T.C.

İSTANBUL KÜLTÜR ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YÜZ İFADESİ TANIMADA ÖNYARGISIZ VE AYRIŞTIRILMIŞ TEMSİLLERİN ÖĞRENİLMESİ

YÜKSEK LİSANS TEZİ

Büşra KOCAÇINAR

1600003900

Anabilim Dalı: Bilgisayar Mühendisliği

Program: Bilgisayar Mühendisliği

Tez Danışmanı: Doç. Dr. Fatma PATLAR AKBULUT

HAZİRAN 2024

T.C.

İSTANBUL KÜLTÜR ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YÜZ İFADESİ TANIMADA ÖNYARGISIZ VE AYRIŞTIRILMIŞ TEMSİLLERİN ÖĞRENİLMESİ

YÜKSEK LİSANS TEZİ

Büşra KOCAÇINAR

1600003900

Anabilim Dalı: Bilgisayar Mühendisliği

Program: Bilgisayar Mühendisliği

Tez Danışmanı: Doç. Dr. Fatma PATLAR AKBULUT

Jüri Üyeleri: Doç. Dr. Akhan AKBULUT

Doç. Dr. Can EYÜPOĞLU

HAZİRAN 2024

ÖNSÖZ

Bu tez çalışmasında yüz ifade tespit sistemlerindeki önyargıyı azaltma ve daha adil modeller elde etme amacıyla ayrıştırılmış temsil yöntemiyle seçilen özelliklerin ayrıştırılma performansları incelenmiştir.

Çalışmanın gerçekleştirilmesinde bilgi ve tecrübeleriyle yol gösteren, bana rehberlik eden değerli tez danışmanım Doç. Dr. Fatma Patlar Akbulut'a desteği için içten ve sonsuz teşekkürlerimi sunarım. Tüm eğitim hayatım boyunca bana her konuda destek olan başta babam olmak üzere sevgili aileme ve bana olan güvenlerini hiç kaybetmeyen arkadaşlarıma teşekkür ederim.

İçindekiler

KISALTMALAR iii

ŞEKİLLER LİSTESİ iv

TABLOLAR LİSTESİ v

KISA ÖZET vi

ABSTRACT vii

1. GİRİŞ 1

2. ÖN BİLGİLER ve LİTERATÜR ARAŞTIRMASI 4

2.1. Duygu Tanıma Çalışmaları 4

2.2. Ayrıştırılmış Temsil Öğrenimi 6

a) Değişimsel oto kodlayıcılar (VAE) ve Beta-VAE 7

b) Çekişmeli Üretici Ağlar (Generative Adversarial Networks – GAN) 8

2.3. Duygu Tanıma Ayrıştırılmış Temsil ve Önyargı 9

3. GEREÇ VE YÖNTEMLER 14

3.1. Araştırma Tasarımı 14

3.2. Veri Seti ve Ön İşleme 15

3.3. Ayrıştırılmış Temsil Öğrenimi 18

4. DENEYLER 19

4.1. Kullanılan Veri Seti 19

4.2. Ayrıştırılmış Temsil Öğrenmesi ile Önyargısız Duygu Tanıma Sistemi 21

4.2.1. Araştırma sorusu 1: Görünüm Faktörlerinin Etkisi 21

4.2.2. Araştırma sorusu 2: Ayrıştırılmış (Disentangled) Temsil Yönteminin Etkisi 22

4.2.3. Araştırma sorusu 3: Algoritma Seçimi ve Optimizasyonu 28

5. SONUÇLAR 38

6. TARTIŞMA 41

7. ÇALIŞMANIN SINIRLARI VE GELECEK PLANI 43

8. KAYNAKLAR 45

KISALTMALAR

FER: Facial Expression Recognition - Yüz İfade Tanıma

VAE: Variational AutoEncoder - Varyasyonel Otokodlayıcı

CNN: Convolutional Neural Network - Evrişimsel Sinir Ağı

RNN: Recurrent Neural Network - Yinelemeli Sinir Ağı

KL Uzaklığı: Kullback - Leibler Uzaklığı

LSTM: Long Short-Term Memory - Uzun-Kısa Süreli Bellek

GAN: Generative Adversarial Network - Çekişmeli Üretici Ağ

ŞEKİLLER LİSTESİ

Şekil 1. CelebA Veri Seti Örnek Görüntüler 15
Şekil 2. AffectNet Veri Seti Örnek Görüntüler 16
Şekil 3. Çalışma [6]'da kullanılan mimari tasarım örneği.....18
Şekil 4. Temel Sınıflandırıcı ResNet-18 Mimarisi 21
Şekil 5. Beta-VAE Ayrıştırılmış Temsil Model Mimarisi 23
Şekil 6. Kodlayıcı Mimarisi 23
Şekil 7. Çözücü Mimarisi 24
Şekil 8. Veri setinden rastgele seçilen bir görüntünün kodlayıcı çıktısının görselleştirilmesi..25
Şekil 9. Sarhan ve Arkadaşları tarafından önerilen yapı [33] 26
Şekil 10. 64x64 Boyutlu CelebA Veri Setinin Beta Değerlerine Göre Beta-VAE Çıktıları 29
Şekil 11. 128x128 Boyutlu CelebA Veri Setinin Beta Değerlerine Göre Beta-VAE Çıktıları..31
Şekil 12.64x64 Boyutlu AffectNet Veri Setinin Beta Değerlerine Göre Beta-VAE Çıktıları...34
Şekil 13. 128x128 Boyutlu AffectNet Veri Setinin Beta Değerlerine Göre Beta-VAE Çıktıları
.....36

TABLOLAR LİSTESİ

Tablo 1. Ayrıştırılmış Temsil Öğrenimi Literatür Karşılaştırması....12
Tablo 2. CelebA Veri Seti Hassas Özellik Dağılımları....15
Tablo 3. CelebA Veri Seti Etiket Örnekleri....16
Tablo 4. AffectNET Veriseti Hassas Özellik Dağılımları....17
Tablo 5. Alt Veri Kümeleri Gülümseme- Hassas Etiket Dağılımı....19
Tablo 6.CelebA Belirlenen Alt Veri Setlerinde Bulunan Diğer Hassas Özelliklerin Dağılımları 20
Tablo 7. AffectNet Dataseti Duygulara Göre Hassas Özellik Dağılımları 20
Tablo 8. Temel Sınıflandırıcı Hassas Özellik Test Sonuçları....22
Tablo 9. B-VAE ile Eğitilen Duygu Tanıma Model Sonuçları....24
Tablo 10. Ayrıştırılmış Temsil Yöntemi ile Duygu Tanıma Eğitim Sonuçları....27
Tablo 11. 64x64 Boyutlu CelebA Veri Setinin Beta Değerlerine Göre Eğitim Sonuçları....28
Tablo 12.128x128 Boyutlu CelebA Veri Setinin Beta Değerlerine Göre Eğitim Sonuçları....30
Tablo 13. CelebA Veri Setinin Beta Değerlerine ve Hassas Özelliklere Göre Test Doğrulukları 32

