEEE Access*, 11, 33061 – 33068. DOI: 10.1109/ACCESS.2023.3263670.
- Wang, W., Zhang, Y., Liu, L., Shao, Y., Zhang, L., Deng, N. (2021). Robust 2DPCA by Tl_1 Criterion Maximization for Image Recognition. *IEEE Access*, 9, 7690 – 7700. DOI: 10.1109/ACCESS.2021.3049535.
- Wang, W., Zhang, Y., Yan, S., Zhang, Y., Jia H. (2012). Parallelization and performance optimization on face detection algorithm with OpenCL: A case study. *Tsinghua Science and Technology*, 17(3), 287-295. DOI: 10.1109/TST.2012.6216758.
- Xiang, Z., Tan, H., & Ye, W. (2018). Dense grid histograms of oriented gradients for face recognition. *IEEE Access*, 6, 29306-29319. DOI: 10.1109/ACCESS.2018.2813395.

- Xiao, Y., Li, Z., Zhang, D., Teng, L. (2021). Detection of Pin Defects in Aerial Images Based on Cascaded Convolutional Neural Network. *IEEE Access*, 9, 73071 – 73082. DOI: 10.1109/ACCESS.2021.3079172.
- Yang, J., Zhang, X., Lei, Z., Li, S.Z. (2014). The Fastest Deformable Part Model for Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 23-28 Haziran 2014.
- Yang, S., Luo, P., Loy, C.C., Tang, X. (2016). WIDER FACE: A Face Detection Benchmark. *IEEE Conference on Computer Vision and Pattern Recognition CVPR*, Las Vegas, NV, USA, 27-30 Haziran 2016.
- Zeng, G., Zhou, J., Jia, X., Xie, W., Shen, L. (2018). Hand-Crafted Feature Guided Deep Learning for Facial Expression Recognition. *13th IEEE International Conference on Automatic Face & Gesture Recognition FG 2018*, Xi'an, China, 15-19 Mayıs 2018.
- Zhu, X., Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild. *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, 16-21 Haziran 2012.
- Zhou, N., Liang, R., Shi, W. (2020). A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection. *IEEE Access*, 9, 5573 – 5584. DOI: 10.1109/ACCESS.2020.3046715.



EKLER

Ek-A

Kullanılan Algoritmaların Teknik Detayları

EfficientNetB7, derin öğrenme tabanlı görüntü sınıflandırma modelleri arasında öne çıkan bir yapıdır. Bu model, Google tarafından geliştirilen ve Convolutional Neural Networks (CNNs) mimarilerinin verimliliğini artırmayı hedefleyen EfficientNet ailesinin en büyük ve en karmaşık üyesidir (Tan ve Le, 2019). EfficientNetB7'nin temel teknik özellikleri ve avantajları şunlardır:

EfficientNet modellerinin başarısı, "Compound Scaling" adı verilen özel bir ölçeklendirme stratejisine dayanır.

- Geleneksel modellerde, sadece ağın derinliği (depth) veya genişliği (width) artırılırken, EfficientNet bu iki özelliği ve görüntü çözünürlüğünü (resolution) dengeli bir şekilde artırır.
- EfficientNetB7 modeli, derinlik, genişlik ve çözünürlüğün optimize edilmiş bir kombinasyonunu sunar, bu da daha yüksek doğruluk oranı elde edilmesini sağlar.

EfficientNetB7, MobileNetV2'nin taban bloğu olan inverted residual block ve depthwise separable convolution yapılarını kullanır.

- Depthwise Separable Convolution: Geleneksel konvolüsyon işlemlerine kıyasla hesaplama yükünü önemli ölçüde azaltır.
- Inverted Residual Block: Daha verimli bilgi aktarımını sağlayarak modeli derinleştirir ve doğruluğu artırır.

EfficientNetB7, yaklaşık 66 milyon parametre ve 400 MB model boyutuna sahiptir.

- Bu boyut, daha küçük EfficientNet modellerine kıyasla yüksek doğruluk için daha fazla kapasite sunar, ancak daha fazla hesaplama gücü gerektirir.

EfficientNetB7, FER-2013 gibi küçük boyutlu gri tonlamalı yüz ifadeleri içeren veri setlerinde başarılı şekilde uyarlanabilir.

- Transfer Learning: Pretrained (önceden eğitilmiş) EfficientNetB7 modelinin, ImageNet gibi büyük veri setleri üzerinde öğrenilmiş özellikleri, duygu analizi gibi özelleşmiş bir görevde yeniden kullanılabilir.
- Input Boyutu: FER-2013 için gri tonlamalı görüntüler 48x48 piksel olduğundan, bu görüntüler EfficientNetB7 için RGB formatına dönüştürülerek uygun çözünürlükte yeniden boyutlandırılır (genellikle 224x224 piksel).

EfficientNetB7, diğer büyük CNN mimarilerine kıyasla daha az hesaplama yüküyle yüksek doğruluk elde eder:

- Daha az parametreyle daha iyi performans.
- Özellikle yüz ifadesi analizi gibi görevlerde yüksek duyarlılık ve doğruluk oranları sunar.

