TOBB EKONOMİ VE TEKNOLOJİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

JEST VE MİMİKLERDEN YAPAY SİNİR AĞLARI İLE DUYGU SINIFLANDIRMA

YÜKSEK LİSANS TEZİ

Büşra KARATAY

Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı: Doç. Dr. Fatma Betül ATALAY SATOĞLU

Eş Danışman: Prof. Dr. Tansel ÖZYER

NİSAN 2022

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, alıntı yapılan kaynaklara eksiksiz atıf yapıldığını, referansların tam olarak belirtildiğini ve ayrıca bu tezin TOBB ETÜ Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırlandığını bildiririm.

Büşra KARATAY

ÖZET

Yüksek Lisans Tezi

JEST VE MİMİKLERDEN YAPAY SİNİR AĞLARI İLE DUYGU SINIFLANDIRMA

Büşra KARATAY

TOBB Ekonomi ve Teknoloji Üniveritesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Fatma Betül ATALAY SATOĞLU

Eş Danışman: Prof. Dr. Tansel ÖZYER

Tarih: Nisan 2022

Metin, resim, video ve konuşma gibi farklı veri kaynaklarından doğru duyguyu sınıflandırmak, çeşitli disiplinlerden araştırmacılar için ilham verici bir alanı olmuştur. Videolardan ve fotoğraflardan otomatik duygu algılama, denetimli ve denetimsiz makine öğrenimi yöntemleri kullanılarak üzerinde çalışılan zorlu konulardan biridir. Bu tez çalışmasında bir takım ön işleme adımları ve yeni bir derin öğrenme mimarisi ile videolardan duygu analizi yöntemi sunulmaktadır. Videolardan OpenPose aracı kullanılarak elde edilen yüz ve vücut pozisyon bilgileri modellerde kullanılmak üzere poz tanımlayıcılara dönüştürüldü, ardından LSTM ve Dönüştürücü modelleri bu veri ile eğitilerek performansları karşılaştırıldı. Ardından LSTM ve Dönüştürücü modellerine bir CNN bloğu ön katman olarak eklenmiş poz tanımlayıcılarla beslenen CNN bloğunun çıktısı LSTM ve Dönüştürücü modelleri için girdi olarak kullanıldı. Model doğruluklarını iyileştirmek amacı ile Video Çoklama, Anahtar Kare Seçimi ve Gauss Karışım Merkezi yaklaşımları ön işleme adımları olarak eklenmiş ve deneyler bu farklı yaklaşımların kombinaysonları için tekrarlandı. Yapılan kapsamlı deneylerin ardından sonuçlar karşılaştırıldı ve önerilen iki katmanlı sınıflandırıcı yapısı ve ön

işleme adımlarının etkileri gözlemlendi. Sonuçlar ayrıca aynı veri kümesini kullanan güncel, yüksek doığruluk oranlarına sahip diğer yöntemlerle de karşılaştırıldı. FABO ve CK+ olmak üzere iki yaygın veri kümesi kullanılarak gerçekleştirilen deneyler, FABO veri seti için video çoklama uygulanmış CNN-Dönüştürücü yapısının %99 doğruluk oranı ile, diğer modellerden daha iyi bir performansa sahip olduğunu gösterdi. Her iki veri kümesi için de bir çok versiyonda önerilen model %90 üzerinde doğruluğa ulaşarak kayda değer başarımlar elde etti.

Anahtar Kelimeler: Duygu sınıflandırma, Video analizi, Görüntü işleme, Yapay sinir ağları, Derin öğrenme.

ABSTRACT

Master of Science

EMOTION CLASSIFICATION WITH ARTIFICIAL NEURAL NETWORKS FROM FACIAL EXPRESSIONS AND GESTURES

Büşra KARATAY

TOBB University of Economics and Technology
Institute of Natural and Applied Sciences
Department of Computer Engineering

Supervisor: Doç. Dr. Fatma Betül ATALAY SATOĞLU

Co-Supervisor: Prof. Dr. Tansel ÖZYER

Date: April 2022

Classifying the right emotion from different data sources such as text, images, video, and speech has been an inspiring field for researchers from various disciplines. Automatic emotion detection from videos and photos is one of the challenging topics being studied using supervised and unsupervised machine learning methods. In this thesis, several preprocessing steps and a new deep learning architecture and emotion classification method from videos are presented. The face and body position information obtained from the videos using the OpenPose tool was converted into pose descriptors for use in the models, then the LSTM and Transformer models were trained with this data and their performances were compared. Then, the output of the CNN block fed with pose descriptors was used as input for the LSTM and Transformer models. Video Generation, Keyframe Selection, and Gaussian Mixture Center approaches were added as preprocessing steps to improve model accuracy and the experiments were repeated for combinations of these different approaches. After extensive experiments, the results were compared and the effects of the proposed two-layer classifier structure and preprocessing steps were observed. Results were also

compared with other recent, high-accuracy methods using the same dataset. Experiments using two common datasets, FABO and CK+, showed that the CNN-Transformer structure with video generation approach for the FABO dataset outperforms the other models, with an accuracy of %99. For both datasets, the proposed method in many versions achieved remarkable success, reaching an accuracy of over %90.

Keywords: Emotion classification, Video analysis, Image processing, Neural networks, Deep learning.

TEŞEKKÜR

Çalışmalarım boyunca değerli yardım ve katkılarıyla beni yönlendiren hocalarım Prof. Dr. Tansel Özyer ve Doç. Dr. Fatma Betül Atalay Satoğlu'na, kıymetli tecrübelerinden yararlandığım, bana her konuda yardımcı olan TOBB Ekonomi ve Teknoloji Üniversitesi Bilgisayar Mühendisliği Bölümü öğretim üyelerine, lisans ve yüksek lisans eğitimim boyunca bana burs sağlayan TOBB Ekonomi ve Teknoloji Üniversitesi'ne ve desteklerini hiç esirgemeyen biricik ailem ve arkadaşlarıma çok teşekkür ederim. Bu araştırmada yer alan tüm/kısmi nümerik hesaplamalar TÜBİTAK ULAKBİM, Yüksek Başarım ve Grid Hesaplama Merkezi'nde (TRUBA kaynaklarında) gerçekleştirilmiştir. Katkılarından dolayı teşekkürler.

İÇİNDEKİLER

ŞEKİL LİSTESİ

Sayfa

RESİM LİSTESİ

Sayfa

ÇİZELGE LİSTESİ

Sayfa

KISALTMALAR

CK+ : Genişletilmiş cohn-kanade veri kümesi

CNN : Evrişimli Yapay Sinir Ağları (Convolutional Neural Networks)

FABO : Sözsüz, otomatik duygu analizi için yüz ve vücut hareketleri veritabanı

LSTM : Uzun Kısa Vadeli Bellek (Long Short-term Memory)

GKM : Gauss Karışım Merkezi

RNN : Özyineli Sinir Ağları (Recurrent Neural Networks)

ReLU : Doğrultulmuş Doğrusal Birim (Rectified Linear Unit)

YSA : Yapay Sinir Ağları (Artificial Neural Networks)

1. GİRİŞ

Evrimsel, biyolojik, nörolojik, sosyal, psikolojik, ruhsal ve diğer bir çok farklı açıdan çeşitli alanlardan akademisyenler tarafından "Duygu" terimi tanımsal olarak yıllardır araştırılmaktadır. Bir bireyin davranış ve ifadelerinden duygu tahmini, duyguların çok yönlü oluşu sebebiyle karmaşık bir işlemdir [1]. Kültür, cinsiyet, etnik köken, konum, kişilik, çevre, durum, olay ve diğer sosyal, zamansal, mekansal ve kişisel unsurlar duyguların ifade ediliş şeklini etkileyebilir. Kişiler birden fazla duyguyu aynı anda hissedebilir ve bunların hepsini kelimelerle veya jest ve mimikleriyle ifade etmekte zorlanabilir. Bunun yanında günümüz teknolojisi sayesinde sanal iletişim ve çevrimiçi sosyal ağların kullanımının yaygınlaşmasıyla birlikte her yaş ve kesimden insan yazılı, sesli veya görüntülü olarak bir çok farklı yolla duygularını ifade etmektedir. Bu da farklı kanallardan gelen büyük miktarlarda veri oluşmasına ve elde edilen bu veriden "Duygu" kavramının ticari, kişisel, güvenlik vb. amaçlarla Makine Öğrenmesi alanında da incelenmesine yol açmıştır.

Müşteri veya çalışanlarla yapılan yazılı, sesli, videolu görüşmelerin veya sosyal medya üzerinden alınan yorum ve geri bildirimlerin analiz edilmesiyle, sunulan hizmet ve ürünlerin veya işleyişin eksikliklerini gidermek suretiyle verim artırmak, sanal eğitim sistemlerinde öğrencilerin duygularını tanıma yoluyla öğrenim süreçlerini iyileştirmek, sağlık sektöründe hasta takibi veya güvenlik sektöründe olası problemlerin önüne geçmek duygu tespitinin kullanılabileceği başlıca alanlardır. Bu alanda yüksek başarımlı otomatik sistemler üzerine son yıllarda bir çok çalışma yapılmıştır [2, 3, 4]. Çevrimiçi platformlarda reklamcılık ve hedef kitle pazarlaması için de duygu analizinden faydalanı labilir. Kullanıcıların ağdaki diğer insanlara karşı duygularına, ilgi alanlarına, fikirlerine, davranışlarına ve diğer kişilik temelli özelliklere dayalı olarak sosyal ağ platformlarında çeşitli öneri sistemleri uygulanabilmektedir [5]. İntihara meyilli veya duygusal sorunlar yaşayan kişileri tespit etmek gibi kritik durumları erkenden tanımak ve dolayısıyla ciddi kayıpları önlemek için kullanıcıların sosyal ağ gönderilerinden duygularını tespit etme üzerine çeşitli yapay zeka sistemleri geliştirilmektedir [6].

1.1 Tezin Katkıları

Yüz ifadeleri ve vücut hareketleri duyguları ifade etmenin başlıca yolları olduğundan ve son zamanlarda fotoğraf ve videolu iletişim arttığından görüntüden duygu tespiti popüler bir araştırma alanı olmuştur [7]. Ancak görüntü işleme ve derin öğrenme yöntemleri ile gerçekleştirilen bu sistemler hala gelişime açıktır. Bu tez çalışmasında kapsamlı bir veri ön işleme aşaması ile güncel derin öğrenme yöntemleri kullanılarak benzer çalışmalara göre daha yüksek başarımlı bir duygu tahmin sistemi geliştirilmiştir. Girdi olarak kullanılan video verisinden öncelikle öznitelik çıkarımı yapıldı. Bunun için OpenPose [8] isimli araçla görüntülerdeki yüz ve vücutlardan anahtar noktalar elde edildi ve bunlar özniteliklere dönüştürüldü. Tahmin mekanizması için ilk katmanı Evrişimli Sinir Ağı (CNN) ikinci katmanı ise ilk olarak Uzun Kısa Süreli Bellek (LSTM) ve ikinci olarak Çok-İmleçli İlgi Tabanlı Dönüştürücü (Transformer) olmak üzere iki farklı sinir ağı yapısı oluştutularak başarımları karşılaştırılmıştır. Çalışmanın devamında başarımı artırmak üzere görüntüden elde edilen özniteliklere ek olarak Gauss Karışım Merkezleri (GMC) elde edilmiş sisteme dahil edilmiş, ayrıca daha sonra ayrıntılarıyla bahsedeceğimiz "Video Çoklama" ve "Anahtar Çerçeve Seçimi" yöntemleri uygulanmıştır. Bu yöntemler sonucu işlediğimiz veri ile CNN-LSTM ve CNN-Transofmer yapıları ayrı ayrı tekrar beslenmiş ve başarımları karşılaştırılmıştır.
Bu tezin en büyük katkısı, CNN ve Dönüştürücü yapılarının bir araya getirildiği çok katmanlı bir DNN yapısı ve beraberinde bu yapının başarımını artıran OpenPose, Video Çoklama, Anahtar Çerçeve Seçimi ve GMC adımlarından oluşan ön işleme aşamasıyla yüksek başarımlı duygu tahmin sistemi sunmasıdır. İki yaygın kullanılan veri kümesinde yaptığımız deneyler sonucu Dönüştürücü tabanlı modelin çoğu versiyonu güncel DNN modellerine göre daha iyi performans gösterdiği gözlemlenmiştir.

1.2 Literatür Araştırması

Davranış ve duygu analizi üzerine yakın zamanda yapılan çalışmaların incelendiği [9] ve [10]' de görüldüğü üzere duygu analizi gelişen teknoloji ile akıllı sistemler, sosyal medya, güvenlik, sağlık, ticari ve benzeri bir çok farklı uygulama alanı olan ve farklı disiplin ve yöntemlerle üzerine sıkça çalışılan bir alandır. Gelişen teknoloji ve görüntülü

iletişimin artışıyla videodan yüksek doğruluk oranlarıyla duygu analizi yapabilmek büyük önem kazanmıştır. Nonis vd. yaptığı çalışmada [11] var olan farklı ifade tanıma yöntemlerinin avantaj ve kısıtlamalını incelemiştir.

Görüntülerden duygu tespiti için Doğrusal Ayrımcılık Analizi (LDA), Değiştirilmiş Temel Bileşen Analizi (PCA), Destek Vektör Makinası, k-en yakın komşuluk, çok katmanlı algılayıcı veya Rastgele Orman gibi geleneksel yöntemler de kullanılmıştır. Ancak son zamanlarda görüntü işleme alanında gösterdikleri başarıdan dolayı derin öğrenme modelleri bu problemde de sıkça uygulanmaya başlamıştır. Yapılan diğer bir çalışmada [12] duygu analizi alanında geliştirilen yeni ve orijinal yapay sinir ağları yaklaşımları incelenmiştir. Buna göre derin öğrenme ile görüntülerden duygu analizinin iki ana zorluğu: (i) aşırı uymayı önlemek için büyük veri setinin gerekliliği ve (ii) yaş, cinsiyet, kültür, köken gibi kişisel özelliklerin bireylerin duygularını ifade ediş biçimlerinde farklılıklara sebep olmasıdır.

Duygu tespiti ile ilgili en son çalışmalar çoğunlukla sinir ağlarına dayandığından ve bu tez çalışmasında da yapay sinir ağı tabanlı bir yöntem sunulduğundan bu alanda yapılan benzer çalışmalar derlenmiştir. Mungra vd. [13], CK+ ve beraberinde farklı veri kümeleri ile test ettiği yüz ifadelerini yedi temel duygu sınıfına göre sınıflandıran PRATIT adlı CNN tabanlı, %64-%76 arasında test başarımı elde eden bir duygu tanıma sistemi önermiştir. İlgili çalışmada ön işleme olarak görüntülere histogram eşitlemesi ve veri çoklama yöntemleri uygulamış ve performansa etkilerini incelemişlerdir.

[14]'da bir Çok-kanallı Evirişimsel Yapay Sinir Ağı (MCCNN) önerilmiş, FABO veri kümesi ile test edilen model %91.3 doğruluk ile en güncel modellerden daha yüksek performans göstermiştir. Yazarlar, etkilerini görmek sadece yüz, sadece vücut ve bu ikisinin bir arada kullanıldığı farklı deneyler yürütmüşlerdir.

Bu alandaki bazı araştırmalar daha çok öznitelik çıkarımı, parametre belirleme ve modellerin yapılarına odaklanmıştır. Örneğin, geleneksel Ölçek Değişmez Unsur Dönü şümü (scale-invariant feature transform, SIFT) ile sığ, CNN modeli ile derin özniteliklerin çıkarıldığı özgün bir CNN-SIFT yapısı ile duygu analizi yöntemi [15]'de sunulmuştur. Bu sığ ve derin öznitelikleri kullanan bir Destek Vektör Mekanizması, CK+ ve beraberin de farklı veri kümeleri ile tets edilmiş ve mevcut modellerin bir kısmından daha iyi sonuçlar elde edilmiştir.

Yapılan başka bir çalıışmada [16] Evrişimli Yapay Sinir Ağlarında çekirdek ve filtre boyutlarındaki değişimin duygu tanıma üzerinde etkileri incelenmiştir. Önerilen 2 farklı CNN yapısı kullandıkları veri kümesinde benzer performanslar sergilemiş, yaklaşık %65 doğruluk elde etmiştir.

Videolardan duygu tespiti için CNN-LSTM tabanlı bir model [17]'de önerilmiştir. Yazarlar, ardışık video kareleri arasında zamansal bir ilişki oluşturarak videodaki yüz ifadelerini çıkaran Yerel Gelişmiş Hareket Geçmişi Görüntüsü (LEMHI) adlı yeni bir yaklaşım önermiştir. Duygu sınıflandırması için veri kümesi ile Zamansal Segment LSTM (TS-LSTM) ve Görsel Geometri Grubu (VGG) ağları ayrı ayrı kullanılmış ve sonuçları, nihai duygu tahminlerini elde etmek amacıyla rastgele arama ağırlıklı toplama stratejisi kullanılarak birleştirilmiştir. Deneyler CK+ dahil olmak üzere veri setleri ile tekrarlanmıştır. Modellerin duygu sınıflandırma doğrulukları %51.2 ile %93.9 arasında değerler almış, en yüksek başarımlı model, mevcut modellerden en az %5 daha yüksek doğruluk elde ederek daha iyi performans göstermiştir.

Yine videolardan duygu tespiti için CNN-LSTM tabanlı başka bir model Abdullah vd. tarafından geliştirilmiştir [18]. Öğrenme aktarımı ile, kapsamlı bir veri kümesi kullanılarak eğitilen model farklı veri kümeleri ile tekrar eğitilmiştir. Öznitelik çıkarımı aşamasında videolardan yüzleri algılayarak ayıklamış, sadece yüzlerin görünür olduğu kareler ele alınmıştır. Model %61 doğruluk elde etmiş ancak kullanılan veri kümesi üzerinde görzel özniteliklerle yapılan başka bir çalışma olmadığı için kıyaslama yapılama mıştır.

[19]'da önerilen İç içe LSTM (STC-LSTM) ile Uzaysal-Zamansal Evrişimli öznitelikler yaklaşımı CK+ dahil farklı veri kümeleri ile test edilmiş yeni bir yaklaşımdır. 3 boyutlu bir CNN modeli kullanılarak yüz ifadelerinden Uzaysal-Zamansal Evrişimli öznitelikler üretilmiş, 3 boyutlu CNN katmanları üzerinde çalışan LSTM'ler ile bu özniteliklerden daha yalın özniteliker elde edilmiştirtir. Önerilen yapı ile farklı veri kümelerinde %84.53 ile %99.8 arası doğruluk oranları görülmüş ve modelin temel metodlardan daha iyi performans sergilediği gözlemlenmiştir. Önerilen yöntem ara katmanlardan çok seviyeli öznitelikleri kullanmasına rağmen, bazı veri kümeleri için, veri kümesi yeterince büyük olmadığından, özellikle 5'ten fazla evrişimli katman kullanıldığında iyi performans sergilenememiştir.