GENEL BİLGİLER

Enstitü : Lisansüstü Eğitim Enstitüsü

Anabilim Dalı : Bilgisayar Mühendisliği

Program : Bilgisayar Mühendisliği

Tez Danışmanı : Doç. Dr. Fatma PATLAR AKBULUT

Tez Türü ve Tarihi : Yüksek Lisans - Haziran 2023

KISA ÖZET

YÜZ İFADESİ TANIMADA ÖNYARGISIZ VE AYRIŞTIRILMIŞ TEMSİLLERİN ÖĞRENİLMESİ

Sözsüz iletişimin bir parçası olan yüz ifadelerinin tanınmasına yönelik geliştirilen sistemler incelendiğinde yüzde bulunan kişisel özelliklerin ve aksesuarların duygu tanıma sistemlerini olumsuz etkilediği tespit edilmektedir. Bu problemin ortadan kaldırılmasına yönelik literatürde kullanılan yöntemler incelendiğinde ayrıştırılmış temsil öğrenimi öne çıkmaktadır. Hassas değişkenlerin, hedef temsilinden ayrıştırılması değişimsel oto kodlayıcılar ile mümkün olmaktadır. Çalışmada bir temel sınıflandırıcı, bir beta-vae kullanılarak eğitilen bir sınıflandırıcı, son olarak da temsillerin doğru bir şekilde ayrıştığını kontrol edebilmek için ortogonalite kaybı kullanan bir mimari ile yüz ifadesi sınıflandırma modelleri eğitilmiştir. Hassas olarak seçilen özellikler yüz ve boyun bölgesinde bulunan kişisel özellikler ve aksesuarlar olarak belirlenmiştir. Kullanılan veri setine göre seçilen duygu gülümseme üzerinden değerlendirilmektedir. Elde edilen model sonuçları temel sınıflandırıcı modeli ile karşılaştırılmıştır. Belirlenen hassas özellikler için ayrıştırılmış temsil yönteminin doğruluğu olumlu yönde etkilediği gözlemlenmiştir. Bunun yanı sıra hassas özelliklerin hedef temsilden çıkarılması ile daha adil bir sınıflandırıcı elde edilmiştir.

Anahtar Kelimeler: Ayrıştırılmış Temsil Öğrenimi, B-VAE, Duygu Tanıma Sistemleri, Duygu Tanıma Sistemlerinde Önyargı

GENERAL INFORMATION

University : İstanbul Kültür University

Institute : Institute of Graduate Studies

Department : Computer Engineering

Programme : Computer Engineering

Supervisor : Assoc. Prof. Dr. Fatma PATLAR AKBULUT

Degre Awarded and Date : Master's degree – June 2024

ABSTRACT

LEARNING DEBIAS AND DISENTANGLED REPRESENTATION IN FACIAL EXPRESSION RECOGNITON

When the systems developed for the recognition of facial expressions, which are a part of non-verbal communication, are examined, it is determined that characteristics features and accessories on the face negatively affect emotion recognition systems. When the methods used in the literature to eliminate this problem are examined, disentangled representation learning stands out. Separating sensitive variables from the target representation is possible with variational autoencoders. In the study, facial expression classification models were trained with a base classifier, a classifier trained using beta-vae, and finally an architecture that uses orthogonality loss to check that the representations are correctly disentangled. The features selected as sensitive features were determined as characteristic features and accessories located in the face and neck area. Depending on the dataset used, the selected emotion is evaluated through smile. The obtained model results were compared with the base classifier model. It has been observed that the disentangled representation method has a positive impact on accuracy for the identified sensitive features. In addition, a fairer classifier was obtained by removing sensitive features from the target representation.

Keywords: Disentangled Representation Learning, B-VAE, Facial Expression Recognition, Bias in Facial Expression Recognition System

1. GİRİŞ

Yüz ifadeleri, sözlü ifadeleri tamamlarken sözsüz etkileşimlerde de güçlü ipuçları vererek, insan iletişimini doğrudan etkiler. Bireyler arasında yüz ifadelerinin yoğunluğu farklılık gösterse de farklı kültürlerde gözlemlenebilen evrensel olarak tanımlanabilir ortak duygusal ifadeler vardır. Bu ifadeler, bireyin içsel duygusal durumunu ifade etme ve sosyal anlayış ile empatiyi geliştirme gibi çeşitli amaçlara hizmet eder. Yüz kaslarındaki incelikli değişiklikler sayesinde bireyler, mutluluk, coşku, üzüntü ve endişeyi kapsayan çok çeşitli duyguları ifade etme yeteneğine sahip olur ki bu da karşısındaki ile bağ kurmayı ve etkileşimi kolaylaştırır. Mevcut araştırmalar [1], insanların yüz ifadelerini yorumlama konusunda dikkate değer bir yeteneğe sahip olduklarını, dilsel ve kültürel engellerle karşılaştıklarında dahi çoğunlukla içgüdüsel ve son derece doğru bir anlama sergilediklerini gösteriyor. Yüz ifadeleri, kişilerin duygularını ve niyetlerini kelimelere gerek duymadan aktarabilmelerini sağladığı için kişilerarası iletişimde oldukça önemlidir. Bu sebeple yüz ifadesi tanıma güvenlikten, eğitime, sağlıktan eğlenceye birçok alana entegre edilmiştir ve edilmeye de devam etmektedir. Bununla birlikte, yüz ifadesi tanıma sistemlerinin etkinliği, özellikle çeşitli insan görünümleri ve bunların içindeki nüanslar ile karşı karşıya kaldığında, bazı zorluklardan da uzak değildir.

Kat edilen tüm ilerlemelere rağmen eğitim için kullanılan veri kümelerinin doğasında bulunan önyargılar modellerin başarımını önemli ölçüde etkilemektedir. Saç modeli, aksesuarlar, makyaj ve görünümle ilgili diğer özellikler gibi faktörler, tanınma modellerinin performansını ve adilliğini önemli ölçüde etkileyebilecek karmaşıklıkları ortaya çıkarır. Örneğin, çeşitli saç stillerinin veya aksesuarların varlığı, duyguların tespiti için önem taşıyan belirli yüz özelliklerini potansiyel olarak gizleyebilir ve bu da yanlış yorumlamalara veya yanlış sınıflandırmalara yol açabilir. Benzer şekilde, makyaj seçimi yüzün görsel sunumu üzerinde bir etkiye sahip olabilir; bu da modelin gerçek duygusal ifadeleri yüzeysel değişikliklerden ayırt etme kapasitesini karmaşıklaştırabilir. Ayrıca, bu özelliklerin sonuçları salt teknik engelleri aşarak