EfficientNetB7, yüksek doğruluk gerektiren aşağıdaki görevlerde sıklıkla tercih edilir:

- Yüz ifadelerinden duygu analizi
- Görüntü sınıflandırma.
- Nesne algılama ve tanıma.

ConvLSTM (Convolutional Long Short-Term Memory), zaman serisi verilerini veya uzamsal-temporal (spatio-temporal) verileri modellemek için geliştirilmiş bir derin öğrenme mimarisidir. Bu yapı, özellikle video analizi, hareket tahmini ve uzamsal-temporal duygu analizi gibi alanlarda kullanılır. ConvLSTM, klasik LSTM'nin sırayla çalışan hücre yapılarını, konvolüsyonel katmanlarla birleştirerek hem uzamsal hem de zamansal verileri işleyebilir.

ConvLSTM, geleneksel LSTM hücrelerinin yerine konvolüsyonel işlemleri yerleştirir.

- LSTM Hücresi: LSTM, uzun süreli bağımlılıkları öğrenme yeteneğiyle bilinen bir tekrarlayan sinir ağı (RNN) mimarisidir. ConvLSTM, bu hücreleri daha etkili hale getirmek için uzamsal veri işleme özelliği ekler.
- Konvolüsyonel Kapılar: Giriş, unutma ve çıktı kapıları (gates), matris çarpımları yerine konvolüsyonel işlemlerle gerçekleştirilir.

- Uzamsal-Temporal Veriler: ConvLSTM, uzamsal veri (ör. görüntü kareleri) ve zamansal veri (ör. bir video dizisi) kombinasyonunu işleyebilir.

ConvLSTM, FER-2013 gibi yüz ifadeleri analizi için kullanılan bir veri seti üzerinde uygulanabilir.

- Zamansal Dizi Verileri: Eğer yüz ifadeleri zaman içinde değişiklik gösteriyorsa (örneğin, video kareleri), ConvLSTM bu tür verileri işlemek için uygundur.
- Uzamsal Özellikler: FER-2013 veri setindeki her bir yüz görüntüsünden uzamsal özellikler (kenar, yapı, doku) çıkarılarak ConvLSTM tarafından işlenebilir.

ConvLSTM hem uzamsal hem de zamansal bilgileri aynı anda öğrenme yeteneğine sahiptir:

- Uzamsal Bilgiler: Görüntüdeki yerel özellikleri (ör. yüz yapısı, göz hareketleri) konvolüsyonel katmanlarla öğrenir.
- Zamansal Bağlantılar: Bir video sekansındaki ardışık karelerdeki duygu değişimlerini analiz eder.
- ConvLSTM, geleneksel LSTM'lere kıyasla daha fazla parametre içerir çünkü her hücrede konvolüsyonel filtreler kullanılır.
- Girdi boyutu ve hücre sayısı, modelin kapasitesini ve karmaşıklığını belirler.

ConvLSTM, duygu analizi gibi spatio-temporal görevlerde aşağıdaki avantajları sunar:

- Zaman Serisi Tahmini: Video tabanlı duygu analizi için uygundur.
- Uzamsal Doğruluk: Görüntüden özellik çıkarmada başarılıdır.
- Zamansal Tutarlılık: Bir duygu durumunun değişimlerini tutarlı bir şekilde analiz eder.

ConvLSTM aşağıdaki uygulama alanlarında kullanılabilir:

- Video Duygu Analizi: Video kayıtlarındaki yüz ifadelerinin zaman içindeki değişimlerini analiz eder.
- Hareket Tahmini: İnsan hareketlerinin analizine entegre edilebilir.
- Sağlık ve Psikoloji: Uzun süreli duygu durumlarını takip etmek için kullanılabilir.
- Reklam ve Medya: Video reklamlarına kullanıcı tepkisini ölçmek için kullanılabilir.

KİŞİSEL YAYIN VE ESERLER

Özcan, H., **Yıldırım, İ.** (2024). Analysis and Methods for Environmental Emotion Density Using Image Processing and Artificial Intelligence, *International Yeditepe Scientific Research Congress*, Çevrimiçi, Türkiye, 7-8 Eylül 2024.



ÖZGEÇMİŞ

İlk, orta ve lise öğrenimini Erzincan'da tamamladı. 2013 yılında girdiği Kocaeli Üniversitesi Kocaeli Mühendislik Fakültesi Mekatronik Mühendisliği Bölümü'nden 2019 yılında Mekatronik Mühendisi olarak mezun oldu. 2019-2020 yılları arasında Experto firmasında teknik uzman olarak görev almıştır. 2021 ve 2022 arasında Arge mühendisi olarak Akratek firmasında görev almıştır. 2022'den beri MetropolCard firmasında yazılım geliştirici görevine ve 2023 yılı Bahar yarıyılında başladığı Kocaeli Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği yüksek lisans öğrenimine devam etmektedir.