Duygu algılama araştırmalarında sadece yüz ifadeleri değil vücut hareketleri üzerinde yapılan çalışmalar bulunmaktadır. Videolardan altı temel duygunun tespiti [20]'de iskelet hareketlerine odaklanılmıştır. Bu videolarda bireylerin eklemlerin sırası, konumla rı ve yönelimleri analiz edilmiştir. 3 boyutlu eklem konumları ve oryantasyonları, normalize edilmiş ve videolardan en önemli bilgileri içeren anahtar kareler seçilmiştir. Deneyler için CNN, RNN ve RNN-LSTM modellleri geliştirilmiş, %54 ile %82 arasında doğruluk lar elde edilmiştir. Model performansları Rastgele ağaçlar, k-yakın komşuluk, destek vektör makinaları ve çok katmanlı perseptronlar gibi klasik temel modellerle karşılaştırıl mış ve hepsinden daha başarılı olduğu gözlemlenmiştir.

Laban Hareket Analizi (LMA) ile dans hareketlerinden yedi farklı duyguyu tespit etmek için [21]'de bir CNN-LSTM modeli kullanılmıştır. Öznitelikler, insan vücudunun uzuv yapısı, uzuvların uzamsal yönelimi ve yerçekimi, uzay, zaman ve akıcılığın kuvvet etkisi olmak üzere üç karakteristik göz önünde bulundurularak üzerinden çıkarılmış ve önerilen modelle %97'ye yakın doğruluk skoru elde edilmiştir bu da mevcut modellerde daha yüksek performanslı bir yöntem olduğunu ispatlamıştır.

Başka bir jest tabanlı duygu tanıma sistemi, Ly vd. tarafından önerilmiştir [22]. 3 boyutlu bir CNN ile LSTM modeli yapısı FABO veri seti ile test edilmiştir. Yazarlar ayrıca ön işleme olarak rastgele kare seçimi ve anahtar kare seçimi yöntemleri uygulamış videolardan rastgele 40 veya 16 anahtar kare seçmişlerdir. 16 anahtar kare seçilerek eğitilen 3D-CNN-LSTM %66.6 doğruluk elde etmiştir. Model klasik ve güncel yapay öğrenme modelleri ile karşılaştırılmış ve bir çoğundan daha iyi performans sergilediği gözlemlenmiştir.

1.3 Tez Taslağı

Tez kapsamında yapılan çalışmalar şu şekilde sıralanmıştır: Bölüm 2.1'de, yapay sinir ağları, bunların türleri ve duygu sınıflandırma alanındaki kullanımları anlatılmıştır. Bölüm 2.2'nda, Dönüştürücü modelinin ve ilgi mekanizmasının ayrıntılarına yer verilmiş tir. Bölüm 3'de, kullanılan veri kümelerinden, ön işleme aşamalarından, model mimarisi ve yapılan deneylerden bahsedilmiştir. Bölüm 4'da ise deney sonuçları analiz edilmiştir. Son olarak Bölüm 5'de, yapılan çalışmalar özetlenmiş ve gelecek çalışmalardan bahsedil miştir.

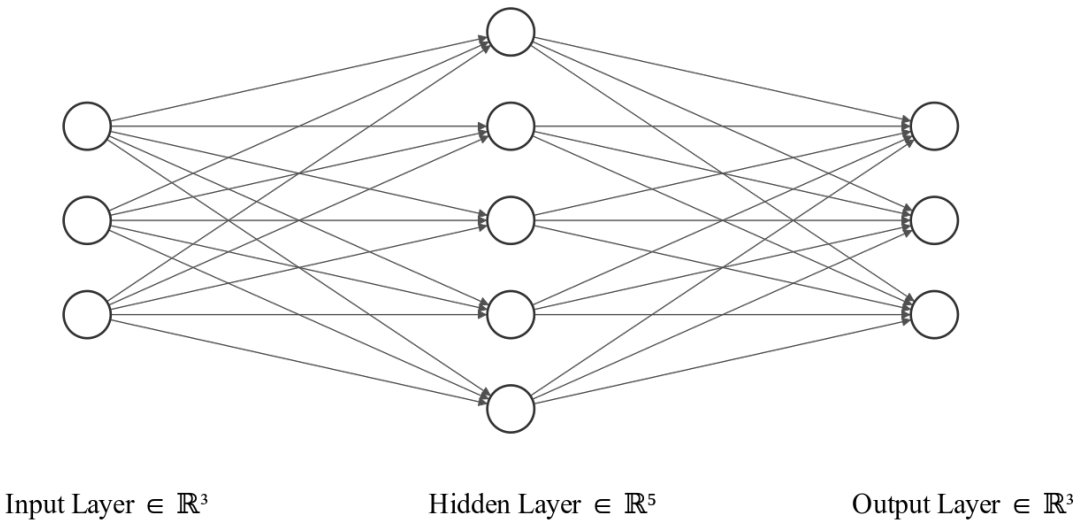
2. ÖN BİLGİ

2.1 Yapay Sinir Ağları ve Türleri

Bu bölümde, tez çalışmasında kullanılan Yapay Sinir Ağlarının (YSA) yapısı, çalışma mekanizmasına ait bilgiler verilmiştir.

2.1.1 Yapay sinir ağları

Yapay Sinir Ağları, temeli sinir ağlarının çalışma şeklini matematiksel olarak ifade etme [23] fikrine dayanan, insanların bilgileri öğrenme, işleme ve üretme yapısından esinlenilerek geliştirilmiş yapay öğrenme modelleridir. Aldığı girdiden bir çıktı üreten yapay sinir hücrelerinin bir araya gelmesi ile bir katman ve katmanların ard arda eklenmesi ile de yapay sinir ağı oluşturulmuş olunur. Şekil 2.1'da örnek bir Tam Bağlantılı Yapay Sinir Ağı gösterilmiştir. Görüldüğü üzere aynı katmandaki nörünların birbirleri ile bağlantısı olmayıp bir önceki katmandan girdi alıp bir sonraki katman iletmektedirler. Bu şekilde kurulan, tüm nöronlar arası bağların bulunduğu ağlara tam bağlantılı denir.



Şekil 2.1: Örnek bir tam bağlı yapay sinir ağı

Her nöronda, endisine bir önceki katmandan iletilen girdileri içeren girdi vektörünün,

katmanın ağırlık vektörü ile çarpımından elde edilen sonucun nörunun içindeki aktivas yon fonksiyonundan geçirilmesi ile çıktı üretilir ve bir sonraki katmana iletilir. YSA'larda Identity, Sigmoid, Tanh, ReLU vb. gibi farklı aktivasyon fonkısiyonları kullanılabilir. Bu tez çalışmasında ReLu kullanılmıştır. Açılımı Doğrultulmuş Doğrusal Birim (Rectified Linear Unit) olan ReLu, girdi negatif ise 0, pozitif ise kendi değerini alan bir fonksiyondur. Hesaplama yükünü azalttığı için derin yapay sinir ağlarında sıkça tercih edilir.

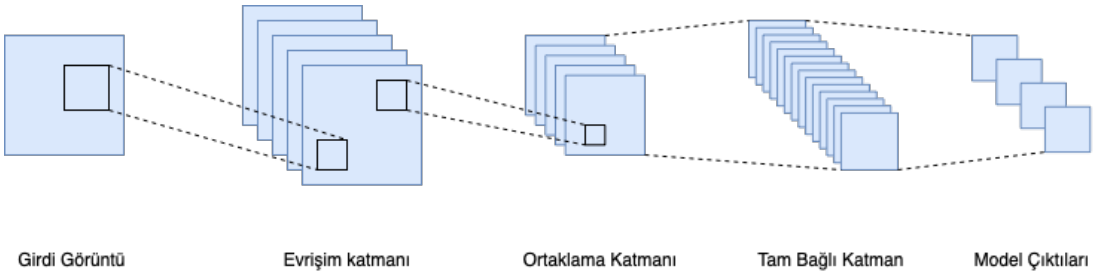
Yapay Sinir Ağlarında öğrenme işlemi katman ağırlıkları güncellenerek yapılır. Modelin performansını artırmak için belirlene bir hata fonksiyonu ile YSA'nın hatası hesaplanır ve model ağırlıkları bu hatayı azaltacak şekilde güncellenir. Bu işlem performans tatmin edici bir seviyeye gelene kadar, hatadaki iyileşme durup kendini tekrar etmeya başayana kadar veya başlankıçta belirlenen miktar kere tekrarlanır. Probleme ve modellenen sinir ağına göre bir çok farklı hata fonksiyonu kullanılabilir. Tez çalışmasında çok sınıflı bir sınıflandırma işlemi yapıldığı için hata fonksiyonu olarak Kategorik Çapraz Entropi kullanılmıştır.

2.1.2 Evrişimli sinir ağları (CNN)

Evrişimli Sinir Ağları (Convolutional Neural Networks - CNN), çoğunlukla görüntü ve video işleme problemlerinde kullanılan bir YSA türüdür. İlk olarak Fukushima tarafından 1980 yılında önerilmiştir [24]. Çoğunlukla evrişim katmanı, ortaklama (pooling) katmanı ve tam bağlı katmanlardan oluşur. Elde edilen çıktıya öznitelik haritası adı verilir. Örnek bir CNN yapısı Şekil 2.2'de gösterilmiştir. Evrişim katmanında belirlenen filtre boyutuna göre girdi taranarak evrişimsel öznitelikler üretir. Ardından ReLu ile doğrusallık giderilir. Ortaklama katmanında ise evrişim katmanında yapılan uzamsal değişmezliğin alt örneklemi alınır. Maksimum ortaklama ve ortalama ortaklama şeklinde çeşitleri bulunur. Tam bağlı katmanda ize örneklenmiş öznitelik haritasıı düzleştirilir ve model çıktıları üretilir.

2.1.3 Uzun kısa vadeli bellek (LSTM)

Uzun Kısa Vadeli Bellekler (Long-Short Term Memory - LSTM), 1997 yılında Hochreiter tarafından geliştirilen bir Özyineli Sinir Ağı türüdür [25]. Özyineli Sinir Ağları'nın



Şekil 2.2: Evrişimli Yapay Sinir Ağı Örneği

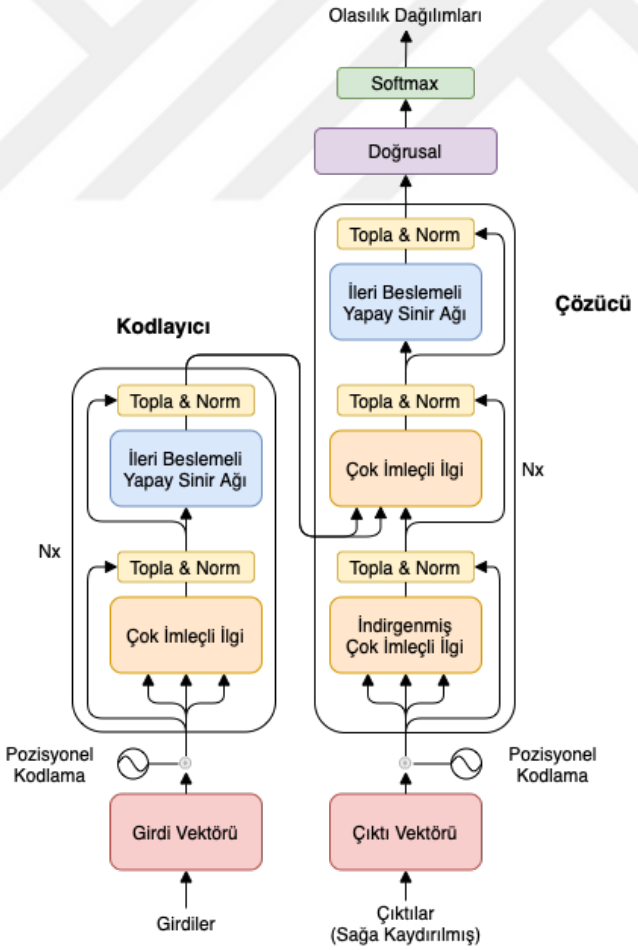
(Recurrent neural network - RNN) yapısı ileri beslemeli ağlardan biraz farklıdır. Şekil 2.1 deki gibi ileri beslemeli ağlarda girdi katman katman ilerleyerek bir çıktı oluşur, her katmanda aşamada girilen bilgi kullanılır. Özyineli Ağlarda ise nöronlar geri besleme döngüsüne sahiptir. Bu sayede katmanların çıktıları sadece ileri değil geri besleme amacıyla da kullanılır. Bu mekanizma ile zaman serisi, cümle, video gibi sekans halinde ve önce ve sonrasıyla ilişkili girdi tiplerinde bir önceki girdi de hatırlanmış ve etkisi de dikkate alınmış olur. RNN'lerde sekans ilerledikçe eski girdiler unutulur, önemli olan bilgiler kaybedilmiş olabilir. Bunun üstesinden gelmek için LSTM'ler kullanılır. LSTM'lerin nöronlarında onlara hafıza özelliği katan bir takım işlemler yapılır ve bu sayede önemli bilgiler hafızada tutulurken önemsizler unutulur. Hafıza mekanizması temel RNN'lere göre avantaj sağlasada sekans uzadıkça eski bilgiler LSTM'lerde de önemini yitirmeye devam eder bu sebeple çok uzun girdilerde performansları düşebilir.

2.2 Dönüştürücü ve İlgi Mekanizması

2.1 Bölümünde sekans şeklinde veriler için LSTM mekanizması kullanmanın avantajlı olduğu ancak uzun sekanslarda bilginin unutulmasının bu modelin bir kısıtı olduğu anlatıldı. İlgi Mekanizmalı Dönüştürücü modelleri bu sorunu ortadan kaldırarak özellikle doğal dil işleme problemlerinde yüksek performans gösteren yeni bir yaklaşımdır [26]. Bu bölümde Dönüştürücü, İlgi Mekanizması ve Pozisyonel Kodlamaya dair bilgiler sunuldu.

Dönüştürücüler (Transformer), [26]'de sunulduğu üzere, İlgi Mekanizması kullanarak özellikle otomatik çeviri gibi doğal dil işleme uygulamalarında, derin öğrenme modelleri nin performansını yüksek oranda iyileştiren bir mimaridir. Yapısında içerisinde İlgi katmanı bulunan kodlayıcı ve çözücü yığıtlarından oluşur. Burada İlgi mekanizmaları o

an gelen girdi işlenirken, bu girdi ile alakalı diğer tüm geçmiş girdilerin de önemlerine göre hesaba takılmasını sağlar. Bu sayede sıralı girdilerde, sekans uzasa bile Dönüştürücü girdiler arasında varsa ilişki kurabilir ve unutmanın önüne geçilmiş olur. Pozisyonel kodlama ve İlgi mekanizması ile bir sekans işlenirken her bir adımda sekansdaki diğer konumlara bakılarak, ilgili pozisyon için daha iyi bir kodlama elde etmeyi sağlar. Tek imleçli İlgi mekanizması ile bu ilişkilendirmeyi sekanstaki tek pozisyonla yapabiliyorken, çok imleçli İlgi mekanizması sekansın farklı pozisyonlarına da odaklana bilme imkan sağlar. Şekil 2.3'da Çok İmleçli İlgi Mekanizmalı Dönüştürücü şeması verilmiştir. Bu yapıyı kullanabilmek için ilk olarak sıralı girdiler modelin işleyebileceği şekilde temsil edilir. Örneğin bir cümle işleniyorsa bu cümlenin kelimeleri kelime gömmelerine dönüştürülür. Ardından Dönüştürü yapısına kelime gömme vektörleri girdi olarak verilir.



Şekil 2.3: Çok İmleçli İlgi Mekanizmalı Dönüştürücü

Şekil 2.3'da da görüldüğü üzere kelime gömmeleri ile birlikte pozisyonel kodlama ile bu gömmelerin sekansdaki pozisyonlarına ait bilgi de dönüştürücü modeline verilir. Bu sayede yapının girdideki sıra ve uzaklıkları anlaması sağlanmış olur. Kodlayıcı içerisinde bu girdiler teker teker işlenmeye başlar. Kodlayıcı yığıtında İlgi mekanizması ve ileri beslemeli ağdan geçen girdiler çözücü yığıtına iletilir. Burada görüldüğü üzere dönüştürücü modelinde farklı İlgi mekanizmaları bulunmaktadır. Girdi yığıtında bulunan Çok İmleçli İlgi, Öz İlgi yapısıdır; sekanstaki farklı pozisyonlar arasındaki ilişkiyi hesaplamaya yarar. Çözücü yığıtındaki İndirgenmiş Çok İmleçli İlgi katmanı çözme aşamasında o an işlenen kısmın sadece kendinden önce işlenmiş pozisyonlara bakmasını sağlar, sonraki pozisyonlarla ilişki kurulmasını önler. Son olarak Çözücüdeki Çok İmleçli İlgi ise kodlayıcı ve çözücü girdilerini ilişkilendirerek, çözme aşamasında o an işlenen kısmın kodlayıcı girdisinde odaklanılması gereken pozisyonun bulunmasını sağlar. İlgi katmanı çıktıları ileri beslemeli katmana iletilir ve çıktı üretilir. Sekansın sonuna gelene kadar çıktılar bir sağa kaydırılarak (ilerletilerek), çıktı katmanına iletilir. Sekans tamamlandığında olasılık dağılımları elde edilmiş olunur.

Bu mimari otomatik çeviri, metin özetleme, metin üretme, varlık ismi tanıma gibi doğal dil işleme uygulamalında model başarımlarını oldukça artırmış, Dönüştürücü tabanlı bir çok metin modeli geliştirilip, kullanıma hazır halde sunulmuştur [27, 28, 29].

3. METODOLOJİ

3.1 Veri Kümesi

Bahsedildiği üzere tez çalışması jest ve mimiklerden yapay sinir ağları ile duygu sınıflandırması yapmak üzerinedir. Bunun için yaygın olarak kullanılan ve metodolojiye uygun 2 yaygın veri kümesini seçilmiştir - (i) FABO [30], videolardan duygu analizi için kullanıma hazır veri kümesi, ve (ii) genişletilmiş Cohn-Kanade veri kümesi (CK+) [31].