önemli etik tartışmalara yol açmaktadır. Yüz ifadesi tanıma modelleri, eşitsizliklerin devam etmesine ve toplumsal stereotiplerin güçlenmesine potansiyel olarak katkıda bulunabilecek doğal önyargılara sahiptir. Örneğin, belirli saç stilleri veya aksesuarlar belirli demografik gruplarla eşit olmayan bir şekilde bağlantılıysa, model istemeden bu standartlara uymayan kişilere karşı önyargı gösterebilir. Günümüzde duygu tanıma çalışmaları fizyolojik sinyaller, videolar, ses kayıtları ve yüz fotoğrafları gibi çeşitli veriler ile yapılmaktadır. Video ve görsel veri ile yapılan duygu tanıma çalışmalarda bireylerin üst vücut bölgeleri ve yalnızca yüz ifadeleri ile araştırmalar yapılır. Bu durumlarda odaklanılan özellikler mevcut olan doğal önyargılardan etkilenebilir. Yüz görselleri ele alınarak oluşturulan derin öğrenme modelleri ile yapılan duygu tanıma çalışmaları kişinin ten rengi, yaş ve cinsiyet dahil ancak bunlarla sınırlı olmamak üzere, bu sistemlere hatalar ve önyargılar getirme potansiyeline sahip olan doğal bireysel özelliklerin etkisine karşı hassastır. Yüz ve duygu tanıma sistemlerinin doğruluğu çeşitli çalışmalarda [2] kapsamlı bir şekilde incelenmiş ve demografik faktörlerin önemli etkisi ortaya çıkarılmıştır. Bu sorunun tespitiyle birlikte önyargılardan bağımsız daha adil sistemler oluşturmak için yeni yöntemler üzerinde çalışmalar halen devam etmektedir. Bu yöntemlerde hassas özellik olarak nitelendirilen önyargı ve adaletsizliğe sebep olan bu özelliklerin tanıma ve tespit sistemlerine dahil edilmeden bir model oluşturulması amaçlanmaktadır. Bu özelliklerin ayrıştırılması konusunda kullanılan yöntemlerden biri de ayrıştırılmış temsil yöntemidir. Bu yöntemde her bir özellik farklı bir vektör tarafından temsil edilmektedir. Bu sayede model eğitiminde kullanılacak özellikler ayrıştırılarak önyargı sorunu çözülmeye çalışılmaktadır.

Bu çalışmanın amacı kişinin saç şeklinin, kullandığı gözlük, kolye, küpe vb. aksesuarların, makyajlı olmasının duygu tanıma sistemleri üzerindeki etkisinin araştırılması ve bu hassas özelliklerin çıkarılarak önyargısız bir model oluşturulmasıdır. Bunu yapabilmek için ayrıştırılmış temsil yöntemi ile öncelikle yüz özelliklerinin doğru bir şekilde ayrıştırılması yapılmıştır. Çalışma sonuçlarının karşılaştırılabilmesi için öncelikle bir VAE modeli eğitilmiş ve sınıflandırıcı ağ ile

duygular tespit edilmiştir. Bu model belirlediğimiz hassas özelliklere karşı özelleştirilmiş bir hassasiyet taşımamaktadır. Verilen görsellerde VAE modeli için herhangi bir etiket ayrımı yapılmamıştır. Daha sonra hassas özelliklerin eğitilen modele sızmadığının tespiti sağlanmıştır. Bu da ortoganilite ile kontrol edilmektedir. Hassas özellik olarak belirlediğimiz etiketlerin sınıflandırıcı modeline dahil edilmediğinden emin olunarak eğitim gerçekleştirilmiştir. Ayrıca kullanılan ayrıştırılmış temsil yönteminin performans karşılaştırmasını yapabilmek için temel bir ağ eğitilerek duygu tanıma modeli geliştirilmiştir. Bu modellerin karşılaştırmaları olarak sonuçları tartışılmıştır. Tamamlanan model ve çıktılar çeşitli güvenlik uygulamalarında, reklam, satış, hedef müşteri belirleme gibi alanlarda (marketing), çalışanların performansını ve stres durumunu ölçme, çalışma ortamını geliştirme gibi çalışmalarda kullanılmasının ölçümlerdeki doğruluğu artırabileceği gözlemlenmektedir.

2. ÖN BİLGİLER ve LİTERATÜR ARAŞTIRMASI

2.1. Duygu Tanıma Çalışmaları

Duygu tanıma; mimikler, vücut dili, fizyolojik sinyal, ses tonu gibi davranışlardan yola çıkarak anlamlandırılmaya çalışılan duygu durumlarını tespit etme olarak açıklanabilmektedir. Yapay zekâ alanında da aynı amaçla sistemler geliştirilmektedir. Literatürde yapılan duygu tespit çalışmalarında farklı veri kaynakları ve yöntemler kullanılmıştır. Duygu tespit sistemleri yüz görselleri, videolar, ses kayıtları [3,4], fizyolojik sinyaller [5,6,7,8,9] gibi birçok veri kullanılarak geliştirilmektedir.

Çalışmamız kapsamında görsel girdi alarak geliştirilen sistemlere odaklanılmıştır. Bu sistemlerin geliştirilmesinde makine öğrenmesi ve derin öğrenme algoritmalarının öne çıktığı görülmektedir. Görüntü işleme konusunda birçok çalışmada kullanılan evrişimli sinir ağları duygu tanıma sistemlerinde de öne çıkmaktadır. Singh ve Nasoz'un yaptığı çalışmada [10] FER2013 veri seti ile CNN mimarisi kullanılarak deneyler gerçekleştirilmiştir. Fotoğrafların çekildiği ortamın parametrelerinin doğrudan model başarısını etkilediği belirtildiği için ışık durumunun düzeltilmesi amacıyla bir çalışma yapılmış ve histogram eşitleme yöntemi ile düzeltme adımı eklenmiştir. Dlib ile kaşlar, gözler, burun, ağzın iç ve dış kısımları ve çene olmak üzere yüz çıkarımı yapılmıştır. 6 evrişimli katman içeren ve max pooling, drop out uygulanan CNN modeli özellik çıkarımı tekniği kullanılmadan ve hiçbir ön işleme adımı olmadan %67 test doğruluğu elde edildiği açıklanmıştır.

Çalışmalarda kullanılan veri setlerindeki görsellerin benzerlikleri, yüz ifadesi tanıma sistemlerinde özelliklere odaklanılmasını zorlaştırmaktadır. Bu nedenle özellik çıkarma adımlarının eklenmesinin model doğrulukları üzerinde olumlu bir etkisi olmaktadır [11]. Li ve arkadaşları [12] ise CNN ağı ile bir özellik çıkarma yöntemi kullanarak benzer deneyleri gerçekleştirmiştir. Özellik çıkarma tekniği olarak yerel ikili desen olarak da bilinen LBP (Local Binary Patterns) merkezi bir piksel değeri belirleyip komşu piksellerle karşılaştırma yaparak ilgili piksel için bir dizi oluşturduktan sonra bir sayısal değer elde etmektedir. Bu çalışmada önerilen mimaride bir ham girdi bir de LBP özellik haritası olmak üzere iki evrişimli sinir ağından oluşan bir özellik çıkarma modülü tasarlanmıştır. Bu iki ağdan çıkan özellikler bir araya

getirilerek bir sonraki modülde birleştirilir. Özelliklerin ağırlıkları bunlara göre belirlenerek duygu tanıma modülünün odaklanacağı özellikler öne çıkarılmaya çalışılmıştır. FER2013, CK+, Oulu-CASIA ve JAFFE gibi farklı veri setleri ile yapılan karşılaştırmalarda önerilen bu mimarinin diğer yöntemlere göre daha yüksek tanıma oranı elde ettiği belirtilmiştir. Bununla birlikte kullanılan LBP ve dikkat mekanizmasının doğruluk oranı üzerinde olumlu etkileri olduğu, bu modüllerin birlikte kullanılmasının ifade tanımada etkili olduğu belirtilmiştir.

Bu konu özelinde hazırlanan ya da yüz görsellerini içeren veri setleri haricinde doğrudan yeni bir kümeye yönelik çalışmalar da yapılmaktadır. Wu'nun [13] yaptığı çalışmada bu yeni hedef bir sınıf ortamı olarak belirlenmiştir. Ancak modelin eğitim aşamasında Fer2013 veri seti kullanılmıştır. Çalışmanın yapıldığı dönemde hedef ortama yönelik yayınlanan ve genel erişime açık veri seti bulunamadığından dolayı üniversitede sınıf ortamında veri toplanarak modelin validasyon aşamasında kullanılmıştır. Çalışma Resnet50 modeli ve dikkat mekanizması olan Squeeze ve Excitation (SE) modülünden oluşmaktadır. SE modülü konvolüsyonel katmanın çıktısını alarak sıkıştırır ve uyarılma aşamasında da özellik haritası ağırlıklı olarak işler.

Yalnızca evrişimli sinir ağlarının kullanılmasının yanı sıra CNN ve RNN algoritmaları birlikte kullanılarak da duygu tanıma sistemleri geliştirilmektedir. Abdullah ve arkadaşları da [14] uzamsal özellik tespiti için CNN kullanırken, bu uzamsal özelliklerden zamansal özellikleri öğrenmek için de tek katmanlı bir LSTM ağı kullanılan bir model önermektedir. Eğitim aşamasında 3923, test ve validasyon için sırasıyla 491 ve 490 video kullanılmıştır. Bu videolar yedi sınıflı duygu etiketi içeren RAVDESS veri setinden alınmıştır. Çalışmanın yapıldığı zamana kadar bu veri seti üzerinde sadece video karelerini kullanarak bir sınıflandırma çalışması yapılmadığı, bu sebeple de yapılan bu çalışmanın RAVDESS veri seti üzerinde yapılan ilk karşılaştırma çalışması olarak kabul edilebileceği belirtilmiştir. Eğitim sonrasında özellik çıkarımı için kullanılan Xception Net, FER2013 veri seti için %80 validasyon

doğruluğu elde etmiştir. Tek katmanlı LSTM ağı ise %65 validasyon doğruluğuna ulaşmıştır.

Zhang ve arkadaşları [15], 31 Türk katılımcıdan toplanan altı duygu ifadesi ve dört zihinsel durum bilgisi içeren BAUM-1s, 8 kişilik 720 video ve altı duygu ifadesi içeren RML veri seti ve 2894 video içeren MMI veri seti kullanılarak bir sistem geliştirmiştir. Hibrit bir derin öğrenme modeli tasarlanan bu çalışmada veri segmentlerini işleyebilmek için CNN kullanılmaktadır. Daha sonra öğrenilen özellikler füzyon ağında birleştirilerek sistem tamamlanmıştır. Belirtilen üç veri seti ile yapılan testler sonucunda önerilen yöntemin avantajlı olduğu vurgulanmaktadır.

2.2. Ayrıştırılmış Temsil Öğrenimi

İnsanların herhangi bir nesneye baktığında gerçekleşen anlamlandırma sürecinin taklit edilmesi amaçlanan ayrıştırılmış temsil öğrenimi, verilerdeki değişkenler arasındaki ilişkilere odaklanarak, bu ilişkileri birbirinden bağımsız ve anlamlı olacak şekilde latent değişkenlere dönüştürerek açıklanabilen temsiller ortaya çıkarmaktır [16]. Elde edilen temsiller birçok farklı amaçla çeşitli çalışmalarda kullanılabilmektedir. Zhang ve arkadaşlarının [17], öneriler üzerine yaptığı çalışmada daha iyi bir öneri sistemi için ilgi ve sosyal faktörleri ayrıştırmak hedeflenmiştir. Başka bir çalışmada ise [18] metinler için ayrıştırılmış temsiller çıkarılmıştır. Bu sayede stil ve içerik özellikleri ayrılarak metin düzenleme işlemlerini yapabilmektedirler.

Yüz görselleri ele alındığında ayrıştırılmış temsil öğrenimi, yüz koordinatlarının öğrenilmesi [19], yüz ifadelerinin ayrıştırılması, yüz duruşunun öğrenilmesi gibi birçok uygulamada kullanılabilmektedir. Bu uygulamaların yanı sıra adil modeller için de ayrıştırılmış temsil kullanılabilmektedir. Adil ayrıştırılmış temsiller ise ırk, cinsiyet, yaş gibi hassas özelliklere yönelik önyargısız temsilleri ifade etmektedir. Ayrıştırılmış temsillerde, hassas özelliklere karşı dayanıklı, bu özellikleri kapsam dışı bırakmış modeller geliştirebilmek için adilliği dahil etmek önemli bir noktadır. Adil ayrıştırılmış temsiller, hassas özelliklerin karar verme sürecinde rol oynamadığından emin olmayı içermektedir [20].

a) Değişimsel oto kodlayıcılar (VAE) ve Beta-VAE

Oto kodlayıcı girdi alarak aldığı veriyi olabilecek en az kayıpla tekrar üretmeyi ve bu işlem sırasında veri ile ilgili önemli temsilleri öğrenmeyi hedefleyen gözetimsiz bir sinir ağıdır [21]. Kodlayıcı (encoder) ve kod çözücü(decoder) olmak üzere iki temel ağı bulunmaktadır. Kodlayıcı ve kod çözücü arasında darboğaz katmanı bulunmaktadır. Bu katman girdinin ne kadar fazla sıkıştırılacağını belirtmektedir.

Değişimsel oto kodlayıcılar ise kodlayıcı ve kod çözücü ağ arasında bulunan hidden space'deki vektör dağılımını normal dağılıma göre dağılmaya zorlayan gözetimsiz ve üretici bir oto kodlayıcı modelidir [22]. VAE girdi katmanındaki vektörü, gizli temsillerde ortalama ve standart sapmaya dönüştürmektedir. Bu iki parametre ile yeni örnekler üretebilmektedir. Oto kodlayıcıların kayıp fonksiyonu standart ortalama kare hatasıyken (Mean squared error (MSE)) değişimsel oto kodlayıcıların kayıp fonksiyonu ise standart ortalama kare hatası ve Kullback-Leibler ırsaklığıdır (KL-Divergence). İki olasılık dağılımı arasındaki farkı ölçen KL ıraksaklığı, normal dağılıma benzetmeye çalıştığımız gizli temsil ve standard-gaussianı dikkate almaktadır. KL- ıraksaklığı aşırı öğrenmeyi engelleyerek gizli temsillerde önemli özelliklerin tutulmasını sağlayan bir regularization görevi görmektedir.

Değişimsel oto kodlayıcılar (Variational Autoencoder, VAE) ve çekişmeli üretken ağlar (GAN) gibi modeller ile kullanılan ayrıştırılmış temsil öğrenimi, girdiye benzer bir veri üretmesini amaçlamanın yanı sıra varyasyonların altında yatan faktörleri de öğrenmeyi amaçlamaktadır.

Orijinal değişimsel oto kodlayıcılarda bulunan bir hiperparametre olan beta ile gizli uzaydaki öğrenmesini düzenleyen B-VAE modeli öğrenilen temsillerin daha iyi ayrıştırılmasını sağlamaktadır [23]. Yüksek beta değeri, modelin özellik ayrıştırmasında olumlu rol oynarken, daha düşük değerlerde yeniden yapılandırma hatasını minimize etmeye odaklanmasını sağlamaktadır. B parametresinin değeri 1 olarak belirlendiğinde orijinal VAE ile aynı işlevi gerçekleştirmektedir. Ayrıştırılmış temsil yöntemi için genellikle 1'den büyük değerler kullanılmaktadır [23]. Görüntü işleme [24- 26], duygu tanımada adil ve önyargısız model çalışmalarında [27,28] kullanılabilmektedir.