3.1.1 FABO

FABO veri kümesi (The Bi-modal Face and Body Gesture Database for Automatic Analysis of Human Nonverbal Affective Behavior), Hatice Güneş ve Massimo Piccardi tarafından 2005 yılında yayınlanmıştır. Bu veri kümesi poz verirken, lab ortamında, sadece ifadelere odaklanılan, duygulara odaklanılan ve açık kayıtlar olmak üzere 5 farklı kurulumda alınmış yüz ve vücut video kayıtlarını içerir. Farklı etnik köken, yaş ve cinsiyetten 23 kişinin renkli kayıtları bu şekilde alınmıştır. Veri kümesinde nötr, kararsızlık, sinir, şaşkınlık (puzzlement), korku, endişe, mutluluk, şaşkınlık (suprised), iğrenme, sıkılmışlık ve üzgün olmak üzere 11 farklı duyguya ait 1900 civarı video bulunmakta olup bunlar sadece yüz, sadece vücut ve hem yüz hem vücut olarak ayrışmaktadır. Sadece yüz veya sadece vücut için yahut hem yüz hem vücut için tüm bir video serisi ya da bu serilerden alınmış tekil kareler olarak kullanılabilir. Görüntülerin ve duyguların yakınlığından dolayı şaşkınlık (puzzlement) ve kararsızlık sınıfları birleştirilerek kararsızlık olarak tek bir sınıf haline getirildi. Benzer şekilde şaşkınlık (suprised) sınıfı da pozitif ve negatif olmak üzere 2 ayrı alt sınıfa ayrılmış durumdaydı ve bu iki sınıf da şaşkınlık olarak tek sınıf olmak üzere birleştirildi. FABO kümesinde hem yüz hem de vücut içeren 16 farklı kişinin toplamda 361 videosu nötr hariç 9 duyguyu içerecek şekilde etiketlidir, nötr sınıfına ait hem yüz hem de vicut içeren etiketli video bulunmamaktadır. Bu tez çalışmasının kapsamı jest ve mimikler olduğundan hem yüz hem de vücut görüntülerini kapsayan bu 361 video kullanıldı. Başka bir deyişle, tez çalışmasında yukarda bahsedilen, sinir, endişe, sıkılmışlık, iğrenme, korku, mutluluk, üzüntü, şaşkınlık, kararsızlık olmak üzere 9 duygu içeren 361 videodan elde edilen yüz

ve vücut 'tanımlayıcılar' kullanılmıştır. Bu 361 videodan 75 adedi sinir, 27 adedi endişe, 40 adedi sıkılmışlık, 29 adedi iğrenme, 26 adedi korku, 31 adedi mutluluk, 20 adedi üzüntü, 29 adedi şaşkınlık ve 83 adedi de kararsızlık sınıflarına aittir. Sınıflara göre video dağılımları Çizelge 3.1'de verilmiştir. Resim 3.1'da bu veri kümesine ait şaşkınlık sınıfından bir örnek verilmiştir.

Çizelge 3.1: FABO veri kümesinden kullanılan videoların sınıflara göre dağılımları

sinir	endişe	sıkılmışlık	iğrenme	korku	mutluluk	üzüntü	şaşkınlık	kararsızlık
75	27	40	29	26	32	20	29	83

Resim 3.1: FABO veri kümesinden şaşkınlık sınıfına ait bir örnek. [30]

3.1.2 CK+

CK+ veri kümesi yazarların daha önce 2000 yılında yayınladığı CK veri kümesinin 2010'da yayınlanmış genişletilmiş bir versiyonudur. Sinir, korku, küçümseme, iğrenme, mutluluk, üzüntü ve şaşkınlık olmak üzere 7 duyguyu barındıran ve sadece yüz ifadelerini içeren, 45 adet kızgın, 18 adet küçümseme 59 adet iğrenme, 25 adet korku, 69 adet mutluluk, 28 adet üzüntü ve 83 adet şaşkınlık olmak üzere toplam 593 renkli ve siyah beyaz kayıttan oluşmaktadır. Görüntülerin sınıflara göre dağılımı Çizelge 3.2'de verilmiştir. Bu kayıtlar görüntü serileri şeklinde bireysel kareler halinde sunulmuştur. Hem poz vermiş hem de pozsuz öznelerin yüz ifadeleri nötrden ilgili duyguya geçecek şekilde görüntü serileri bulunmaktadır. Sunulan toplam 593 kaydın 327 adeti etiketlidir bu sebeple tez çalışmaısnda bunlar kullanıldı. FABO'nun aksine CK+ veri kümesindeki görüntülerde vücut bulunmamaktadır bu sebeple sadece yüz ifadelerinden elde edilen

tanımlayıcılar kullanıldı. Resim 3.2'da bu veri kümesine ait üzüntü sınıfından bir örnek verilmiştir.

Çizelge 3.2: CK+ veri kümesinden kullanılan videoların sınıflara göre dağılımları

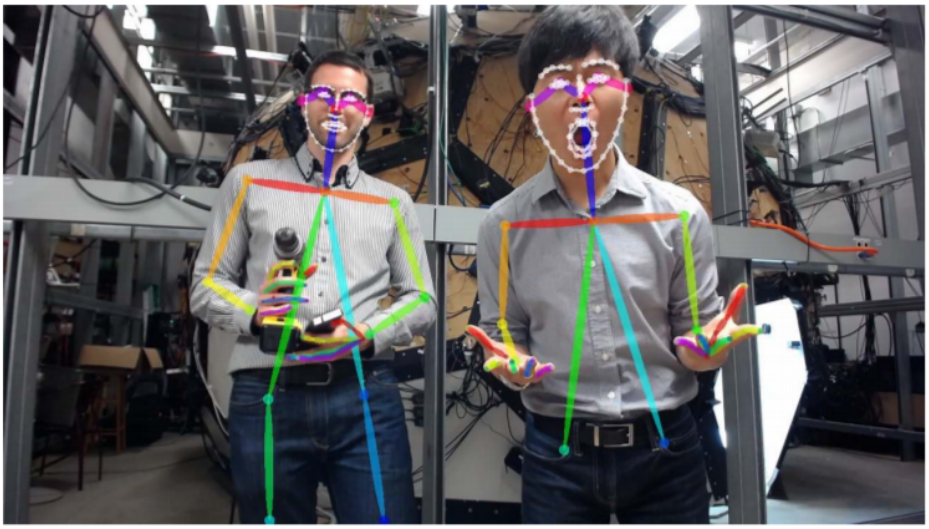
sinir	küçümseme	iğrenme	korku	mutluluk	üzüntü	şaşkınlık
45	18	59	25	69	28	83

Resim 3.2: CK+ veri kümesinden üzüntü sınıfına ait bir örnek. [31]

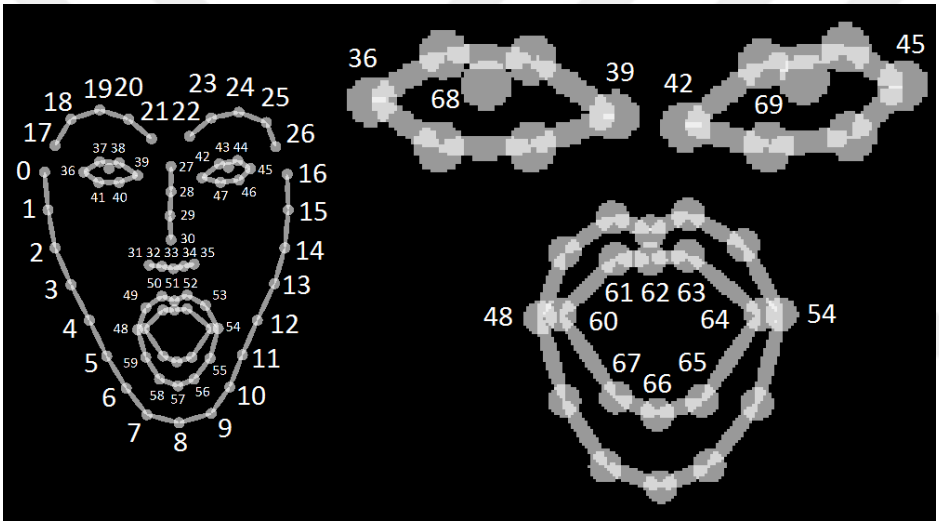
3.2 Ön İşleme ve Öznitelik Çıkarımı

3.2.1 OpenPose

Duygu sınıflandırma modeline beslemeden önce veri kümeleri bir ön işleme aşamasından geçirildi. Burada video ve görüntü dizilerinden öznitelik çıkarma amacıyla OpenPose [8, 32, 33, 34] adlı bir araç kullanıldı. OpenPose, insanların el, yüz, ayak ve vücutlarından 135 farklı anahtar noktayı gerçek zamanlı olarak tespit edebilen bir sistemdir. Farklı işletim sistemlerinde ve donanımlarda, fotoğraf, video, web kamera ve IP camera akışları gibi farklı girdilerle çalışabilir. 70 adet yüz, 21 adet her bir el, 6'sı ayak olmak üzere 15/18/25 vücut ve ayaklar, toplamda 3 farklı blok halinde 135 anahtar nokta tespit edebilir. Resim 3.3'da örnek bir OpenPose çıktısı verilmiştir. OpenPose'da kullanılan yüz vücut ve el anahtar noktaları Şekiller 3.1, 3.2, 3.3'de ayrı ayrı gösterilmiştir.
Tez çalışmasında OpenPose ile görüntülerden anahtar noktalar elde edildikten sonra "Yaygın bilgisayarla görme teknikleriyle insan yüz ifadesini algılama" (Detecting human facial expression by common computer vision techniques [36]) adlı çalışma örnek alınarak anahtar noktalar duygu sınıflandırma modellerinde öznitelik olarak kullanılmak



Resim 3.3: Yüz ve vücut için örnek bir OpenPose çıktsı. [8].

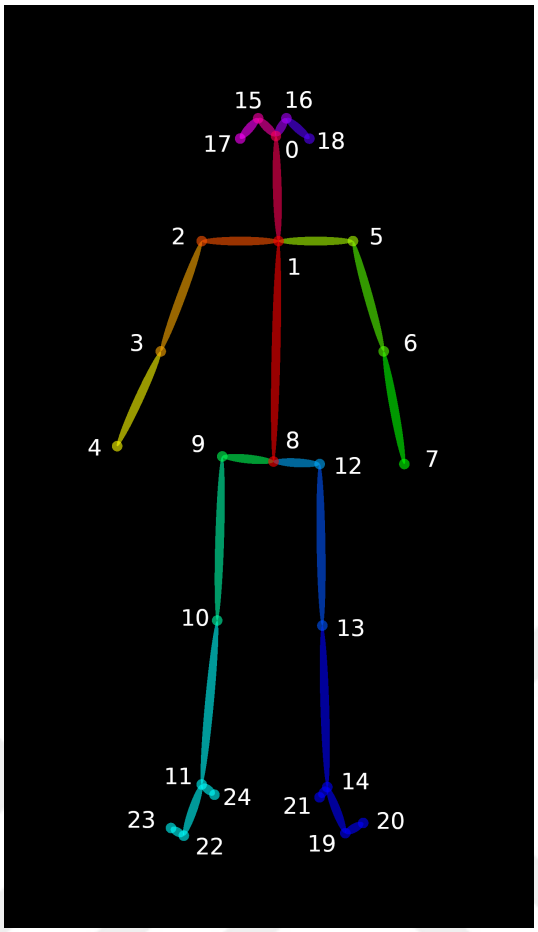


Şekil 3.1: OpenPose Yüz anahtar noktaları. [35]

üzere yüz ve vücut tanımlayıcılarına dönüştürüldü. Oluşturulan tanımlayıcıların listesi Çizelge 3.3'da verilmiştir. 3 farklı tanımlayıcı kümesi oluşturuldu;

- (i) sadece vücuttan elde edilen anahtar noktalarla temel tanımlayıcı
- (ii) temel ve el tanımlayıcılar
- (ii) temel, el ve yüz tanımlayıcılar

OpenPose aracı kullanılmadan önce görüntüler aşağıda maddelenmiş bir takım işlemden geçirildi;

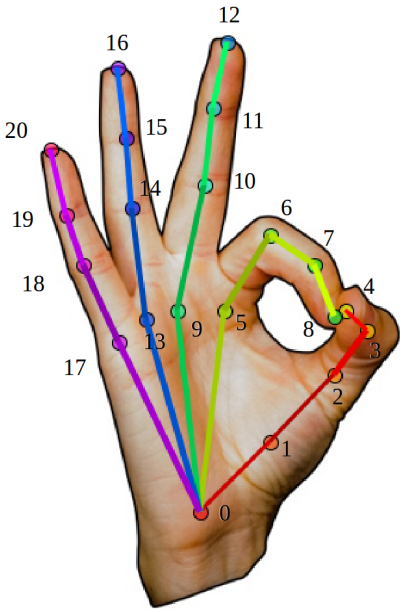


Şekil 3.2: OpenPose Vücut anahtar noktaları. [35]

- FABO veri kümesinde her bir video nötr durumda başlayıp 3 saniye içinde ilgili duygunun zirvesine ulaşmaktaydı. Bu saniyedeki kare sayısı 15 olan bu videolarda ilk 45 kare anlamına gelmektedir. Bu sebeple, çalışmada işlenmemiş videoların ilk 45 kareleri dikkate alındı.

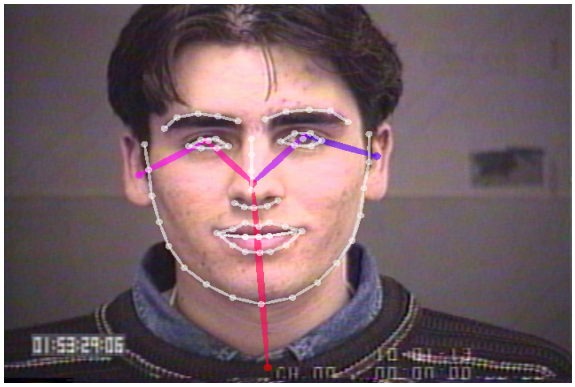
- Bazı videolarda öznelerin kameraya uzaklığı sebebiyle orantısal bir sorun olaşmak taydı. Bu sorunu aşmak için öznelerin iki kulağı arasındaki mesafeyi ölçekleme faktörü olarak kabul ederek görüntüdeki diğer mesafeler bu sabite göre ölçeklendi rildi.

Daha sonra anahtar noktaları elde etmek üzere ilk 45 kareye OpenPose uygulandı. Ardından elde edilen anahtar noktalardan tanımlayıcılar elde edilmek üzere, üst ve alt dudak arası mesafe, göz ve kaş arası mesafe, baş boyun ve omuz arasındaki açı vb. gibi açılar ve mesafeler hesaplandı.



Şekil 3.3: OpenPose el anahtar nokataları. [35]

Benzer yaklaşım CK+ veri kümesi için de sadece yüz için olmak üzere izlendi. CK+ verisetine OpenPose siyah beyaz görüntülerde iyi performans göstermediği için ayrıca renklendirme işlemi uygulandı, bunun için açık kaynak bir araçtan faydalanıldı [37]. Resim 3.4'da CK+ veri kümesine ait renklendirilmiş ve OpenPose uygulanmış örnek bir görüntü verilmiştir.



Resim 3.4: CK+ veri kümesinden renklendirilmiş ve OpenPose uygulanmış bir kare

Öznitelik olarak kullanmak üzere açı ve ikili mesafelerden Vücut (73), Sağ ve Sol Eller (211 x 2), Yüz (126) olmak üzere toplam 621 adet öznitelik elde edildi. Bu öznitelikler üzerinden bir öznitelik seçimi yapılarak gürültü ekleyenler ayıklandı ve toplamda on

altı öznitelik seçildi. Yedi yüz ve dokuz vücut olmak üzere kullanılan tüm öznitelikler Çizelge 3.3'da listelenmiştir.

Çizelge 3.3: Duygu sınıflandırma için kullanılan poz tanımlayıcılar

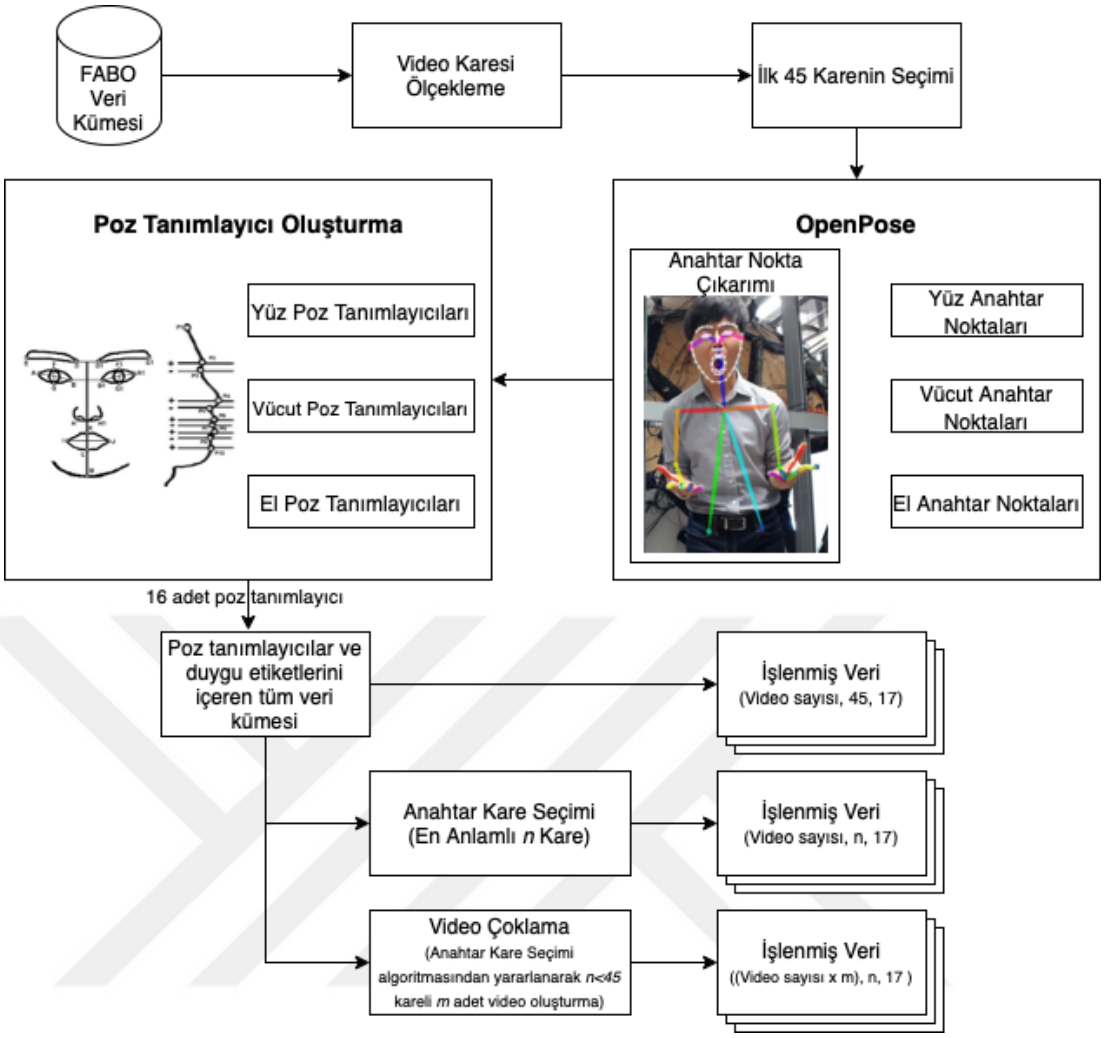
Bölge	Ölçü	Poz Tanımlayıcı
Yüz	Mesafe	Gözler Kaş ve üst göz kapağı Kaş ve alt göz kapağı Burun ve üst dudak Üst ve alt dudak Alt dudak ve çene Ağız genişliği
Vücut & El	Açı	Burun, boyun, sağ omuz Boyun, sağ omuz, sağ dirsek Sağ omuz, sağ dirsek, sağ bilek Sağ dirsek, sağ bilek, sağ avuç Burun, boyun, sol omuz Boyun, sol omuz, sol dirsek Sol omuz, sol dirsek, sol bilek Sol dirsek, sol bilek, sol avuç Sağ omuz, boyun, sol omuz

3.2.2 Video çoklama ve anahtar kare seçimi

Çizelge 3.4 FABO veri kümesi için izlenen ön işleme aşamalarını göstermektedir. Veriyi ölçeklendirme ve openpose aşamalarından sonra hem veri miktarını artırmak ve sınıf dağılımındaki dengesizliği aşmak hem de veriyi daha anlamlı hale getirmek için "Anahtar Kare Seçimi" ve "Video Çoklama" yöntemleri uygulandı. Bazı kareler hareketin yavaşlığı veya duraksama dolayısıyla birbirine çok yakın olduğundan dolayı görüntü disizindeki tüm 45 kare aynı miktarda önem arzetmemektedir. Bu sebeple 45 kare arasından en çok bilgi içeren anahtar kareleri seçmek üzere bir anahtar kare seçimi yöntemi geliştirildi. Her bir video için kareler arasındaki öklid uzaklığı hesaplandı ve bu mesafeler [0-1] arasına normalize edilerek kareler mesafelerine göre puanladı. Bu sayede en yüksek mesafeli yani en farklı kare daha geniş aralığı kapsayarak en yüksek önceliğe sahip olmuş oldu bu da rastgele seçilme ihtimalini artırmış oldu. Daha sonra yine [0-1] arasında rastgele bir sayı üreterek bu sayıya büyük veya eşit en yakın puana sahip kare puanlanmış kareler arasından seçildi ve bu işlem n kere tekrarlandı. Çok

fazla veri kaybına sebep olmamak açısından farklı değerler de denenerek işlem 30 kere tekrarlandı ve 45 karelik videolardan 30 karelik kısa ve anlamlı videolar elde edildi. Ardından veri miktarını artırmak ve sınıflar arası dengesizliği aşmak üzere videolar çoklandı. Bunun için de her bir videodan bahsedilen "Anahtar Kare Seçimi" yöntemiyle kısmen rastgele seçilmiş *n* adet kareden oluşan *m* farklı kare kombinasyonuyla yeni kısa videolar elde edildi. Farklı kare kombinasyonları seçildiği için videolar tekrarlanmamış oldu. Bu aşamada n ve m sayıları için farklı değerler denendi ve sırasıyla 5 ve 10'da karar kılındı. Anahtar Kare Seçimi veya video çoklamaya tabi tutulmamış hali ile FABO veri kümesi (361, 45, 17) olmak üzere her biri 45 kare olmak ve her bir kare için 17 poz tanımlayıcı olmak üzere 361 videodan oluşan 3 boyutlu bir veriye dönüştürüldü. Video çoklama ile 361 videonun her birinden 30 kareli 5 yeni video oluşturulması sonucu (1085, 30, 17) boyutlarında bir veri kümesi elde edildi. Ardından Anahtar Kare Seçimi sonrası bu veri (1085, 10, 17) olacak şekilde 10 kareye düşmüş oldu.

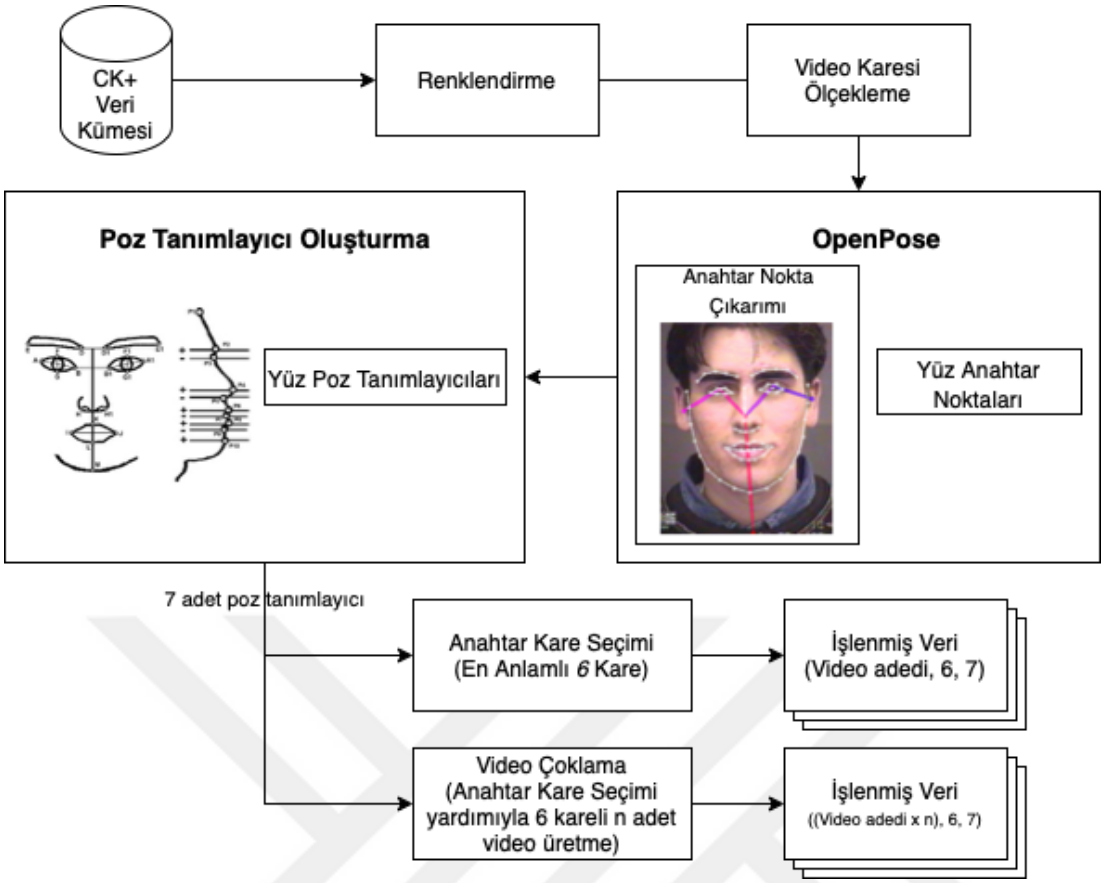
Anahtar kare seçimi ve video çoklama adımlarında CK+ veri kümesi için verinin yapısından dolayı kısmen farklı bir yaklaşım izlemek gerekti. CK+ veri kümesinde örneklerin uzunlukları FABO'nun aksine farklılık göstermektedir. En kısa görüntü dizisi 6 kare olup en uzun da 50 karedir. Burada kısa videoları kareleri tekrar etme veya boş karelerle doldurma şeklinde yöntemler izlenerek kare sayılarını eşitleme yolu denenmiş ancak bu yaklaşımlar performansı düşürmüştür. Bu sebeple CK+ veri kümesi poz tanımlayıcılar seçildikten sonra anahtar kare seçimi işlemine tabi tutulmuş, her bir görüntü dizisinden 6 anahtar kare seçilmiştir. Tüm adımlardan sonra CK+ veri kümesinden 327 video için video çoklama öncesinde (327, 6, 7) boyutlarında 3 boyutlu bir veri elde edilmiştir, ilk boyut her bir video, 2. boyut videonun içerdiği kareler, 3. boyut da öznitelik olarak kullanacağımız poz tanımlayıcılardır. Daha sonra video çoklama aşamasında ise her bir örnekten 6 kare uzunluğunda 5 farklı yeni görüntü dizisi oluşturulmuştur. Burada 6, 7, 8, 9, 10 kare uzunluğundaki görüntüler tekrara düşeceği için bu uzunluktaki girdilerden 6 kare uzunluğundaysa olduğu gibi kullanılmış diğer uzunluktakilerden ise sırasıyla sadece 2, 3, 4, 5 adet yeni görüntü oluşturularak tekrarın önüne geçilmiştir. CK+ veri kümesi için izlenen ön işleme adımları Şekil 3.5'da gösterilmiştir.



Şekil 3.4: FABO veri kümesi için ön işleme adımları. Poz Tanımlayıcı yüz görseli için [36]

3.2.3 Gauss karışım merkezleri

Performansa etkisini gözlemlemek amacıyla veriye Gauss Karışım Merkezleri öznitelik olarak eklendi. Gauss Karışım Modeli kullanılarak veri seti 6 merkez, 9 merkez ve 12 merkez olmak üzere üç farklı şekilde gruplandı. Burada merkez sayıları küme miktarı ve dugu sınıfı mikarı ilişkisini gözlemlemek için bu şekilde seçildi. Testler 6 ve 9 duygu ile yürütüldüğü için Gauss karışım modeli ile verinin 6 gauss karışım merkezine ve 9 gauss karışım merkezine ayrılmış iki versiyonu oluşturuldu. Ardından sınıf sayısından daha yüksek küme miktarının etkisini analiz etmek için 12 merkezli gauss karışım modeli de çalıştırılarak üçüncü versiyon da oluşturuldu. Ardından her bir videoya gauss karışım merkezi öznitelik olarak eklendi.



Şekil 3.5: CK+ veri kümesi için ön işleme adımları. Poz Tanımlayıcı yüz görseli. [36]

3.3 Duygu Sınıflandırma

Duygu sınıflandırma sisteminde iki katmanlı bir yaklaşım izlendi. İlk katmanda bununan CNN bloğuna ön işleme adımlarının sonucunda elde edilen özniteliklerin girdi olarak verildi ve bu katmanın çıktıları ile ikinci katman beslendi. İkinci katman için karşılaştırma amacıyla iki farklı yapay sinir ağı modeli kullanıldı. İlk olarak LSTM ağı ikinci olarak da Çok-İmleçli İlgi Mekanizmalı Dönüştürücü denendi. Tez çalışmasında kurulan sınıflandırıcı yapısının iskeleti Şekil 3.6'de gösterilmiştir. Kapsamlı bir analiz yapabilmek için hem kullanılan verinin, hem modellerin, hem katmanlı yapının farklı kombinasyonları ile farklı yapılar kuruldu ve sonuçlar elde edildi.

Hem FABO hem de CK+ veri kümesinin farklı versiyonları ön işleme bölümünde bahsedildiği şekilde edildi;

- Herhangi bir işlemden geçirilmemiş, poz tanımlayıcılardan oluşan veri kümesi
- Video Çoklama ile genişletilmiş veri kümesi

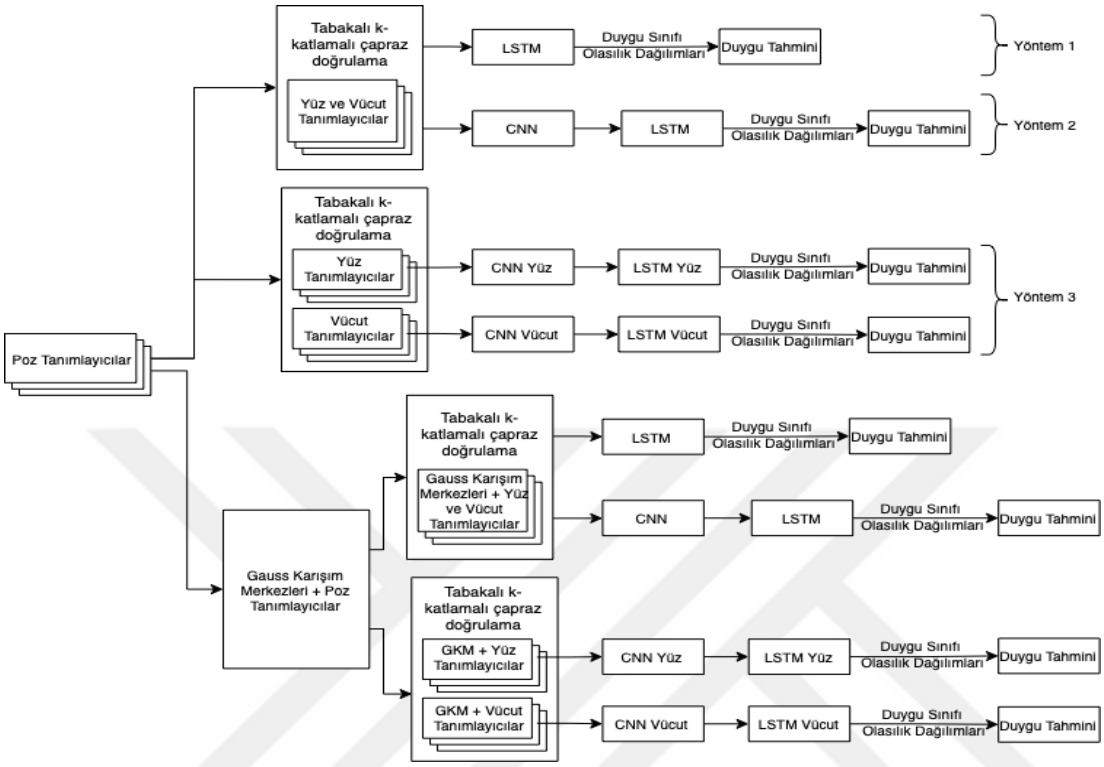
- Anahtar Kare Seçimi yaparak elde edilen yeni veri kümesi

Ardından bu farklı kümeler üzerinde LSTM hem de Dönüştürücü ağı için aşağıdaki 4 yaklaşım izlendi;

- Yöntem 1: Tüm öznitelik kümesi ile temel model eğitimi
- Yöntem 2: Tüm öznitelik kümesi ile ilk katmanı CNN olan iki katmanlı yapı eğitimi
- Yöntem 3: Yüz ve vücut özniteliklerini ayırarak ayrı ayrı 2 katmanlı yapı eğitimi
- Yöntem 4: Son olarak da Gauss Karışım Merkezi eklenerek ilk 3 yaklaşımın tekrarlanması

İlk yaklaşımda ön işleme aşamasında elde edilen tüm özniteliklerle temel model olarak LSTM ve Transformer ağları ayrı ayrı eğitildi ve test edildi. İkinci yaklaşımda yine tüm öznitelik seti ile CNN modeli beslendi ve çıktısı ikinci katman olan LSTM için girdi olarak kullanıldı. Bu işlem LSTM yerine Dönüştürü kullanılarak tekrarlandı. Ardından özniteliklerin başarıma etkisini görmek amacıyla yüz ve vücut öznitelikleri birbirinden ayrılarak iki farklı girdi seti oluşturuldu bunlarla yüz ve vücut için ayrı CNN modelleri beslendi ve çıktıları yine yüz ve vücut için ayrı ikinci katman modelleri için girdi olarak kullanıldı. Son olarak başarıma etkisini gözlemlemek amacıyla veriye 3.2 aşamasında bahsedildiği şekilde elde edilen Gauss Karışım Merkezleri öznitelik olarak eklenerek testler tekrarlandı. Sistemin çıktısı tüm duygu sınıfları için bir olasılık dağılımı olarak belirlendi. En yüksek olasılıklı sınıf ilgili videonun duygu sınıfı olarak kabul edildi.
Şekil 3.6 içerisinde görülen birinci yöntem sadece LSTM kullanılarak yapılan duygu sınıflandırma işlemini göstermektedir. Girdi olarak ön işleme adımından elde edilen poz tanımlayıcılar kullanıldı. Veri $k = 10$ olmak üzere tabakalı k-katlamalı çapraz doğrulama kullanılarak eğitim ve test kümelerine ayrıldı. Burada hem veri kümesinin boyutu küçük hem de sınıflar arası dağılım dengesiz olduğu için tabakalı k-katlamalı çapraz doğrulama tercih edildi. Daha sonra bir LSTM katmanı, bir düğüm seyreltme katmanı ve bir yoğun katmanlı bir LSTM ağı oluşturuldu. LSTM katmanının aktivasyon fonksiyonu olarak 'ReLu', son katmanın aktivasyon fonksiyonu da çok sınıflı biri sınıflandırma yapılacağı için 'softmax' olarak belirlendi, ağın kayıp fonksiyonu olarak Kategorik Çapraz Entropi

ve ’Adam’ optimizasyon yöntemi kullanıldı. LSTM ağı küme büyüklüğü 4 olmak üzere 250 devir eğitim sonucu yine 4 küme büyüklü ile test edildi. Bu eğitimin sonucu temel sonuç olarak baz alındı.



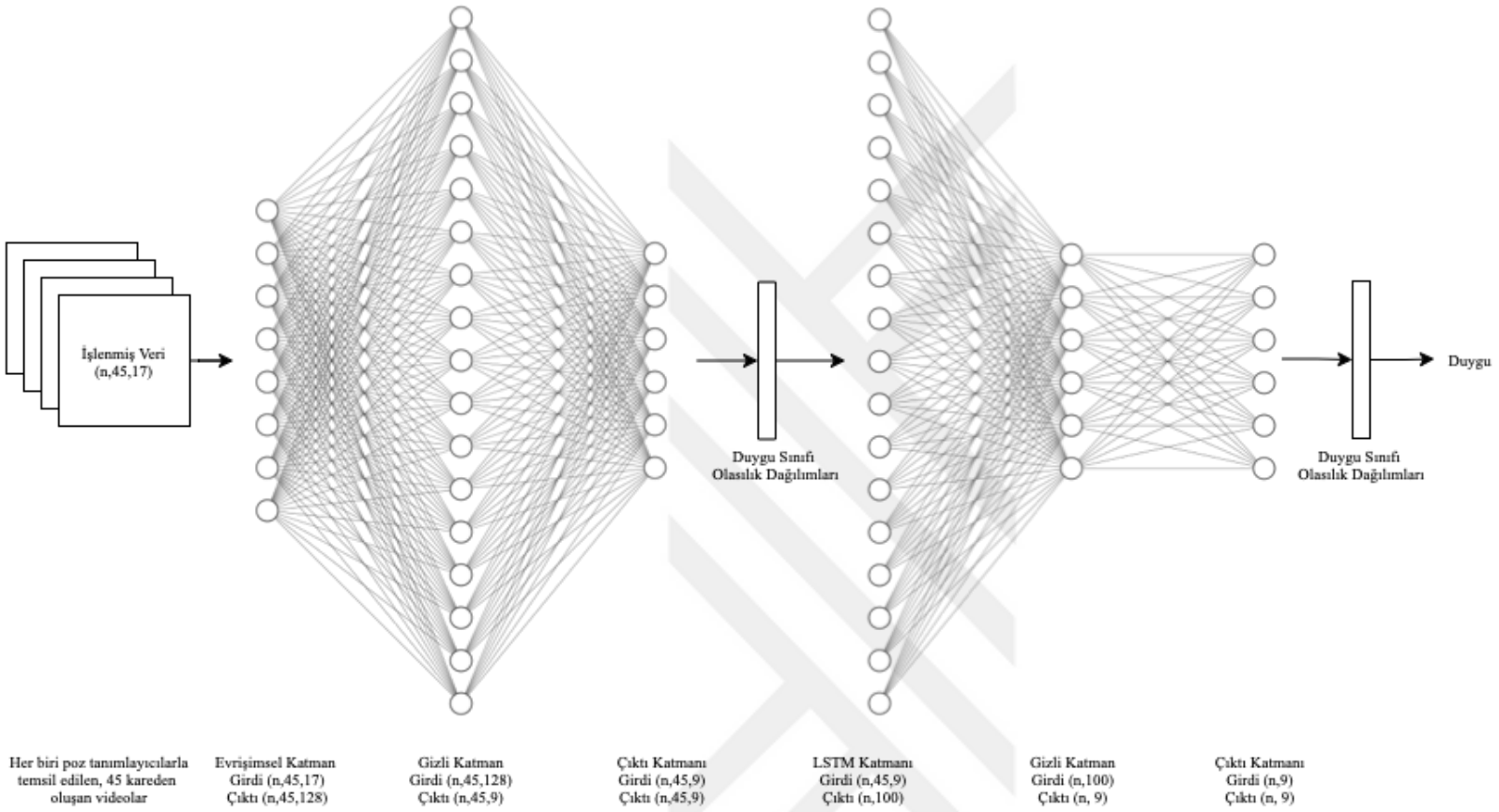
Şekil 3.6: Videolardan duygu sınıflandırma için oluşturulan çerçeve.