Denklem 1'de kayıp hesaplaması verilmektedir. Denklemin ilk kısmında yeniden oluşturma hatası hesaplanmaktadır. Bu değer girdi verisinin yeniden oluşturulma performansını göstermektedir. İkinci kısımda ise KL kaybı verilmektedir. Verilen B değeri 1'e eşit olduğu durumda VAE' ile aynı hesaplama gerçekleşmiş olmaktadır.

$$Kayıp_{\beta-VAE} = E_{q(z|x)}[logp(x \mid z)] - \beta \cdot D_{KL}(q(z \mid x) \parallel p(z)) \quad (1)$$

b) Çekişmeli Üretici Ağlar (Generative Adversarial Networks – GAN)
Çekişmeli Üretici Ağlar iki temel bileşenden oluşmaktadır: üretici ve ayırt edici ağlar [29]. Üretici ağ girdi olarak aldığı rastgele gürültüden gerçekçi veri örnekleri üretmeye çalışmaktadır. Bu sırada ayırt edici ağ da üretilen bu örneklerin gerçek veri setinden olup olmadığını tespit ederek çıktının üretici tarafından oluşturulma durumunu ayırt etmeye çalışmaktadır. Bu iki ağ birbirlerine karşı çalışmaktadır. Üretici ağ ayırt edici ağın tahminlerini zorlaştırırken, ayırt edici ağ da daha iyi bir şekilde gerçek ve üretilen veriyi ayırt etmeyi öğrenmektedir. Ayırt edici ağ aynı zamanda verinin içerisindeki farklı faktörleri de ayırt edebilmektedir. Bu işlemler sırasında her iki ağ da farklı faktörleri temsil eden özellikleri ayrıştırarak bağımsız bir temsil elde etmeye çalışmaktadır.

2.3. Duygu Tanıma Ayrıştırılmış Temsil ve Önyargı

Literatürde yapılan çalışmalarda, yüz görsellerinden ifade tanıma sistemleriyle doğrudan ilişkisi olmayan ya da bu sistemi yanıltabilecek özelliklerin olduğu ve bu özelliklerin doğruluğu önemli bir ölçüde etkilediği belirtilmektedir. Görüntülerdeki yaş, cinsiyet, ten rengi, sakal, bıyık gibi özelliklerin ve yüzde bulunan küpe, gözlük vb. karakteristik özelliklerin ifade tanıma için doğrudan ilişkili özellikler olmadığından yüz ifade tespitinde performansı düşürdüğü vurgulanmaktadır. [30]

Yapılan bu çalışmada [31] kimlik ve poz bilgilerinin verilen görüntü temsillerinden çıkarılarak ifade tespit sistemi geliştirilmesi hedeflenmiştir. Bunun için GAN tabanlı IPD-FER modeli önerilmiştir. Bu modelde ifadesiz yüz özelliklerini üretebilen önceden eğitilen bir yüz tanıma modeli kullanılmıştır. Poz kodlayıcı ve ifade kodlayıcı olmak üzere iki farklı kodlayıcı ifade veri seti ile eğitilmiştir. Kimlik özelliği ve poz birleştirildikten sonra üretken ağ ile kimlik ve pozun olduğu ancak ifadesiz olan görüntüler üretilmesi beklenmiş ve üç özellik birleşimi ile üretilen görüntüler karşılaştırılarak sadece ifade özellikleri kalacak şekilde ayrıştırmanın tamamlanmış olduğu açıklanmıştır. Sonuçlar TDGAN, IF-GAN, DLN, Cross-VAE gibi yöntemlerle karşılaştırılmıştır. Bu yöntemlere göre belirgin bir doğruluk artışı gözlemlenmiştir.

Üç farklı yöntem ile karşılaştırma yapan bu çalışmada [32] yüz ifadesi tanıma çalışmalarında ön yargı ve adalet incelenmiştir. RAF-DB ve CelebA veri setlerini kullanarak yapılan bu çalışmada başarılar doğruluk ve adalet metrikleri ile ölçülmüştür. Önyargı ölçümleri farklı demografik gruplar arasındaki doğruluk oranlarındaki farklılıkların incelenmesiyle yapılmıştır. Birinci yöntem olan temel model eğitiminde CNN ağı kullanılarak yüz ifadelerinin tespit modeli geliştirilmiştir. İkinci yöntem belirlenen yaş, cinsiyet ve ırk gibi hassas özelliklere duyarlı olacak şekilde tasarlanmıştır. ResNet-18 ağı kullanılarak özellik vektörü çıkarılmıştır. Duyarlılık vektörü ile toplanan özellik vektörü sınıflandırma katmanına verilerek daha adil bir model eğitilmeye çalışılmıştır. Son yöntem olarak da ayrıştırılmış yöntem kullanılmıştır. Bu yöntemde daha önce belirlenen hassas özelliklerin modelden çıkarılması ve tahmin işlemine dahil edilmemesi amaçlanmıştır. İkinci yöntemde

olduğu gibi ResNet-18 ağı kullanılarak özellik vektörü elde edilmiştir. Sonrasında biri temel sınıflandırma, diğeri de duyarlı özellikleri tahmin edemeyecek şekilde olan iki paralel daldan oluşan bir ağ tasarlanmıştır. Bu yöntem ile önyargının azaltılması ve gerçek dünyada daha adil kullanılabilen model çalışması amaçlanmıştır. Makalede belirtilen ayrıştırılmış temsil yaklaşımında kullanılan kayıp fonksiyonu hassas özelliklerin öğrenilmiş temsillerden çıkarılmasını sağlamaktadır.

Ayrıştırılmış temsillerin belirlenen hassas özelliklerden doğru bir şekilde arındırıldığından emin olmak için çeşitli yöntemler geliştirilmiştir. Temsil vektörlerinin birbirini etkileyemeyen, bağımsız olduğunu ifade etmek için kullanılan bir kısıtlama yöntemi de ortogonalitedir. [33] Matematiksel olarak iki vektörün ortogonal olması demek, bu iki vektörün arasındaki açının 90 derece olduğunu ifade etmektedir. Temsil vektörleri için ele alındığında iki konseptin birbirine dik olduğu dolayısıyla bir vektör üzerindeki değişikliğin diğerini etkilemeyeceği kabul edilmektedir. Yapılan bir çalışmada [33] bu yöntem kullanılarak anlamlı ve hassas temsilleri birbirinden ayırmak için yeni bir ayrıştırma yaklaşımı sunulmaktadır. Uygulanan ortogonalite kısıtlamaları sayesinde ve anlamlı temsillerin hassas bilgilere karşı bağımsız olmasını sağlamak için entropi maksimizasyonu gibi teknikler kullanılmıştır. Aynı zamanda adil bir model oluşturmak da amaçlanmıştır. Beş farklı veri seti üzerinde yapılan deneyler sonucunda üç veri seti üzerinde mevcuttaki modellerden daha iyi performans sergilediği ve diğer iki veri seti için de benzer sonuçlar elde edildiği belirtilmektedir.

Literatürde bu amaçla yapılan çalışmaların bir kısmında da GAN kullanılmıştır [34-37]. Önerilen TDGAN modeli [38] iki daldan oluşan ayrıştırılmış bir üretken ağdır. Çalışmada yüz ifadeleri ile diğer yüz özelliklerinin ayrıştırılması hedeflenmiştir. İki ayrı ayırt edici ile kimlik ve ifade sınıflandırması yapılmaktadır. Rekabetçi öğrenme sayesinde de önerilen model, belirlenecek ifadeyi bir başka yüze aktarabilmektedir. Aynı zamanda da her ifade görüntüsü için diğer yüz ifadelerinden ayrılmış bir temsil öğrenebilmektedir. Sonuçlar karşılaştırıldığında önerilen modelin hem günlük hem de

laboratuvar ortamında elde edilen görseller ile yapılan testlere göre mevcut yöntemlerle rekabet edebildiği vurgulanmıştır.