Şekil 3.6’daki ikinci yöntem ise 2 katmanlı yapıyı göstermektedir. Yine birinci katman daki veri ile bu yöntemde önce CNN ağı beslendi. CNN bloğu 1 evrişimli katman, 1 seyreltme katmanı ve 1 yoğun katmandan oluşmaktadır. Evrişimli katmanın filtre büyüklüğü 128, çekirdek büyüklüğü 8 olarak belirlendi ve aktivasyon fonksiyonu olarak LSTM bloğuna benzer şekilde ’ReLu’ kullanıldı. Yine tabakalı 10-katlamalı çapraz doğrulama ile 12 küme büyüklüğü ile 6000 devir eğitildi. CNN bloğundan duygu sınıfı dağılımlarını içeren bir vektör elde edildi ve bu çıktı ilk yöntemde anlatılan yapıya sahip bir LSTM bloğu için girdi olarak kullanıldı. Ardından LSTM bloğu bu girdi ile ilk yöntemde olduğu gibi eğitim ve teste tabi tutuldu.

Üçüncü olarak Şekil 3.6 yöntem 3’de görülen yüz ve vücut için ayrı deneyler yapıldı. İlk aşamada poz tanımlayıcılar yüz ve vücut olmak üzere ayrıldı ve ikinci yöntemde anlatılan iki katmanlı CNN ve LSTM yapısı yüz ve vücut öznitelikleri için ayrı ayrı eğitildi ve sonuçlar elde edildi. Son olarak da tüm veriye gauss karışım merkezleri

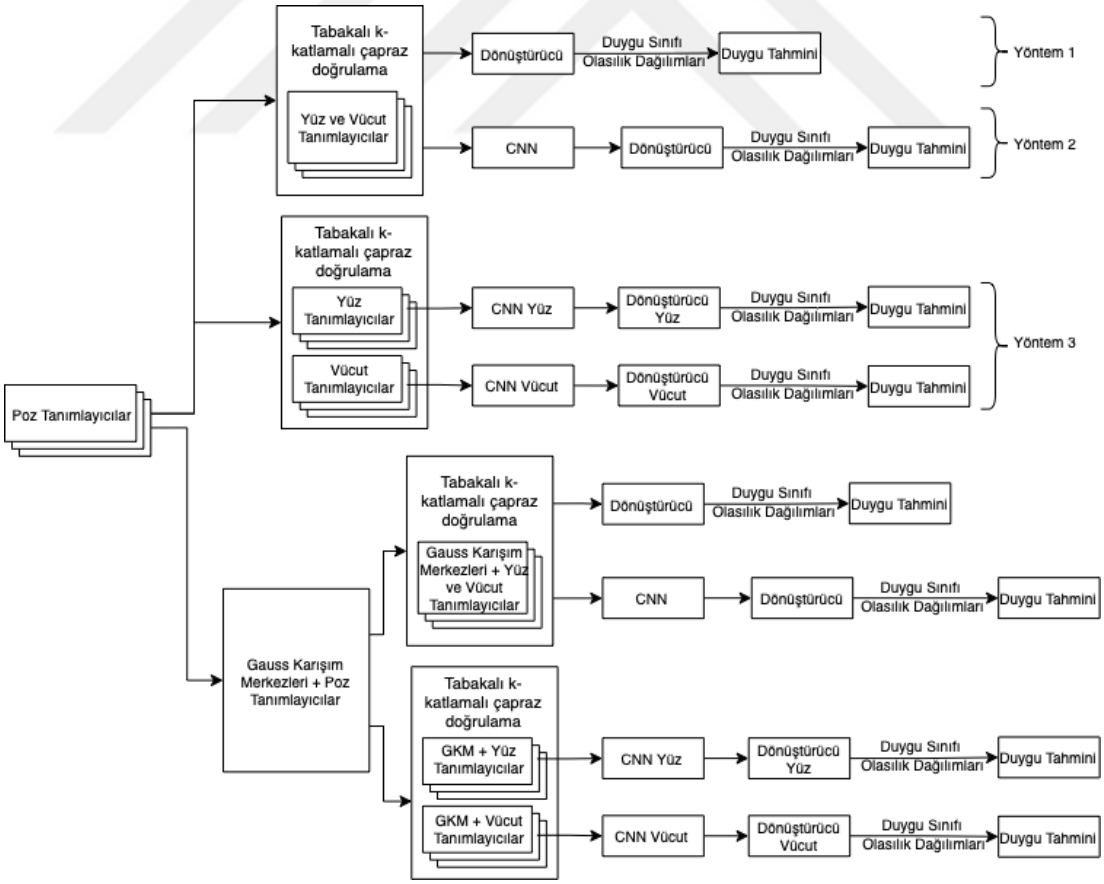
öznitelik olarak eklenerek birinci, ikinci, ve üçüncü yöntemler tekrarlandı. Duygu sınıfı çeşitliliğinin etkisini gözlemlemek üzere bu yaklaşımların hepsi FABO veri kümesi için hem 9 duyguyu içeren tüm veri seti hem de sadece 6 temel duygu içeren bir alt küme için tekrarlandı; CK+ veri kümesi için de benzer şekilde hem 7 duyguyu içeren tüm veri hem de sadece 6 duyguyu içeren bir alt kümede tekrarlandı. Burada 6 temel duygu seçilmesinin sebebi Ekman tarafından sunulan çalışmada [38] yapılan duygu tanımlarıdır. Her bir duygu için örneğin İğrenme için burun kırışıklığı, üst dudak kaldırıcı, çene kaldırıcı; Şaşkınlık için kaş kaldırma, duday ayrımı, düşük çene gibi, farklı duygu için eylem birimleri tanımlanmıştır.

Şekil 3.7'de iki katmanlı duygu sınıflandırma modeli gösterilmiştir. Bu örnekte FABO veri kümesi video çoklama ve anahtar kare seçimi olmadan çalıştığı durumdaki girdi ve çıktı boyutları görülmektedir. *n* girdi miktarı yani video sayısı olmak üzere, ek adımlardan geçmemiş bu videoların her biri 45 kareden oluşmakta ve her karenin Gauss Karışım Merkezi bilgisi ve 16 poz tanımlayıcı ile toplam 17 öznitelikten oluşan bir vektörle temsil edilmektedir. Dokuz sınıflı bir tahmin yapıldığı için CNN bloğundan sınıf olasılık dağılımları (1,9) noyutlarında elde edilir ve her bir kare için elde edilen bu dağılımlarla (n,45,9) boyutlarına sahip LSTM girdisi oluşturulur ve LSTM bloğu beslenir. Bu bloktan da sınıf olasılık dağılımları elde edilir ve en yüksek olasıklı sınıf duygu tahmini olacak şekilde çıktı elde edilir. Girdi ve çıktı şekilleri takip edilen farklı ön işleme adımlarına ve duygu sınıfı miktarı a göre değişiklik göstermektedir. Örneğin Gauss Merkezleri olmadan, Video Çoklama yöntemi izlenerek çalıştırıldığında girdi şekli (n, 30, 16) olmaktadır. Benzer şekilde sistem sadece 6 duygu için çalıştırıldığında CNN bloğunun çıktısı (n, 45, 6) olup LSTM bloğunun çıktısı da (1, 6) olur. CK+ veri kümesi üzerinde duygu sınıflandırılması yapıldığında da aynı şekilde değişiklik olmaktadır. Katmanların input ve output şekilleri parametrik olduğundan kullanılan veri izlenen ön işleme adımlarına göre otomatik olarak değişmektedir.



Şekil 3.7: Video çoklama ve anahtar kare seçimi olmadan Gauss Karışım Merkezli, dokuz duygu sınıfı için CNN-LSTM duygu sınıflandırma yapısı. (Görseldeki YSA'nın hücreleri semboliktir, girdi ve çıktı boyutları belirtilmiştir.)

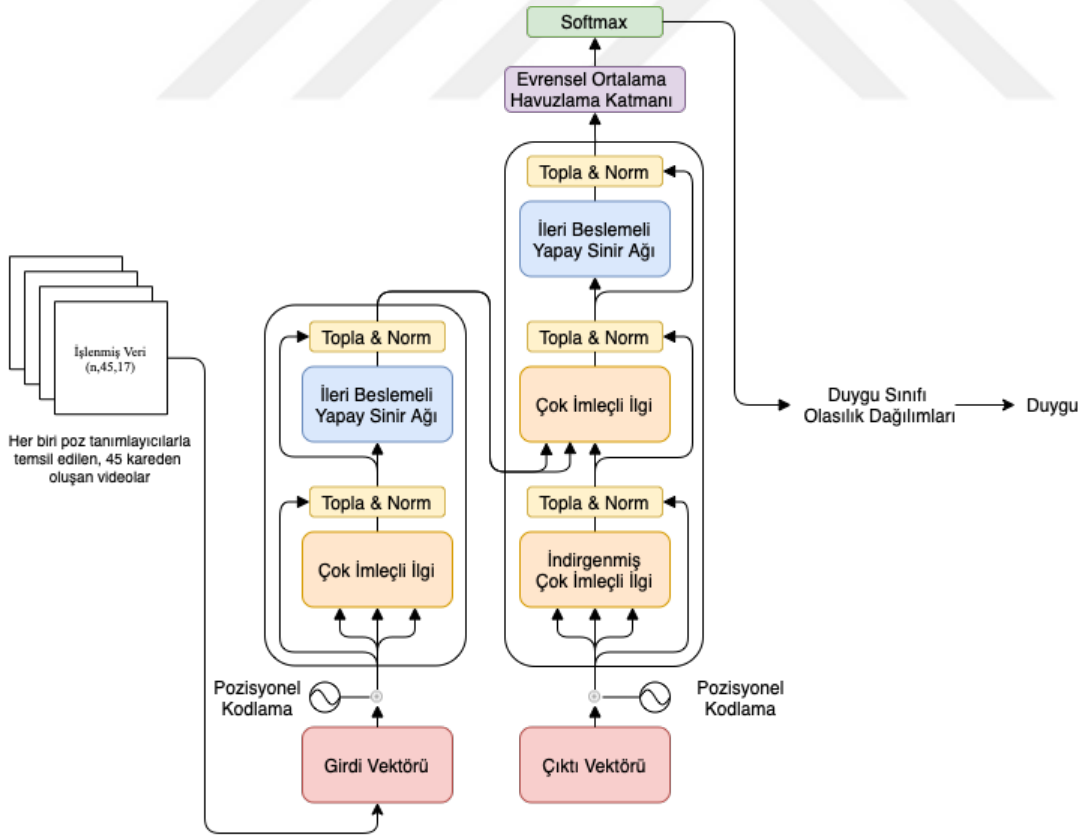
LSTM ağı her ne kadar bir hafıza yapısına sahip olsa da yapay sinir ağlarında ardışık girdilerde seri uzadıkça baştaki veriler önemini yitirmektedir. Bu tez çalışmasında da asıl amaç video gibi ardışık kareler içeren bir veri tipinde zaman içinde oluşan önem kaybını gidererek başarımı yükseltmektir. Bu sebeple bu sorunu Doğal Dil İşleme problemlerinde ortadan kaldırmış Çok-İmleçli İlgi Mekanizmalı Dönüştürücü ağı videodan duygu sınıflandırma problemine uyarlandı. Burada videolardan her bir kare için elde ettiğimiz poz tanımlayıcılar, Dönüştürücünün Doğal Dil İşleme uygulamaların da kullanılan kelime gömmeleri ne benzer işlen görmekte, her bir kare için oluşturulan poz tanımlayıcılar vektörü o karenin gömmesi olmaktadır. Yine benzer şekilde konumsal kodlama doğal dil işleme problemlerinde kelimenin cümledeki pozisyonu iken burada karenin videodaki sırası olmuştur. Özetle tüm videolar Çok-İmleçli İlgi Mekanizmalı Dönüştürücünün işleyebileceği, kare gömmeleri sekansına dönüştürüldü ve bu sayede tarihsel bilgi kaybı önlenmeye çalışıldı.



Şekil 3.8: Videolardan duygu sınıflandırma için Dönüştürücü ile oluşturulan çerçeve.

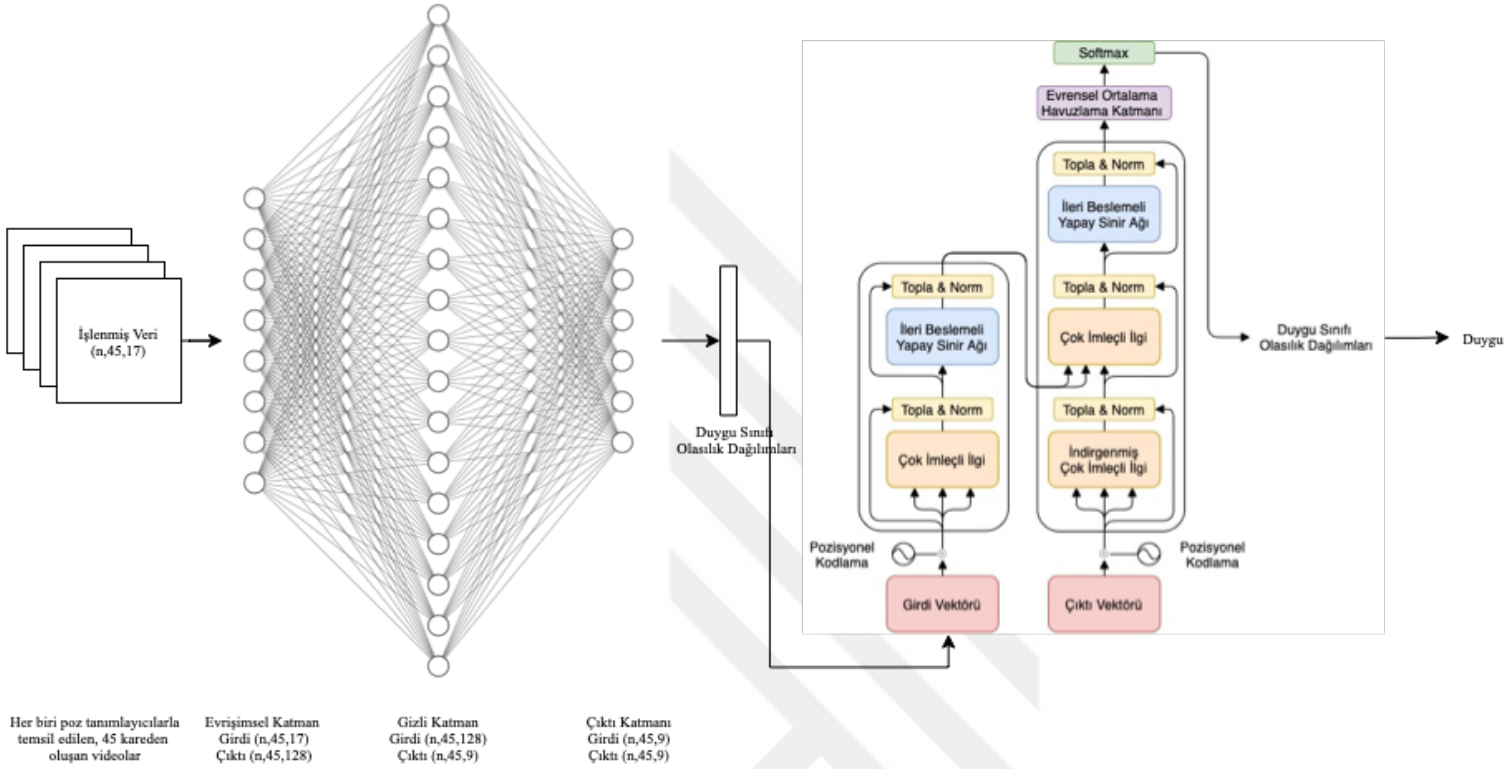
Dönüştürücü bloğu için bir çok-imleçli ilgi mekanizmalı dönüştürücü katmanı, ardından bir küresel ortalama havuzlama katmanı, bir düğüm seyreltme katmanı ve son olarak bir yoğun katman ile bir ağ oluşturuldu. 10 küme büyüklüğü ile 10 devir ve LSTM modeli ile aynı şekilde tabakalı 10-katlamalı çapraz doğrulama ile eğitildi ve test edildi. Aynı şekilde iki katmanlı yaklaşım, Yüz ve Vücut ayrı yaklaşım, gauss merkezli yaklaşım ve farklı duygu kümeli testler Şekil 3.8'de görülen şekilde LSTM yerine Dönüştürücü kullanılarak tekrarlandı.

Dönüştürücü modelinin duygu sınıflandırma amacıyla kullanımı Şekil 3.9'de gösterilmiştir. Gauss merkezi eklenmiş, video çoklama veya anahtar kare seçimi adımları uygulanmamış FABO veri kümesinden elde edilen (n, 45, 17) boyutlarındaki *n* video içeren veri ile öznitelikler (Poz Tanımlayıcılar ve Gauss Karışım Merkezi) Girdi Vektörü, kare pozisyonları Pozisyonel Kodlama olacak şekilde Dönüştürücü beslendi. Her bir video için kare gömmelerinden (öznitelik vektörü) duygu sınıfı olasılık dağılımları elde edildi ve en yüksek olasılığa sahip sınıf ilgili videonun duygusu olarak belirlendi.



Şekil 3.9: Dönüştürücü ile duygu sınıflandırma modeli

Şekil 3.10'da da CNN ön katmanıyla iki katmanlı CNN-Dönüştürücü modeli gösterilmiş tir. Burada da Dönüştürücü'nün Girdi Vektörü olarak CNN bloğundan elde edilen olasılık dağılımları kullanıldı, başka bir deyişle bu yapıda bu olasılık dağılımları kare gömmesi olarak kullanıldı; karelerin ilgili videodaki pozisyonu da yine pozisyonel kodlama görevi gördü. Dönüştücü bloğundan aynı şekilde olasılık dağılımı elde edildi ve ilgili video için tahmin edilen duygu sınıfı atandı.



Şekil 3.10: CNN-Dönüştürücü yapısı ile duygu sunuflandırma modeli (Görseldeki YSA'nın hücreleri semboliktir, girdi ve çıktı boyutları belirtilmiştir.)

4. SONUÇLAR

4.1 Deneysel Sonuçlar

Bölüm 3'te bahsedilen yapay sinir ağları ve ön işleme aşamalarının farklı kombinasyonla rıyla deneyler yürütüldü. Gauss Karışım merkezleri olmadan; tüm veri kümesi ile, genişletilmiş veri kümesi ile, seçilen karelerden oluşan veri kümesi ile ve bunların yüz ve vücut için ayrılmış verisoyunları ile LSTM, Dönüştürücü, CNN-LSTM ve CNN-Dönüştürücü modelleri eğitildi ve test edildi. Ardından verilerin 6, 9 ve 12 Gauss Karışım Merkezi eklenmiş versiyonu ile deneyler tekrarlandı. Benzer şekilde tüm bu deneyler FABO için altı ve dokuz farklı duygu sınıfıyla, CK+ için altı ve yedi farklı duygu sınıfıyla gerçekleştirildi

Çizelge 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 ve 4.8'de deneylerin sonucunda elde edilen model performansları detaylıca listelenmiştir. Duygu sınıflandırıcının eğitimin ardından ve test aşamasında yaptığı tahminlerden veri kümelerindeki örnekleri atadığı sınıflar elde edildi ve bunlarla sınıflandırıcı performansını ölçmek için her bir sınıfa ait aşağıdaki istatistikler çıkarıldı.

- DP (Doğru Pozitif): Gerçekte bir sınıfa ait olan örneklere tahmin sonucunda gerçek sınıflarının atanmış olması
- YP (Yanlış Pozitif): Gerçekte bir sınıfa ait olmayan örneklere tahmin sonucunda o sınıfın atanmış olması
- DN (Doğru Negatif): Gerçekte bir sınıfa ait olmayan örneklere tahmin sonucunda o sınıfın atanmamış olması
- YN (Yanlış Negatif): Gerçekte bir sınıfa ait olan örneklere tahmin sonucunda o sınıfın atanmamış olması

DP, YP, DN ve YN değerleri kullanılarak model performanslarını karşılaştırmak için Doğruluk 4.1, Kesinlik 4.2, Duyarlılık 4.3 ve F1-Skoru 4.4'te gösterilen formüllerle hesaplandı. Doğruluk, doğru tahmin edilen girdilerin tüm girdilere oranıdır. Model başarımı ölçümünde en yaygın kullanılan metrik olsa da tek başına yeterli değildir.

Özellikle sınıf dağılımlarının dengesiz olduğu durumlarda yanıltıcı olabilir. Bu sebeple diğer metrikler de hesaba katıldı. Kesinlik, bir sınıfa ait olarak tahmin edilen girdilerin gerçekte ne kadarının ilgili sınıfa ait olduğunu gösteren metriktir. Yanlış pozitif olarak adlandırılan hataların gözlenebilmesini sağlar. Duyarlılık ise gerçekte bir sınıfa ait olan girdilerin ne kadarını doğru tahmin ettiğimizi gösteren metriktir, yanlış negatif hataları görmemizi sağlar. F1-Skoru ise Kesinlik ve Duyarlılığın harmonik ortalaması alınarak elde edilir. Bu sayede daha güvenilir bir başarım ölçümü yapılabilir.

$$Dogruluk = (DP + DN)/(DP + YP + DN + YN) \tag{4.1}$$

$$Kesinlik = DP/(DP + YP) \tag{4.2}$$

$$Duyarlilik = DP/(DP + YN) \tag{4.3}$$

$$F1 - Skoru = 2x(KesinlikxDuyarlilik)/(Kesinlik + Duyarlilik) \tag{4.4}$$

Çizelge 4.1 ve 4.2'de görüldüğü üzere neredeyse tüm versiyonlarda henüz ön katman eklenmemiş durumda bile Dönüştürücü, LSTM'den daha iyi performans göstermektedir. Çizelge 4.1'de en yüksek performansa sahip modelin genişletilmiş veri seti ile eğitilmiş Dönüştürücü olduğunu görülmektedir. Çizelge 4.2'de ise GMK kullanılmayan LSTM modeli çok az farkla en iyi performansa sahip olsa da bir çok durumda Dönüştürücü daha iyi performans sergilemiştir.

Çizelge 4.3 ve 4.4, FABO ile eğitilen CNN-LSTM yapısının performansını göstermekte dir. Çizelge 4.3 yüz ve vücut özniteliklerinin birlikte kullanıldığı yapının sonuçlarını içerirken, Çizelge 4.4 bu özniteliklerin ayrı arı kullanıldı versiyonun sonuçlarını içermek tedir. Birleşik öznitelik kümesi için CNN-LSTM yapısının Doğruluğunun 6 duygu sınıfı için %44 ile %99 arasında, 9 duygu sınıfı için %31 ile %98 arasında uygulanan ön işleme adımlarına bağlı olarak değişiklik gösterdiği gözlemlendi. İki durumda da en yüksek performans video çoklama ile elde edilirken, 6 Duygu sınıfından oluşan veri kümesinde 9 duygu sınıfı içeren veriye göre az farkla daha yüksek skorlar gözlemlendi. 9 duygu

Çizelge 4.1: FABO Veri Kümesi kullanılarak CNN ön katmanı olmayan LSTM ve Dönüştürücü Modelleri için farklı yöntemlerle yapılan deney sonuçları

Yöntem				6 Duygu				9 Duygu			
Video Çoklama	Anahtar Kare Seçimi	Gauss K. M	YSA	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru
-	-	-	LSTM	0.418	0.595	0.295	0.394	0.331	0.404	0.085	0.140
		6 Merkez	LSTM	0.392	0.525	0.202	0.292	0.292	0.367	0.065	0.110
		9 Merkez	LSTM	0.397	0.545	0.232	0.325	0.221	0.227	0.054	0.087
		12 Merkez	LSTM	0.391	0.527	0.206	0.297	0.258	0.204	0.084	0.119
		-	Dönüştürücü	0.456	0.461	0.451	0.456	0.286	0.298	0.280	0.289
		6 Merkez	Dönüştürücü	0.465	0.479	0.460	0.469	0.334	0.348	0.317	0.332
		9 Merkez	Dönüştürücü	0.506	0.512	0.491	0.501	0.348	0.355	0.334	0.344
		12 Merkez	Dönüştürücü	0.427	0.436	0.423	0.429	0.309	0.315	0.292	0.303
+	30 Anahtar Kare	-	LSTM	0.579	0.586	0.572	0.579	0.301	0.350	0.228	0.276
		6 Merkez	LSTM	0.692	0.735	0.653	0.691	0.248	0.282	0.164	0.208
		9 Merkez	LSTM	0.422	0.444	0.396	0.419	0.180	0.180	0.148	0.162
		12 Merkez	LSTM	0.545	0.556	0.527	0.541	0.181	0.183	0.164	0.173
		-	Dönüştürücü	**0.983**	**0.983**	**0.983**	**0.983**	**0.975**	**0.975**	**0.973**	**0.974**
		6 Merkez	Dönüştürücü	**0.978**	**0.979**	**0.977**	**0.978**	**0.968**	**0.968**	**0.968**	**0.968**
		9 Merkez	Dönüştürücü	**0.975**	**0.975**	**0.975**	**0.975**	0.948	0.952	0.947	0.949
		12 Merkez	Dönüştürücü	**0.981**	**0.981**	**0.981**	**0.981**	0.944	0.945	0.943	0.944
+	10 Anahtar Kare	-	LSTM	0.967	0.969	0.964	0.967	0.943	0.949	0.939	0.944
		6 Merkez	LSTM	0.951	0.954	0.949	0.951	0.919	0.924	0.915	0.919
		9 Merkez	LSTM	0.907	0.913	0.904	0.908	0.895	0.907	0.890	0.898
		12 Merkez	LSTM	0.953	0.957	0.950	0.953	0.918	0.924	0.911	0.917
		-	Dönüştürücü	0.927	0.929	0.927	0.928	0.887	0.889	0.882	0.886
		6 Merkez	Dönüştürücü	0.951	0.951	0.948	0.949	0.870	0.878	0.867	0.873
		9 Merkez	Dönüştürücü	0.939	0.940	0.938	0.939	0.853	0.863	0.844	0.854
		12 Merkez	Dönüştürücü	0.935	0.939	0.935	0.937	0.853	0.861	0.844	0.852

Çizelge 4.2: CK+ Veri Kümesi kullanılarak CNN ön katmanı olmayan LSTM ve Dönüştürücü Modelleri için farklı yöntemlerle yapılan deney sonuçları

Model				6 Duygu				7 Duygu			
Video Çoklama	Anahtar Kare Seçimi	Gauss K. M	YSA	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru
-	6 Anahtar Kare	-	LSTM	0.713	0.763	0.700	0.730	0.711	0.738	0.665	0.699
		6 Merkez	LSTM	0.655	0.686	0.648	0.666	0.618	0.645	0.600	0.622
		9 Merkez	LSTM	0.641	0.659	0.635	0.647	0.619	0.643	0.616	0.629
		12 Merkez	LSTM	0.674	0.688	0.639	0.662	0.622	0.640	0.610	0.624
		-	Dönüştürücü	0.716	0.725	0.687	0.706	0.671	0.697	0.640	0.668
		6 Merkez	Dönüştürücü	0.710	0.735	0.694	0.714	0.664	0.713	0.646	0.678
		9 Merkez	Dönüştürücü	0.674	0.713	0.645	0.677	0.655	0.694	0.630	0.661
		12 Merkez	Dönüştürücü	0.661	0.691	0.635	0.662	0.664	0.698	0.618	0.656
+	6 Anahtar Kare	-	LSTM	**0.838**	**0.854**	**0.832**	**0.843**	**0.837**	**0.854**	**0.819**	**0.836**
		6 Merkez	LSTM	0.757	0.773	0.753	0.763	0.770	0.781	0.752	0.766
		9 Merkez	LSTM	0.745	0.761	0.739	0.750	0.709	0.735	0.706	0.720
		12 Merkez	LSTM	0.726	0.737	0.722	0.730	0.698	0.715	0.680	0.697
		-	Dönüştürücü	**0.826**	**0.830**	**0.819**	**0.824**	**0.791**	**0.815**	**0.774**	**0.794**
		6 Merkez	Dönüştürücü	**0.784**	**0.790**	**0.774**	**0.782**	**0.743**	0.761	0.728	0.744
		9 Merkez	Dönüştürücü	0.763	0.784	0.747	0.765	0.741	0.770	0.698	0.732
		12 Merkez	Dönüştürücü	0.770	0.784	0.751	0.767	0.761	0.787	0.735	0.760

sınıfı için en yüksek performansa %98.58 doğruluk olmak üzere ile GKM olmadan video çoklama ile ulaşılılırken GMK uygulamak 6 duygu sınıfı için performansı arttırdı. Tüm versiyonlar arasında en iyi performansı gösteren model %99 ve üzeri doğruluk ile video çoklama ve GMK aşamaları uygulanmış 6 duygu sınıfı ile çalıştırılan modellerdir. Yüz ve vücut öznitelikleri ayrılarak ayrı modeller eğitildiği zaman CNN-LSTM modeli nin doğruluk değeri yaklaşık %20-%30 arası düşüş gösterdi. Çizelge 4.4'te görüldüğü gibi, bu durum sadece video çoklama ve anahtar kare seçimi yapılmadığı durumda gözlemlendi, tüm veri setiyle elde edilen sonuçların aksine burada ön işleme yapılma

Çizelge 4.3: FABO veri kümesi üzerinde farklı ön işleme adımları uygulanarak yapılan deneyler sonucu CNN-LSTM ağının performansı

Yöntem				6 Duygu				9 Duygu			
Video Çoklama	Anahtar Kare Seçimi	Gauss K. M	YSA	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru
-	-	-	CNN	0.4516	0.5301	0.4028	0.4578	0.3278	0.4175	0.2124	0.2816
			LSTM	0.5593	0.6355	0.5055	0.5631	0.3740	0.4150	0.3541	0.3822
		6 Merkez	CNN	0.4476	0.4870	0.4040	0.4416	0.3103	0.3715	0.2156	0.2728
			LSTM	0.5205	0.5689	0.4810	0.5212	0.3715	0.3732	0.3431	0.3575
		9 Merkez	CNN	0.4529	0.5057	0.3982	0.4455	0.3161	0.3765	0.2240	0.2808
			LSTM	0.5300	0.6133	0.4564	0.5233	0.4137	0.4348	0.3826	0.4070
		12 Merkez	CNN	0.4423	0.4872	0.3988	0.4386	0.3137	0.3632	0.2113	0.2672
			LSTM	0.4955	0.5433	0.4221	0.4751	0.3769	0.4179	0.3627	0.3883
+	30 Anahtar Kare	-	CNN	0.9341	0.9374	0.9305	0.9339	0.7734	0.8416	0.7005	0.7646
			LSTM	0.9598	0.9626	0.9578	0.9602	**0.9858**	**0.9870**	**0.9853**	**0.9861**
		6 Merkez	CNN	0.9364	0.9402	0.9330	0.9366	0.7823	0.8502	0.7146	0.7765
			LSTM	**0.9971**	**0.9980**	**0.9971**	**0.9975**	0.9610	0.9629	0.9599	0.9614
		9 Merkez	CNN	0.9395	0.9435	0.9370	0.9402	0.7998	0.8577	0.7417	0.7955
			LSTM	**0.9961**	**0.9970**	**0.9951**	**0.9961**	0.8988	0.9004	0.8960	0.8982
		12 Merkez	CNN	0.9422	0.9453	0.9392	0.9423	0.7810	0.8538	0.7079	0.7740
			LSTM	**0.9951**	**0.9970**	**0.9941**	**0.9956**	0.7674	0.7676	0.7674	0.7675
+	10 Anahtar Kare	-	CNN	0.9216	0.9222	0.9213	0.9217	0.7881	0.7977	0.7795	0.7885
			LSTM	0.9706	0.9773	0.9657	0.9714	0.9042	0.9116	0.8946	0.9030
		6 Merkez	CNN	0.9218	0.9228	0.9217	0.9222	0.7986	0.8090	0.7908	0.7998
			LSTM	0.9706	0.9782	0.9667	0.9724	0.9127	0.9258	0.8974	0.9114
		9 Merkez	CNN	0.9225	0.9231	0.9220	0.9225	0.7814	0.7944	0.7696	0.7818
			LSTM	0.9725	0.9754	0.9686	0.9720	0.8969	0.9195	0.8760	0.8972
		12 Merkez	CNN	0.9234	0.9248	0.9230	0.9239	0.7952	0.8073	0.7840	0.7955
			LSTM	0.9696	0.9733	0.9657	0.9695	0.9060	0.9141	0.8969	0.9054

dığında elde edilen sonuçlar diğer versiyonlara daha yakındır. En iyi sonuçlar video çoklama ile, en iyi 10 anahtar kare seçimi yapıldığında yaklaşık %78 doğruluk değeri şeklinde gözlemlenmiş, yine 6 duygu sınıfıyla yapılan deneyin sonuçlarının 9 duygu sınıfından daha yüksek olduğu görüldü. Bunun yanında ayrı öznitelik setleri ile yapılan deneyde Gauss Karışım Merkezleri'nin performansa kayda değer bir etkisi olmadığı da gözlemlendi. Bunun yanında özniteliklerin bir arada olduğu veri setinde olduğu gibi 6 duygunun aksine, 9 duygu sınıfı içeren 30 kareli videolarda Gauss Karışım Merkezi miktarı artırıldığında LSTM başarısının düştüğü gözlemlendi. Yine benzer şekilde GMK kullanılan 9 duygu içeren veri kümelerinde 10 anahtar kareli videolarda serigelen performans 30 anahtar kareli videolardan daha yüksektir. Örneğin Çizelge 4.4'de, 12 merkezli GMK, 30 anahtar kareli videolar kullanıldığında %76 doğruluk elde edilirken, 10 anahtar kareli videolarda doğruluk %90'dır.

CNN-Dönüştürücü yapısı için FABO veri kümesi tüm veri kümesi ve yüz ve vücut özniteliklerinin ayrıldığı veri kümeleri ile yapılan deneylerin sonuçları Çizelge 4.5 ve 4.6'da sırasıyla listelenmiştir. Tüm özniteliklerin bir arada kullanıldığı deneylerde CNN-Dönüştürücü Modeli, hem 6 duygu hem de 9 duygu sınıfı kullanıldığında 30 anahtar kareli Video Çoklama ile %99 doğruluk elde etti. Bu da CNN-Dönüştürücü

Çizelge 4.4: FABO veri kümesi üzerinde farklı ön işleme adımları uygulanarak ve Yüz ve Vücut Tanımlayıcıların ayrı ele alarak yapılan deneyler sonucu CNN-LSTM ağının performansı