Duygu tanıma sistemlerinde yüz morfolojisi, poz gibi etkenlerden kaynaklanan performans etkisini ayrıştırılmış temsil yöntemiyle ortadan kaldırmayı hedefleyen çalışmalar bulunmaktadır. Bu çalışmalardan biri olan [39] 2D yüz görüntüleri içeren veri seti kullanmaktadır. Çalışmada çok ölçekli bir kontraktif evrişimli ağ (CCNET) kullanılmıştır. Bu ağ sayesinde özellik temsilleri üretilebilmektedir. Kontraktif otokodlayıcı ile de duygu faktörlerinin diğer etkenlerden ayrıştırılması sağlanmaktadır. Çalışmanın sonucunda literatürdeki sonuçları yaklaşık %2,5 oranında iyileştirdiği görülmektedir.

Hassas özellik olarak kimlik bilgilerini ele alan bir diğer çalışma [40]'da çekişmeli üretici ağ mimarisine sahip olan TER-GAN (Transfer-based Expression Recognition Generative Adversarial Network) önerilmiştir. Sonuçlar CNN modeli ve TER-GAN için CK+, Oulu-CASIA, BU-3DFE, MMI veri setleri üzerinde değerlendirilmiştir. Buna göre TER-GAN'ın kimlik özelliklerini ayrıştırarak yüz ifadesi tanıma görevinde daha başarılı olduğu belirtilmektedir.

İki farklı dallanma kullanarak kimlik ve yüz ifadeleri tanıma yöntemi kullanan bir çalışmada [41] elde edilen yüz temsilinden kimlik temsili çıkarılarak duygu temsilini ayırmak amaçlanmıştır. Bu çerçevede sapma öğrenme ağı olarak da bilinen DLN (Deviation Learning Network) önerilmektedir. Bu ağ sapma modülü, yüksek dereceli modül, kitle katmanı ve hiyerarşik anotasyon gibi bölümleri içermektedir. Bunlarla birlikte gerçekleştirilen çalışmalarda doğruluk oranı ele alındığında FECNet ve AffectNet'ten daha iyi sonuçlar alındığı vurgulanmıştır.

CK+, MMI, Oulu-CASIA veri setlerini kullanan bir diğer çalışmada [42] üç farklı daldan oluşan ADFL (Adversarial Disentangled Feature Learning) önerilmektedir. Çalışmada nötr ve ifade özelliklerini ayrıştırmak hedeflenmektedir. Bu nedenle belirli ortak katmanlar kullanıldıktan sonra model üç farklı dala ayrılmaktadır. İfade dalı gerekli ifade bilgilerini çıkarmaktan sorumludur. Nötr ifade dalı ise çapraz entropi kaybı ve negatif entropi kaybı kullanıp nötr özellikleri çıkarırken ifade temsillerini

çıkarmaktadır. Son olarak iki dal arasındaki farkı hesaplayarak ifade temsilini çıkaran üçüncü bir dal tasarlanmıştır. Mevcut yöntemlerle kıyaslandığında önerilen yöntemin daha iyi bir performans gösterdiği belirtilmektedir.

Birden fazla özelliğin aynı anda ayrıştırılabilmesi için yapılan çalışmada [43], çekişmeli transfer yöntemi ele alınmıştır. Dikkat mekanizması kullanan bu modelde laboratuvarda oluşturulan veri setleri ve doğal ortamda alınan veri setleri ile deneyler gerçekleştirilmiştir. Sonuçlarına bakıldığında her iki grup için de sonuçların iyi bir performans sergilediği vurgulanmıştır.

Tablo 1'de ayrıştırılmış temsil öğrenimi metoduyla yapılan çalışmalar ve sonuçları paylaşılmıştır. Benzer ve farklı veri setleri üzerindeki metotlar incelenmiştir.

Tablo 1. Ayrıştırılmış Temsil Öğrenimi Literatür Karşılaştırması

Çalışma	Veri Kümesi	Örnek Sayısı	Özellik Sayısı	Hassas Özellik Sayısı	Hedef Özellik Sayısı	Sınıflandırma Doğruluğu	Hassas Doğruluk	Hedef Doğruluk
Sarhan ve arkadaşları [18]	CIFAR-10	60000	10	2	10	-	0.1907	0.9725
	CIFAR-100	60000	100	2	100	-	0.1447	0.7074
	YaleB	2414	190	5	38	-	0.52	0.85
	Adult	45222	14	1	1	-	0.6826	0.85
	German	1000	20	1	1	-	0.71	0.77
Xu ve arkadaşları[6]	RAF-DB	14388	40	3	7	0.65	-	0.74
	CelebA	202599	40	3	2	0.93	-	0.92
Önerilen Çalışma	AffectNet	440000	40	8	2	0.87	0.79	0.83
	CelebA	202599	40	8	2	0.81	0.85	0.85

Tabloda ilk sırada verilen [33] çalışmada seçilen veri setleri bizim problemimizle benzer datasetler üzerinde çalışmamıştır. Ancak önerilen yöntem ile belirlenen hassas özelliklerin ayrıştırılma doğruluğu karşılaştırılabilmektedir. CIFAR-10 ve CIFAR-100, sırasıyla on ve yüz sınıftan oluşan görsellerden oluşan veri setleridir. Bu çalışmada hassas doğruluk ne kadar düşükse özelliklerin o kadar iyi ayrıştırıldığı kabul edilmektedir. Bu doğrultuda çalışmanın CIFAR veri setleri üzerinde iyi bir performans

gösterdiği belirtilmektedir. YaleB, Adult ve German veri setleri üzerinde ise ayrıştırılma belirli bir oranda sağlanmaktadır.

Diğer çalışmada ise [44], çalışmamızda ortak olarak kullanılan CelebA veri seti bulunmaktadır. Bu çalışma ile paralel olan amacımız hassas özelliklerin çıkartılmasıdır. Çalışma kodlayıcı-çözücü mimarisi kullanmadan bir derin öğrenme ağı ile özellikleri ayrıştırmayı hedeflemektedir. CelebA veri seti için seçilen özellikler cinsiyet, yaş (yaşlı, genç) ve cinsiyet ve yaşı birlikte değerlendirecek şekilde belirlenmiştir. Çalışmada ele alınan üç farklı yöntem sonucunda tabloda gülümseme etiketine göre elde edilen ortalama doğruluk sonuçları paylaşılmıştır. Sınıflandırma doğruluğu olarak verilen sonuç temel model olarak ele aldıkları deney sonucunda ortalama olarak elde edilmiştir. Hedef etiket doğruluğu ise yapılan deneyler sonucunda veri artırımı yöntemi kullanılarak elde edilmiştir.