Yöntem					6 Duygu				9 Duygu			
Video Çoklama	Anahtar Kare Seçimi	Gauss K. M	Öznitelik	YSA	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru
-	-	-	Yüz	CNN	0.551	0.737	0.343	0.468	0.412	0.705	0.145	0.240
			Vücut	CNN	0.549	0.733	0.338	0.463	0.412	0.702	0.143	0.238
			Yüz	LSTM	0.548	0.733	0.338	0.463	0.411	0.701	0.143	0.238
			Vücut	LSTM	0.548	0.733	0.338	0.463	0.411	0.701	0.143	0.238
		6 Merkez	Yüz	CNN	0.566	0.742	0.365	0.489	0.411	0.703	0.155	0.254
			Vücut	CNN	0.565	0.740	0.364	0.488	0.412	0.701	0.156	0.255
			Yüz	LSTM	0.565	0.741	0.364	0.489	0.411	0.702	0.156	0.255
			Vücut	LSTM	0.565	0.741	0.364	0.489	0.411	0.702	0.156	0.255
		9 Merkez	Yüz	CNN	0.554	0.727	0.348	0.471	0.408	0.700	0.149	0.246
			Vücut	CNN	0.554	0.725	0.348	0.470	0.410	0.698	0.150	0.247
			Yüz	LSTM	0.553	0.725	0.348	0.470	0.409	0.698	0.150	0.247
			Vücut	LSTM	0.553	0.725	0.348	0.470	0.409	0.698	0.150	0.247
		12 Merkez	Yüz	CNN	0.570	0.744	0.369	0.493	0.419	0.705	0.164	0.265
			Vücut	CNN	0.571	0.741	0.370	0.493	0.421	0.704	0.164	0.267
			Yüz	LSTM	0.570	0.741	0.369	0.493	0.420	0.703	0.165	0.267
			Vücut	LSTM	0.570	0.741	0.369	0.493	0.420	0.703	0.165	0.267
+	30 Anahtar Kare	-	Yüz	CNN	0.672	0.762	0.545	0.636	0.487	0.696	0.231	0.347
			Vücut	CNN	0.678	0.764	0.552	0.641	0.493	0.698	0.234	0.351
			Yüz	LSTM	0.677	0.763	0.551	0.640	0.492	0.697	0.233	0.350
			Vücut	LSTM	0.677	0.763	0.551	0.640	0.492	0.697	0.233	0.350
		6 Merkez	Yüz	CNN	0.679	0.771	0.559	0.648	0.486	0.700	0.234	0.351
			Vücut	CNN	0.685	0.772	0.566	0.653	0.493	0.702	0.238	0.355
			Yüz	LSTM	0.685	0.772	0.566	0.653	0.491	0.701	0.237	0.354
			Vücut	LSTM	0.685	0.772	0.566	0.653	0.491	0.701	0.237	0.354
		9 Merkez	Yüz	CNN	0.679	0.772	0.559	0.648	0.484	0.700	0.234	0.350
			Vücut	CNN	0.687	0.774	0.569	0.656	0.491	0.701	0.237	0.354
			Yüz	LSTM	0.687	0.774	0.569	0.656	0.489	0.701	0.237	0.354
			Vücut	LSTM	0.687	0.774	0.569	0.656	0.489	0.701	0.237	0.354
		12 Merkez	Yüz	CNN	0.677	0.767	0.557	0.645	0.483	0.701	0.231	0.348
			Vücut	CNN	0.683	0.769	0.565	0.651	0.490	0.702	0.235	0.352
			Yüz	LSTM	0.683	0.769	0.564	0.651	0.488	0.702	0.234	0.351
			Vücut	LSTM	0.683	0.769	0.564	0.651	0.488	0.702	0.234	0.351
+	10 Anahtar Kare	-	Yüz	CNN	0.779	0.854	0.769	0.809	0.541	0.737	0.420	0.535
			Vücut	CNN	**0.782**	**0.856**	**0.774**	**0.813**	0.541	0.738	0.427	0.541
			Yüz	LSTM	**0.785**	**0.856**	**0.774**	**0.813**	0.544	0.738	0.427	0.541
			Vücut	LSTM	**0.785**	**0.856**	**0.774**	**0.813**	0.544	0.738	0.427	0.541
		6 Merkez	Yüz	CNN	0.771	0.851	0.760	0.803	0.543	0.739	0.425	0.540
			Vücut	CNN	0.774	0.853	0.766	0.807	0.543	0.740	0.433	0.546
			Yüz	LSTM	0.778	0.853	0.766	0.807	0.547	0.740	0.431	0.545
			Vücut	LSTM	0.778	0.853	0.766	0.807	0.547	0.740	0.431	0.545
		9 Merkez	Yüz	CNN	0.778	0.857	0.768	0.810	0.536	0.737	0.412	0.528
			Vücut	CNN	**0.782**	**0.860**	**0.774**	**0.815**	0.536	0.738	0.418	0.534
			Yüz	LSTM	**0.786**	**0.860**	**0.775**	**0.815**	0.540	0.738	0.418	0.534
			Vücut	LSTM	**0.786**	**0.860**	**0.775**	**0.815**	0.540	0.738	0.418	0.534
		12 Merkez	Yüz	CNN	0.772	0.854	0.759	0.804	0.534	0.734	0.408	0.524
			Vücut	CNN	0.776	0.856	0.766	0.809	0.534	0.735	0.415	0.531
			Yüz	LSTM	0.779	0.856	0.766	0.808	0.537	0.734	0.413	0.529
			Vücut	LSTM	0.779	0.856	0.766	0.808	0.537	0.734	0.413	0.529

yapısının en FABO veri kümesi üzerinde yapılan deneylerde en iyi performansa sahip model olduğunu gösterdi. CNN-LSTM yapısına ve temel modellere benzer şekilde seçilen anahtar kare sayısı azaltıldığında başarımın düştüğü gözlemlendi.

Çizelge 4.6'da listelendiği üzere, toplu veri kümesine benzer şekilde 6 duygu sınıfı ile yapılan Video çoklamalı deneyler 30 kare seçildiğinde sadece vücut öznitelikleri kullanıldığında %99 doğruluk elde etti ancak sadece yüzden elde edilen poz tanımlayıcı lar kullanıldığında elde edilen doğruluk &96'da kaldı. 9 duygu sınıfı için de benzer şekilde vücut tanımlayıcıları ile %90 civarı doğruluk elde edilirken, yüz tanımlayıcıları ile %87 doğruluk gözlemlendi. Toplu veri kümesi ile benzer şekilde, seçilen anahtar

Çizelge 4.5: FABO veri kümesi üzerinde farklı ön işleme adımları uygulanarak yapılan deneyler sonucu CNN-Dönüştürücü ağının performansı

Yöntem				6 Duygu				9 Duygu			
Video Çoklama	Anahtar Kare Seçimi	Gauss K. M	YSA	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru
-	-	-	CNN	0.452	0.509	0.386	0.439	0.316	0.414	0.211	0.280
			Dönüştürücü	0.472	0.682	0.319	0.434	0.374	0.457	0.258	0.330
		6 Merkez	CNN	0.444	0.493	0.395	0.439	0.318	0.395	0.222	0.284
			Dönüştürücü	0.505	0.667	0.309	0.423	0.380	0.465	0.292	0.359
		9 Merkez	CNN	0.445	0.493	0.395	0.439	0.331	0.406	0.212	0.279
			Dönüştürücü	0.476	0.678	0.344	0.457	0.399	0.500	0.289	0.366
		12 Merkez	CNN	0.440	0.480	0.390	0.430	0.307	0.365	0.216	0.271
			Dönüştürücü	0.486	0.663	0.324	0.435	0.343	0.449	0.247	0.318
+	30 Anahtar Kare	-	CNN	0.936	0.939	0.933	0.936	0.778	0.847	0.705	0.770
			Dönüştürücü	**0.999**	**1.000**	**0.999**	**1.000**	**0.989**	**0.990**	**0.988**	**0.989**
		6 Merkez	CNN	0.936	0.939	0.932	0.936	0.786	0.851	0.721	0.781
			Dönüştürücü	**0.999**	**0.999**	**0.997**	**0.998**	**0.992**	**0.993**	**0.990**	**0.991**
		9 Merkez	CNN	0.945	0.948	0.942	0.945	0.787	0.851	0.719	0.779
			Dönüştürücü	**0.999**	**0.997**	**0.999**	**0.998**	**0.992**	**0.993**	**0.991**	**0.992**
		12 Merkez	CNN	0.937	0.941	0.934	0.938	0.783	0.852	0.715	0.778
			Dönüştürücü	**0.999**	**0.999**	**0.999**	**0.999**	**0.991**	**0.991**	**0.989**	**0.990**
+	10 Anahtar Kare	-	CNN	0.925	0.926	0.925	0.925	0.792	0.801	0.784	0.792
			Dönüştürücü	0.983	0.987	0.978	0.983	0.930	0.939	0.924	0.931
		6 Merkez	CNN	0.931	0.932	0.931	0.931	0.802	0.811	0.795	0.803
			Dönüştürücü	0.983	0.983	0.980	0.982	0.948	0.962	0.940	0.951
		9 Merkez	CNN	0.925	0.926	0.925	0.925	0.806	0.815	0.799	0.807
			Dönüştürücü	0.980	0.982	0.973	0.977	0.942	0.955	0.934	0.944
		12 Merkez	CNN	0.925	0.926	0.925	0.926	0.786	0.799	0.775	0.787
			Dönüştürücü	0.981	0.987	0.978	0.983	0.941	0.950	0.935	0.942

kare miktarı azaltıldığında model performanslarında düşüş görüldü. Dönüştürücü ile yapılan deneylerde Gauss Karışım Merkezlerinin model performanslarına kayda değer bir etkisi olmadığı görüldü.

Çizelge 4.7 ve 4.8'de sırasıyla CNN-LSTM ve CNN-Dönüştürücü yapıları için CK+ veri kümesi kullanılarak yapılan deneylerin sonuçları listelenmiştir. Ön işleme aşamasında anlatıldığı üzere bu veri kümesinin yapısından dolayı tüm videolar 6 anahtar kare seçilerek 6 kareye indirgenmiş ve deneyler bu şekilde yürütülmüştür. Burada FABO'dan farklı olarak CNN-LSTM yapısı CNN-Dönüştürücü yapısından daha iyi performans sergiledi. Bu veri kümesi için en başarılı model, FABO kümesinde elde edilen değerler den düşük olsa da, %91 ile 9 gauss karışım merkezli, video çoklamalı CNN-LSTM modelidir. CNN-Döüştürücü yapısıyla elde edilen en yüksek doğruluk ise %83 olup, GMK kullanılmadan video çoklmaa ile elde edilmiştir.

Çizelge 4.9, literatürdeki güncel, yüksek performanslı modellerin doğrulukları ile tez çalışmasında önerilen modellerle elde edilen doğrulukların karşılaştırmasını içerir. İlgili modeller, 3D-CNN+LSTM with Keyframes Selection [22], Multichannel CNN (MCCNN) [14] FABO veri setini kullanmış, CNN-SIFT [15], Fusion of LEMHI-VGG, VGG-CTSLSTM [17] ve Nested LSTM (STC-NLSTM) [19] ise CK+ veri kümesi kullanılarak test edilmiş. Bu modellerin duygu sınıflandırma doğruluğu, tez

Çizelge 4.6: FABO veri kümesi üzerinde farklı ön işleme adımları uygulanarak ve Yüz ve Vücut Tanımlayıcıların ayrı ele alarak yapılan deneyler sonucu CNN-Dönüştürücü ağının performansı.

Yöntem					6 Duygu				9 Duygu			
Video Çoklama	Anahtar Kare Seçimi	Gauss K. M	Öznitelik	YSA	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru
-	-	-	Yüz	CNN	0.568	0.760	0.366	0.494	0.416	0.727	0.156	0.257
			Vücut	CNN	0.541	0.732	0.331	0.456	0.404	0.704	0.146	0.242
			Yüz	Dönüştürücü	0.550	0.719	0.285	0.409	0.420	0.575	0.261	0.359
			Vücut	Dönüştürücü	0.389	0.519	0.231	0.320	0.286	0.387	0.159	0.225
		6 Merkez	Yüz	CNN	0.584	0.767	0.390	0.517	0.398	0.713	0.146	0.243
			Vücut	CNN	0.576	0.752	0.386	0.510	0.406	0.702	0.155	0.254
			Yüz	Dönüştürücü	0.466	0.674	0.299	0.414	0.394	0.496	0.223	0.308
			Vücut	Dönüştürücü	0.384	0.442	0.197	0.273	0.249	0.337	0.150	0.208
		9 Merkez	Yüz	CNN	0.547	0.741	0.336	0.462	0.389	0.707	0.134	0.226
			Vücut	CNN	0.546	0.730	0.345	0.468	0.399	0.696	0.145	0.241
			Yüz	Dönüştürücü	0.525	0.668	0.305	0.419	0.388	0.552	0.230	0.325
			Vücut	Dönüştürücü	0.392	0.570	0.211	0.308	0.269	0.368	0.173	0.235
		12 Merkez	Yüz	CNN	0.554	0.740	0.348	0.474	0.381	0.692	0.123	0.209
			Vücut	CNN	0.549	0.726	0.344	0.467	0.393	0.686	0.138	0.230
			Yüz	Dönüştürücü	0.520	0.627	0.298	0.404	0.371	0.483	0.184	0.267
			Vücut	Dönüştürücü	0.369	0.503	0.242	0.327	0.266	0.339	0.150	0.208
+	30 Anahtar Kare	-	Yüz	CNN	0.651	0.758	0.510	0.610	0.456	0.696	0.207	0.319
			Vücut	CNN	0.681	0.766	0.554	0.643	0.475	0.697	0.230	0.345
			Yüz	Dönüştürücü	0.960	0.965	0.957	0.961	0.873	0.894	0.858	0.876
			Vücut	Dönüştürücü	**0.991**	**0.993**	**0.989**	**0.991**	**0.915**	**0.933**	**0.902**	**0.917**
		6 Merkez	Yüz	CNN	0.657	0.763	0.519	0.618	0.455	0.697	0.209	0.321
			Vücut	CNN	0.688	0.771	0.566	0.653	0.469	0.698	0.224	0.340
			Yüz	Dönüştürücü	0.955	0.959	0.947	0.953	0.868	0.887	0.844	0.865
			Vücut	Dönüştürücü	**0.991**	**0.993**	**0.986**	**0.990**	**0.909**	**0.927**	**0.895**	**0.911**
		9 Merkez	Yüz	CNN	0.650	0.758	0.510	0.610	0.458	0.697	0.212	0.325
			Vücut	CNN	0.677	0.763	0.549	0.639	0.480	0.705	0.239	0.357
			Yüz	Dönüştürücü	0.960	0.965	0.953	0.959	0.870	0.893	0.848	0.870
			Vücut	Dönüştürücü	0.987	0.989	0.982	0.986	0.938	0.948	0.922	0.935
		12 Merkez	Yüz	CNN	0.653	0.762	0.513	0.613	0.453	0.695	0.208	0.321
			Vücut	CNN	0.686	0.771	0.561	0.650	0.476	0.705	0.235	0.352
			Yüz	Dönüştürücü	0.965	0.970	0.955	0.962	0.874	0.894	0.853	0.873
			Vücut	Dönüştürücü	**0.993**	**0.996**	**0.992**	**0.994**	**0.912**	**0.926**	**0.898**	**0.912**
+	10 Anahtar Kare	-	Yüz	CNN	0.786	0.838	0.728	0.779	0.555	0.727	0.362	0.483
			Vücut	CNN	0.813	0.854	0.768	0.809	0.574	0.733	0.414	0.530
			Yüz	Dönüştürücü	0.917	0.927	0.902	0.914	0.819	0.837	0.803	0.820
			Vücut	Dönüştürücü	0.934	0.957	0.924	0.940	0.848	0.867	0.837	0.851
		6 Merkez	Yüz	CNN	0.784	0.842	0.722	0.778	0.548	0.726	0.355	0.477
			Vücut	CNN	0.815	0.858	0.771	0.812	0.567	0.732	0.406	0.522
			Yüz	Dönüştürücü	0.891	0.905	0.874	0.889	0.792	0.821	0.781	0.800
			Vücut	Dönüştürücü	0.925	0.940	0.908	0.924	0.836	0.857	0.820	0.838
		9 Merkez	Yüz	CNN	0.774	0.835	0.707	0.766	0.557	0.732	0.374	0.495
			Vücut	CNN	0.806	0.852	0.758	0.802	0.579	0.739	0.427	0.541
			Yüz	Dönüştürücü	0.899	0.918	0.881	0.899	0.789	0.813	0.768	0.790
			Vücut	Dönüştürücü	0.945	0.957	0.932	0.944	0.858	0.878	0.842	0.860
		12 Merkez	Yüz	CNN	0.764	0.832	0.692	0.755	0.550	0.730	0.361	0.483
			Vücut	CNN	0.800	0.850	0.749	0.796	0.568	0.736	0.409	0.526
			Yüz	Dönüştürücü	0.888	0.906	0.876	0.891	0.803	0.822	0.786	0.804
			Vücut	Dönüştürücü	0.945	0.957	0.933	0.945	0.841	0.866	0.828	0.846

çalışmasında önerilen modellerin bazılarından daha iyi performans göstermiş olsa da çoğunlukla önerilen model öne çıkmaktadır. 9 ve 12 Gauss Karışım Merkezli CNN-Dönüştürücü modeli, FABO veri seti kullanılarak oluşturulan veri üzerinde %99 sınıflandırma doğruluğu elde ederek diğer tüm modellerden net bir şekilde daha iyi performans gösterdi. Ayrıca tez çalışmasında önerilen modeller, FABO veri kümesi ve farklı ön işleme yöntemi kombinasyonlarıyla mevcut MCCNN modeline kıyasla %8 daha yüksek bir değere elde ederek %99 doğrulukla en başarılı model olmuş oldu.

Çizelge 4.7: CK+ veri kümesi üzerinde farklı ön işleme adımları uygulanarak yapılan deneyler sonucu CNN-LSTM ağının performansı

Yöntem				6 Duygu				7 Duygu			
Video Çoklama	Anahtar Kare Seçimi	Gauss K. M	YSA	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru
-	6 Anahtar Kare	-	CNN	0.881	0.909	0.855	0.881	0.848	0.895	0.804	0.847
			LSTM	0.881	0.909	0.855	0.881	0.848	0.895	0.804	0.847
		6 Merkez	CNN	0.917	0.935	0.899	0.916	0.895	0.923	0.868	0.895
			LSTM	0.917	0.935	0.899	0.917	0.895	0.923	0.868	0.895
		9 Merkez	CNN	0.915	0.934	0.897	0.915	0.893	0.921	0.865	0.892
			LSTM	0.915	0.934	0.897	0.915	0.893	0.921	0.866	0.893
		12 Merkez	CNN	0.900	0.922	0.879	0.900	0.874	0.909	0.841	0.874
			LSTM	0.900	0.923	0.879	0.900	0.874	0.909	0.841	0.874
+	6 Anahtar Kare	-	CNN	0.913	0.926	0.895	0.910	0.886	0.908	0.856	0.881
			LSTM	0.913	0.926	0.895	0.910	0.886	0.908	0.857	0.882
		6 Merkez	CNN	0.932	0.940	0.917	0.928	0.905	0.922	0.880	0.900
			LSTM	**0.932**	**0.940**	**0.917**	**0.928**	**0.905**	**0.922**	**0.880**	**0.901**
		9 Merkez	CNN	0.919	0.930	0.902	0.916	0.901	0.919	0.875	0.897
			LSTM	0.919	0.931	0.902	0.916	0.901	0.920	0.875	0.897
		12 Merkez	CNN	0.913	0.925	0.896	0.910	0.895	0.915	0.868	0.891
			LSTM	0.913	0.926	0.896	0.910	0.895	0.915	0.868	0.891

Çizelge 4.8: CK+ veri kümesi üzerinde farklı ön işleme adımları uygulanarak yapılan deneyler sonucu CNN-Dönüştürücü ağının performansı

Yöntem				6 Duygu				7 Duygu			
Video Çoklama	Anahtar Kare Seçimi	Gauss K. M	YSA	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru
-	6 Anahtar Kare	-	CNN	0.873	0.904	0.845	0.873	0.852	0.897	0.809	0.850
			Dönüştürücü	0.736	0.802	0.697	0.746	0.705	0.766	0.637	0.695
		6 Merkez	CNN	0.914	0.934	0.895	0.914	0.894	0.924	0.867	0.895
			Dönüştürücü	0.707	0.779	0.612	0.686	0.677	0.743	0.612	0.671
		9 Merkez	CNN	0.898	0.921	0.875	0.898	0.889	0.924	0.858	0.890
			Dönüştürücü	0.752	0.793	0.661	0.721	0.713	0.767	0.655	0.707
		12 Merkez	CNN	0.893	0.916	0.870	0.892	0.885	0.918	0.854	0.885
			Dönüştürücü	0.700	0.770	0.606	0.678	0.646	0.746	0.606	0.669
+	6 Anahtar Kare	-	CNN	0.907	0.923	0.892	0.907	0.883	0.912	0.857	0.883
			Dönüştürücü	**0.876**	**0.889**	**0.863**	**0.876**	**0.839**	**0.874**	**0.822**	**0.847**
		6 Merkez	CNN	0.922	0.936	0.909	0.922	0.900	0.925	0.878	0.901
			Dönüştürücü	0.842	0.853	0.819	0.835	0.830	0.856	0.806	0.830
		9 Merkez	CNN	0.910	0.926	0.896	0.911	0.903	0.924	0.883	0.903
			Dönüştürücü	0.836	0.858	0.819	0.838	0.820	0.847	0.796	0.821
		12 Merkez	CNN	0.908	0.924	0.893	0.908	0.892	0.918	0.869	0.893
			Dönüştürücü	0.844	0.863	0.828	0.845	0.806	0.828	0.791	0.809

4.2 Detaylı Analiz

Deneysel Sonuçlar bölümünde 4.1 bahsedilen sonuçlar, farklı yapay sinir ağı yapılarının, özniteliklerin, ön işleme adımlarının ve duygu sınıdlarının etkilerini açıkça göstermiştir. Araştırmamız tarafından uygulanan tüm modeller arasındaki karşılaştırmalar ve modelle rimiz ile diğer güncel yapıların karşılaştırmaları, bu modellerin her birinin avantajlarını ve sınırlamalarını gözlemlemeyi sağlamıştır. 2 farklı veri seti üzerinde yapılan deneyler de duygu sınıflandırma performanslarına ilişkin kullanılan veriye yönelik yorum yapabilmeye imkan sağlamıştır.

Deneyler sonucu elde edilen bulgular aşağıdaki gibidir;

- FABO veri seti ile yapılan deneylerde, 9 duygu sınıfı ve 6 duygu sınıfı için

Çizelge 4.9: Tez çalışmasında önerilen modelin güncel yüksel başarımlı yöntemlerle kıyaslanması.

Model	Veri Kümesi		Duygu Sayısı			Ön İşleme			Doğruluk
	FABO	CK+	6	7	9	Gauss K.M.	Video Çoklama	Anahtar Kare Seçimi	
3D-CNN+LSTM with Keyframes Selection [22]	+	-	-	-	-	-	-	16 Anahtar Kare	0.733
Multichannel CNN (MCCNN) [14]	+	-	-	-	-	-	-	-	0.913
CNN-SIFT [15]	-	+	-	-	-	-	-	-	0.941
Fusion of LEMHI-VGG, VGG-CTSLSTM [17]	-	+	-	-	-	-	-	-	0.939
Nested LSTM (STC-NLSTM) [19]	-	+	-	-	-	-	-	-	0.998
CNN + LSTM	+	-	-	-	+	-	+	30 Anahtar Kare	0.986
CNN + LSTM	+	-	+	-	-	6	+	30 Anahtar Kare	0.997
CNN + LSTM	+	-	+	-	-	9	+	30 Anahtar Kare	0.996
CNN + LSTM	+	-	+	-	-	12	+	30 Anahtar Kare	0.995
CNN + Dönüştürücü	+	-	-	-	+	6	+	30 Anahtar Kare	0.992
CNN + Dönüştürücü	+	-	-	-	+	9	+	30 Anahtar Kare	0.992
CNN + Dönüştürücü	+	-	-	-	+	12	+	30 Anahtar Kare	0.991
CNN + Dönüştürücü	+	-	+	-	-	9	+	30 Anahtar Kare	**0.999**
CNN + Dönüştürücü	+	-	+	-	-	12	+	30 Anahtar Kare	**0.999**
CNN + LSTM	-	+	-	+	-	6	+	10 Anahtar Kare	0.905
CNN + LSTM	-	+	-	+	-	9	+	10 Anahtar Kare	0.901
CNN + LSTM	-	+	+	-	-	6	+	10 Anahtar Kare	**0.932**
CNN + Dönüştürücü	-	+	+	-	-	-	+	10 Anahtar Kare	0.876
CNN + Dönüştürücü	-	+	+	-	-	6	+	10 Anahtar Kare	0.842
CNN + Dönüştürücü	-	+	-	+	-	-	+	10 Anahtar Kare	0.839

%99'lara ulaşan doğruluklar elde edildi. Ve doğruluğa çok yakın Kesinlik, Duyarlı lık ve F1-Skoru değerleri görüldü. Bu da model doğruluklarında göz ardı edilen yanlış pozitif veya yanlış negatif hataların olmadığını göstermiş oldu. Yüksek doğruluk oranları, Dönüştürücü kullanılan yapılarda, video çoklama ve 30 anahtar kareli ön işleme adımlarıyla elde edildi. LSTM kullanılan yapılar birkaç istisna dışında aynı şartlar saplandığında Dönüştürücü kullanılan modellerin gerisinde kaldı. Bu da çok imleçli ilgi mekanizmalı dönüştürücü yapısının avantajını göstermiş oldu. İlgi mekanizması ile geçmiş verinin unutulmasını önlemek, uzun sekanslarda LSTM'den daha yüksek başarım elde edilmesini sağladı.

- Bunun yanında Anahtar Kare Seçimi yaparken seçilen kare miktarı düşürüldüğün de (10 anahtar kare seçilen deneylerde), Dönüştürücü performansının düştüğü gözlemlendi. LSTM kullanolan yapılarda da performans düşüşü gözlemlenmiş olsa da fark daha azdı. Bunun sebebinin sekans uzunluğunun kısalması olduğu düşünül mekte. Bunun yanında video çoklama işlemi ile veri miktarını artırmanın model performanslarını da iyi yönde etkilediği görüldü.

- Öte yandan, CK+ veri seti ile en yüksek doğruluk, Dönüştürücü modelinden yaklaşık %3 ila %20 arasında daha iyi değerlerle LSTM yapıları tarafından elde

edildi. Bunun sebebi CK+ veri kümesinin hem sadece yüz tanımlayıcılardan oluşması hem de sekans uzunluğunun çok kısa oluşudur, Her bir video sadece 6 kare uzunluğundadır. FABO kümesiyle yapılan deneylere bakıldığında varılan bu sonucun tutarlı olduğu görülmüştür. Bu veri kümesinde de video çoklama aynı şekilde model performanslarını pozitif yönde etkilemiştir. Veri kümesinin kısıtlarından dolayı tüm videolardan 6 kare seçilmesi gerektiğinden, ön işlemeden geçmemiş veri üzerinde deney yapma imkanı kalmadı.

- Sınıflandırmada etkili olan bir diğer faktör de Gauss merkezlerinin eklenmesiydi. Performans skorları farklı sayıda Gauss merkezi ile yapılan deneylerden çok etkilenmemiş olsa da, çoğu durumda ek özellikler olarak daha düşük Gauss merkezi sayısıyla hem LSTM hem de Dönüştürücü'de FABO veri kümesi için daha iyi performans görüldü. CK+ veri kümesiyle yapılan deneylerde de benzer sonuçlar alındı. Ancak duygu sınıflandırma başarımı, hem FABO hem de CK+ veri kümesinde elde edilen en yüksek doğruluklar ele alındığında, Gauss merkezle rini kullanılmadığında daha iyiydi.

- Her iki veri kümesinin sonuçlarından, Dönüştürücünün, büyük özellik setleri (yani, hem yüz hem de vücut özellikleri) ve daha büyük uzunlukta videolada LSTM'den daha iyi performans sergilediği sonucuna kolayca varılabilir. CK+ veri seti için çoğu durumda LSTM ve Transformer modelleri arasındaki performans ölçüleri çok benzer ve karşılaştırılabilir olduğundan, İlgi Mekanizmalı Dönüştürü cünün önerilen ön işleme adımları, Gauss Karışım Merkezleri ve CNN ön katmanıy la görüntülerdeki duyguları sınıflandırma işlemini açıkça iyileştiği söylenebilir.

- Yüz ve Vücut tanımlayıcıları iki farklı veri kümesine ayırarak yapılan deneyler özniteliklerin başarıma etikisi hakkında analiz yapabilmeye olanak sağladı. Birleşik veri kümesi üzerinde hem CNN-LSTM hem de CNN-Dönüştürücü yapısı %99'lara varan doğruluk oranlarına ulaşabildi ancak aynı performans ayrık veri kümeleri ile eğitim yapıldığında gözlemlenemedi. Video çoklama ve 10 kareli anahtar kare seçimi ile, Gauss merkezlerinin etkisi çok olmasa da, ayrık modeller daha iyi performans sergiledi. Genel olarak vücut tanımlayıcılarla duygu sınıflandırma başarımı, yüz tanımlayıcılara göre daha yüksekti. Dönüştürücü kullanılan yapılarda

da benzer durumlar gözlemlendi.

Tez çalışmasında önerilen ön işleme adımları ve derin öğrenme yapılarını farklı kombi nasyonlar denenerek birbirleriyle kıyaslama ve analiz etmenin yanında, elde edilen sonuçlar aynı problem üzerinde çalışan güncel ve yüksek başarımlı diğer yöntemlerle de karşılaştırıldı. FABO veri kümesi ile eğitilmiş CNN-Dönüştürücü yapısı, aynı veri kümesini kullanan en gelişmiş yöntemlerden daha yüksek performans sergiledi. CK+ veri kümesinde daha önce açıklandığı üzere sekans kısalığından dolayı FABO'ya göre daha düşük performans elde edilmiş olsa da tez çalışmasında önerilen iki katmanlı yapı mevcut yapılarla karşılaştırılabilir, çok da düşük olmayan doğruluklar elde etmiştir.

5. DEĞERLENDİRME

Herhangi bir kaynaktan otomatik duygu tespiti, insan duygularının ve ifadelerinin çok yönlülüğü nedeniyle zorlu bir iştir. Son yıllarda bu probleme daha çok Derin Sinir Ağları ile yaklaşılmaktadır. Bu tez çalışmasında, videolardan veya görüntü dizilerinden duyguları sınıflandırmak için bir dizi ön işleme adımı ile birlikte çok katmanlı bir DNN yapısını Gauss Karışım Merkezleri ile birleştirerek yeni bir yapı önerdik. İki popüler veri kümesinden, FABO ve CK+, faydalanılarak, video karelerinden yüz ve vücut özellikleri elde edildi, ardından bunları video çoklama ve anahtar kare seçimi adımlarına tabi tutuldu. Ön işleme adımından elde edilen veriler ile CNN modeli beslendi ve bu modelin çıktısı LSTM ve Çok imleçli İlgi Mekanizmalı Dönüştürücü için ayrı ayrı girdi olarak kullanıldı. Önerilen DNN yapısının performansını iyileştirmek için bu aşamada farklı merkezlere sahip Gauss Karışım Modelleri de çalışmaya dahil edildi. Yapılan deneyler sonucunda Dönüştürücü kullanılan yapıların çoğu senaryoda diğer yapılardan daha iyi performans sergilerdiği gözlemlendi. Önerilen modelin, yapısıyla tutarlı olarak, fazla öznitelik sayısı ve uzun sekanslarda daha başarılı olduğu görüldü. Tez çalışmasında önerilen yöntem ve modellerin farklı kombinasyonlarının karşılaştırma lı analizi ve bunların benzer mevcut duygu sınıflandırma modellerle karşılaştırılması, iki katmanlı yapı, ilgi mekanizmalı dönüştücü, jest ve mimiklerden elde edilen tanımlayıcı lar, gauss karışım merkezleri, video çoklama ve anahtar kare seçimi yöntemlerinin ayrıntılı bir biçimde incelenmesine ve bunların duygu sınıflandırma görevi üzerindeki bireysel ve birleşik etkilerinin iç yüzünü anlamaya olanak sağladı. Bu karşılaştırma ayrıca LSTM ve Dönüştücü modellerinin yapılarını ve farklarını görselleştirme ve inceleme fırsatı sundu.

5.1 Kısıtlar ve Gelecek Çalışmalar

Önerilen CNN-Dönüştürcü modeli, bir çok modelden daha iyi performans göstermesine rağmen, sistemde birkaç sınırlama vardır. Kullanılan iki veri kümesinde de sınıf dağılımla rı dengesizdi, 20 ile 83 arasında değişmekteydi. Dengeli bir veri kümesi ile daha yüksek performans elde etmek mümkün olabilir. Ayrıca, CK+ veri setinde videoların

uzunlukları farklılık göstermekteydi, bunu aşmak için de tüm videolar en kısa videonun uzunluğuna eşitlenmek üzere 6 kareye indirildi. Sonuç olarak büyük miktarda veri kaybı yaşandı. Bunun yanında sonuç olarak, FABO kümesi üzerinde yapılan deneylerde görüldüğü üzere kısa sekans uzunluklarında performansı düşen Dönüştürücü modeli, CK+ kümesinde LSTM kadar başarılı olamadı. Bunlara ek olarak gelecek çalışma olarak OpenPose aracını önerilen sınıflandırıcı yapısına entegre edip gerçek zamanlı bir duygu sınıflandırma aracı planlanan işler arasındadır. Bunun yanında veri kanalı genişletilerek konuşma içeren videolar üzerinde çalışılıp hem görüntü hem yazıdan duygu analizi yapan çok kanallı ve daha geniş kapsamlı bir yapı yine planlanmaktadır.

KAYNAKLAR

[1] **Darwin**, **C.** (2015). The expression of the emotions in man and animals. University of Chicago press.

[2] **Bota**, **P. J.** et al. (2019). A review and current challenges and future possibilities on emotion recognition using machine learning and physiological signals. In: *IEEE Access* 7, pp. 140990–141020.

[3] **Chakraborty**, **B. K.** et al. (2018). Review of constraints on vision-based gesture recognition for human–computer interaction. In: *IET Computer Vision* 12.1, pp. 3–15.

[4] **Poria**, **S.** et al. (2019). Emotion recognition in conversation: Research challenges and datasets, and recent advances. In: *IEEE Access* 7, pp. 100943–100953.

[5] **Santamaria-Granados**, **L.**, **Mendoza-Moreno**, **J. F.**, and **Ramirez-Gonzalez**, **G.** (2021). Tourist Recommender Systems Based on Emotion Recognition—A Scientometric Review. In: *Future Internet* 13.1, p. 2.

[6] **Islam**, **M. R.** et al. (2018). Depression detection from social network data using machine learning techniques. In: *Health information science and systems* 6.1, pp. 1–12.

[7] **Ko**, **B. C.** (2018). A brief review of facial emotion recognition based on visual information. In: *Sensors* 18.2, p. 401.

[8] **Cao**, **Z.** et al. (2019). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE*.

[9] **Özyer**, **T.**, **Ak**, **D. S.**, and **Alhajj**, **R.** (2021). Human action recognition approaches with video datasets - A survey. In: *Knowledge Based Systems* 222, p. 106995.

[10] — (2020). Recent Trends in Emotion Analysis: A Big Data Analysis Perspective. In: *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2020, The Hague, Netherlands, December 7-10, 2020*. IEEE, pp. 710–714.

[11] **Nonis** et al. (2019). 3D approaches and challenges in facial expression recognition algorithms—A literature review. In: *Applied Sciences* 9.18, p. 3904.

[12] **Li**, **S.** and **Deng**, **W.** (2020). Deep facial expression recognition: A survey. In: *IEEE Transactions on Affective Computing*.

[13] **Mungra**, **D.** et al. (2020). PRATIT: a CNN-based emotion recognition system using histogram equalization and data augmentation. In: *Multimedia Tools and Applications* 79.3, pp. 2285–2307.

[14] **Barros**, **P.** et al. (2015). Multimodal emotional state recognition using sequence-dependent deep hierarchical features. In: *Neural Networks* 72, pp. 140–151.

[15] **Sun**, **X.** and **Lv**, **M.** (2019). Facial expression recognition based on a hybrid model combining deep and shallow features. In: *Cognitive Computation* 11.4, pp. 587–597.

[16] **Agrawal**, **A.** and **Mittal**, **N.** (2020). Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy. In: *The Visual Computer* 36.2, pp. 405–412.

[17] **Hu**, **M.** et al. (2019). Video facial emotion recognition based on local enhanced motion history image and CNN-CTSLSTM networks. In: *Journal of Visual Communication and Image Representation, Elsevier* 59, pp. 176–185.

[18] **Abdullah**, **M.**, **Ahmad**, **M.**, and **Han**, **D.** (2020). Facial Expression Recognition in Videos: An CNN-LSTM based Model for Video Classification. In: *2020 International Conference on Electronics, Information and Communication (ICEIC), IEEE*, pp. 1–3.

[19] **Yu**, **Z.** et al. (2018). Spatio-temporal convolutional features with nested LSTM for facial expression recognition. In: *Neurocomputing* 317, pp. 50–57.

[20] **Sapiński**, **T.** et al. (2019). Emotion recognition from skeletal movements. In: *Entropy* 21.7, p. 646.

[21] **Wang**, **S.** et al. (2020). Dance Emotion Recognition Based on Laban Motion Analysis Using Convolutional Neural Network and Long Short-Term Memory. In: *IEEE Access* 8, pp. 124928–124938.

[22] **Ly**, **S. T.** et al. (2019). Gesture-Based Emotion Recognition by 3D-CNN and LSTM with Keyframes Selection. In: *International Journal of Contents* 15.4, pp. 59–64.

[23] **McCulloch**, **W. S.** and **Pitts**, **W.** (Dec. 1943). A logical calculus of the ideas immanent in nervous activity. In: *The bulletin of mathematical biophysics* 5.4, pp. 115–133.

[24] **Fukushima**, **K.** (Apr. 1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. In: *Biological Cybernetics* 36.4, pp. 193–202.

[25] **Hochreiter**, **S.** and **Schmidhuber**, **J.** (Nov. 1997). Long Short-Term Memory. In: *Neural Comput.* 9.8, pp. 1735–1780.

[26] **Vaswani**, **A.** et al. (2017). Attention is all you need. In: *Advances in neural information processing systems*, pp. 5998–6008.

[27] **Devlin**, **J.** et al. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv: 1810.04805.

[28] **Yang**, **Z.** et al. (2019). *XLNet: Generalized Autoregressive Pretraining for Language Understanding*. arXiv: 1906.08237.

[29] **Brown**, **T. B.** et al. (2020). *Language Models are Few-Shot Learners*. arXiv: 2005.14165.

[30] **Gunes**, **H.** and **Piccardi**, **M.** (2006). A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. In: vol. 1, pp. 1148–1153.

[31] **Lucey**, **P.** et al. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: pp. 94–101.

[32] **Simon**, **T.** et al. (2017). Hand Keypoint Detection in Single Images using Multiview Bootstrapping. In: *CVPR*.

[33] **Cao**, **Z.** et al. (2017). Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. In: *CVPR*.

[34] **Wei**, **S.-E.** et al. (2016). Convolutional pose machines. In: *CVPR*.

[35] **Hidalgo**, **G.** et al. (n.d.). *OpenPose, The first real-time multi-person system to jointly detect human body, hand, facial, and foot keypoints*. URL: https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/md_doc_02_output.html.

[36] **admin**, **ialab** (2017). *Detecting human facial expression by common computer vision techniques*. URL: http://www.interactivearchitecture.org/detecting-human-facial-expression-by-common-computer-vision-techniques.html.

[37] **mlhubber** (n.d.). *Colorize black and white photos*. URL: https://github.com/mlhubber/colorize.

[38] **Ekman**, **P.** (1992). An argument for basic emotions. In: *Cognition and emotion* 6.3-4, pp. 169–200.