3. GEREÇ VE YÖNTEMLER

3.1. Araştırma Tasarımı

Çalışmada yüz görsellerinde aksesuar ve makyaj, sakal, saç gibi etkenlerin yüz ifadesi tanıma sistemine olan etkisini araştırarak, ayrıştırılmış temsil yöntemi ile özellik vektörlerinden çıkarılması sağlanarak sınıflandırıcının daha adil olmasını sağlamak amaçlanmaktadır. Hedeflerimizden biri, ayrıştırılmış gösterim tekniğini kullanarak bu faktörleri özellik vektörlerinden metodik olarak ortadan kaldırmaktır. Bu nedenle öncelikle kullanılacak veri setlerine karar verilmiştir. "Gülümseme" duygu durumunu ifadeleri içeren geniş yüz görüntüleri koleksiyonu ve 202599 görseli ile bilinen CelebA veri seti ve yedi farklı duygu sınıfı ile etiketlenmiş 420299 tane yüz görseli olan AffectNET veri setleri kullanılmıştır. Bunu takiben, deneylerin yürütülmesi sürecini kolaylaştırmak için bu veri kümelerinin daha küçük bölümleri oluşturulmuştur. Bu modellerin performansını değerlendirmek için doğruluk, validasyon doğruluğu, önyargı ve varyans dahil bir dizi ölçüm kullanılmıştır. Sarhan ve arkadaşlarının [33] yaptığı çalışma ile tutarlı olarak, deneysel metodolojimiz belirli araştırma hedeflerimizi etkili olarak ele alacak şekilde özelleştirilmiştir. Daha sonra, kapsamlı deneylere olanak sağlamak için bu veri kümelerinden belirli bölümler dikkatlice seçilmiştir. Araştırmamız, önerdiğimiz metodolojiden elde edilen sonuçların, önceki bir araştırmada açıklanan alternatif bir yaklaşımla elde edilenlerle ilişkili olarak incelenmesini kapsamaktadır. Araştırmamıza rehberlik edecek üç araştırma sorusu ve hipotez ileri sürülmüştür.

Araştırma sorusu 1: Yüz ifadesi tanıma sistemlerinin doğruluğu aksesuarlar, makyaj, sakal, bıyık, gözlük, küpe, kolye, şapka ve saç şekli gibi çeşitli faktörlerden etkilenir mi?

- Hipotez 1: Aksesuarların varlığı, yüz ifadesi algılamanın doğruluğunu önemli ölçüde etkiler.

Araştırma sorusu 2: Ayrıştırılmış (Disentangled) gösterim yöntemi kullanılarak hassas özelliklerin ortadan kaldırılmasının tanıma doğruluğu üzerindeki etkisi nedir?

- Hipotez 2: Disentangled yöntemi, hassas özelliklerin ayrılmış vektörler içerisine sızmasını etkili bir şekilde önler.

Araştırma sorusu 3: Yüz ifadesi tanıma sistemlerinin verimliliğini artırmak için algoritma seçmenin ve optimize etmenin sonuçları nelerdir?

- Hipotez 3: Optimal algoritmaların seçimi yüz ifadesi tanıma sistemlerinde performans oranlarının artmasını sağlayacak.

3.2. Veri Seti ve Ön İşleme

Kullanılan veri setlerinin ilki CelebA veri setidir. Bu veri seti toplamda 10177 farklı kişinin toplamda 202599 yüz görüntüsünü içermektedir. Her bir görsel 40 farklı özellik için etiketlenmiştir. Şekil 1'de veri setinden örnek görüntüler verilmektedir.

Şekil 1. CelebA Veri Seti Örnek Görüntüler

Çalışmalarda kullanılırken duygu tespiti için veri setinde bulunan "Gülümseme (Smiling)" sınıfı hedef sınıf olarak dikkate alınmaktadır. Bu duygu durumunu tespit etmeye çalışırken bıyık, sakal, gözlük, yoğun makyaj gibi sınıflar hassas özellik olarak belirlenmiştir. Tablo 1' de seçilen bütün hassas özellikler ve CelebA veri setine göre duygu etiketi olan "Gülümseme" dağılımları verilmektedir. Hassas özelliklere sahip olan görseller için hedef sınıf sayıları tespit edilmiştir.

Tablo 2. CelebA Veri Seti Hassas Özellik Dağılımları

Özellik / *Gülümseme*	**Perçem**	**Gözlük**	**Bıyık**	**Sakal**	**Küpe**	**Şapka**	**Kolye**	**Makyaj**
Var	*16709*	*5302*	*2699*	*11883*	*25199*	*3342*	*14969*	*46499*
Yok	*14000*	*7891*	*5718*	*21558*	*13077*	*6476*	*9944*	*31891*

Verilen örnek görüntülerin etiketleri Tablo 2' de gösterilmektedir. Bu tabloda "1" ile ifade edilen özellikler ilgili görselin taşıdığı bir özellik olduğu anlamına gelmektedir. Aynı şekilde "-1" ifadesi de o özelliğin görselde bulunmadığını ifade etmektedir.

Tablo 3. CelebA Veri Seti Etiket Örnekleri

Perçem	-1	1	-1	-1	-1
Gözlük	1	-1	-1	-1	-1
Bıyık	-1	-1	-1	-1	-1
Sakal	-1	-1	-1	1	1
Gülümseme	1	1	-1	1	-1
Küpe	-1	1	-1	-1	-1
Şapka	-1	-1	-1	-1	1
Kolye	-1	-1	-1	-1	-1
Makyaj	-1	1	-1	-1	-1

İkinci veri setimiz olan AffectNet [15] ise manuel olarak yedi farklı duygu ile etiketlenen yaklaşık olarak 440 bin görsel içermektedir. Eğitim ve validasyon seti olarak ayrılarak yayınlanmıştır. Şekil 2'de eğitim veri setinden örnek görüntüler verilmiştir.

Şekil 2. AffectNet Veri Seti Örnek Görüntüler

Veri setin içerisinde 82915 tane yüz tespit edilememiş görsel bulunmaktadır. Çalışmaya bu yüzler dahil edilmemiştir. AffectNet veri seti sadece 0 (nötr), 1 (mutlu); 2 (üzgün), 3 (şaşkın), 4 (korku), 5 (tiksinti), 6 (kızgın), 7 (küçümseme) duygu

durumlarını içeren bir veri seti olduğundan dolayı hassas özelliklerin etiketleme işlemi CelebA veri setinde olduğu gibi 40 özellik olacak şekilde eğitim seti üzerinde ayrıca bir etiketleme işlemi yapılmıştır. Bu işlem için AIML- Human Attributes Detection with Facial Feature Extraction [45] çalışması kullanılmıştır. Etiketleme işlemi ile birlikte sonuç olarak elimizde 199127 tane görsel olmaktadır. Tablo 3'ten yeniden etiketleme işlemi yapılan seçilen hedef değişken olan "Gülümseme" etiketine göre hassas etiket dağılımları görülmektedir. Yoğun makyaj olan görsellerin diğer hassas özelliklere sahip görsellere göre daha fazla sayıda olduğu gözlemlenmektedir. Ancak kullanılan iki veri seti içerisinde birden fazla hassas özelliğe sahip görseller de bulunmaktadır. Örneğin gözlük takan bir kişinin aynı zamanda bıyığı ya da küpesi de olabilmektedir. Çalışmamızda bu özelliklerin birlikte olduğu durumlar gruplandırılmaya çalışıldığında tek bir hassas özelliği taşıyan görsel sayısının deneylerde yetersiz kalması göz önüne alınılarak sadece seçilen hassas özelliğin durumuna odaklanılarak tekli bir gruplandırma yapılmıştır.

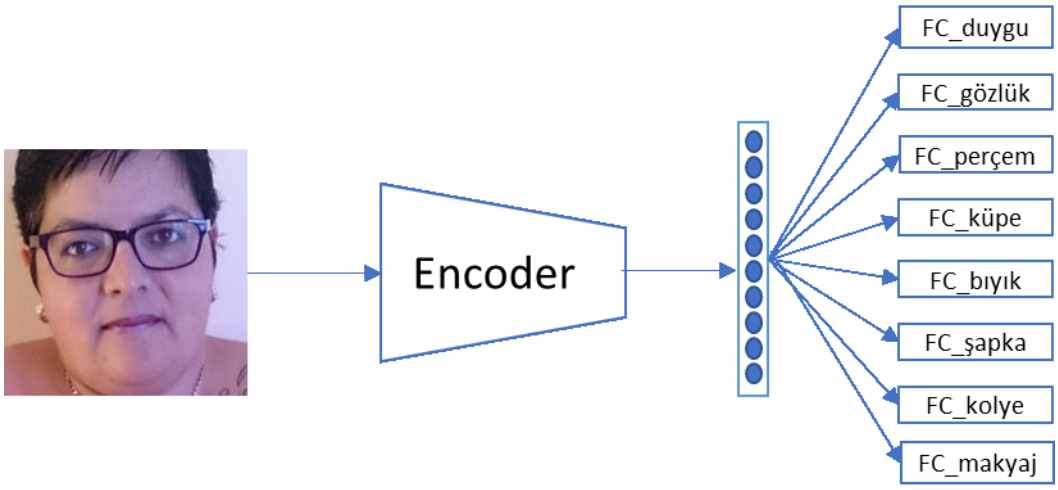
Tablo 4. AffectNET Veriseti Hassas Özellik Dağılımları

Özellik / *Gülümseme*	**Perçem**	**Gözlük**	**Bıyık**	**Sakal**	**Küpe**	**Şapka**	**Kolye**	**Makyaj**
Var	*12876*	*9502*	*4960*	*7764*	*24384*	*3541*	*4549*	*39171*
Yok	*11074*	*8391*	*9549*	*14908*	*10253*	*4899*	*417*	*28084*

3.3. Ayrıştırılmış Temsil Öğrenimi

Ayrıştırılmış temsiller bahsedilen problemleri çözme konusunda önemli bir rol oynamaktadır. Bunun sebebi de bu gösterim ile daha anlaşılabilir, genelleştirilebilir ve az çapraz bilgi içeren temsiller oluşturulabilmesidir. Ayrıştırılmış yöntem ile faktörler bağımsız olarak temsil edilebildiğinden dolayı daha anlaşılabilir temsiller sağlamaktadır. Elimizdeki veriye ait farklı özellikleri de ayrıştırarak modelin görmediği veri örneklerine uyarlamasını olumlu yönde etkilemektedir. Bu özelliklerinden dolayı ayrıştırılmış temsillerin makine öğrenmesi modellerinde daha etkili ve başarılı olabileceği belirtilmektedir [46].

Yapılan çalışmalarda temsillerin ayrıştırılması için farklı metotlar kullanılmaktadır. VAE kullanılarak yapılan bazı çalışmalarda kodlayıcının son katman haricindeki bütün katmanları ortak olarak kullanılabilmektedir. Son katmanda her özellik için ayrı katman kullanılmaktadır [44]. Bu şekilde istenilen özelliklerin ve hassas özelliklerin kayıp hesapları ayrı ayrı yapılarak eğitim gerçekleştirilebilmektedir. Önerilen mimarinin bu çalışma için tasarlanmış hali Şekil 3'te gösterilmektedir.



Şekil 3. Çalışma [6]'da kullanılan mimari tasarım örneği

4. DENEYLER

Bu çalışmada iki farklı yöntem ile duygu tanıma sistemi geliştirilmiştir. İlk olarak bu iki veri seti kullanılarak derin öğrenme yöntemleri ile model eğitilmiştir. Burada doğrudan bir özellik çıkarımı yöntemi kullanılmamıştır. Sonrasında değişimsel oto kodlayıcı kullanılarak bir sınıflandırıcı modeli eğitilmiştir. Son olarak VAE modelinin doğru bir şekilde hassas özelliklerden ayrıştırılmasını sağlamak amacıyla önerilen Sarhan ve arkadaşlarının [33] önerdiği model eğitilmiştir. Bu üç deneyin sonuçları iki veri kümesi için de karşılaştırılmıştır.

4.1. Kullanılan Veri Seti

CelebA ve tekrar etiketlenen AffectNet veri kümelerinin "Gülümseme" etiketi ile kullanılmak 64,128 ve 256 olmak üzere farklı boyutlarda önişlemeler yapılmıştır. Ancak yapılan deneye göre veri seti ve etiketleri değiştiğinden eğitilen modele uygun olacak şekilde eğitim/test performansı ve donanım kısıtlamaları düşünülerek eşit dağılımlı alt kümeler oluşturulmuştur. Dağılımlar dikkate alınırken öncelikle "Gülümseme" etiketi ele alınmıştır. Hedef etiket olarak belirlenen bu duygu durumunun belirlenen hassas özelliklerden (gözlük, bıyık, küpe, kolye, şapka, makyaj, perçem, sakal) en iyi şekilde ayrıştırılması amaçlanarak yapılan çalışmada seçilen tekli hassas özelliklerin olma ve olmama durumları için de aynı olacak şekilde planlanarak alt veri setleri tanımlanmıştır. Bu veri setleri oluşturulurken sadece bir hassas özellik dikkate alınmış olup belirtilen diğer özellikleri taşıyıp taşımama durumu değerlendirilmemiştir. Tablo 5'te CelebA ve AffectNet veri setinden oluşturulan alt veri setlerinin "Gülümseme" etiketine göre dağılımı paylaşılmıştır. Her dağılım kendi içerisinde eşitlenmiştir.

Tablo 5. Alt Veri Kümeleri Gülümseme- Hassas Etiket Dağılımı

	Gözlük	***Bıyık***	***Küpe***	***Kolye***	***Şapka***	***Makyaj***	***Perçem***	***Sakal***
CelebA	*5300*	*2699*	*13078*	*9944*	*3342*	*31891*	*14000*	*11883*
AffectNet	*8390*	*4960*	*10252*	*416*	*3541*	*28084*	*11074*	*7764*

Aynı görsel birden fazla hassas özelliğe sahip olabilmektedir. Bu sebeple sonuçların yorumlanmasında yararlanmak üzere Tablo 6'da hazırlanan alt veri setlerindeki diğer hassas özellik dağılımları da verilmektedir.

Tablo 6.CelebA Belirlenen Alt Veri Setlerinde Bulunan Diğer Hassas Özelliklerin Dağılımları

	Bıyık	***Gözlük***	***Kolye***	***Küpe***	***Şapka***	***Sakal***	***Makyaj***	***Perçem***
Bıyık	-	925	170	227	679	5.237	2	174
Gözlük	1.111	-	738	722	1.085	3.214	304	780
Kolye	214	677	-	6.759	542	553	11.488	4.563
Küpe	261	639	6.489	-	784	665	18.887	4.841
Şapka	693	912	426	676	-	2.325	598	179
Sakal	23.738	22.513	20.429	18.583	22.887	-	12.857	19.621
Makyaj	2	270	12.860	22.129	712	62	-	12.892
Perçem	285	853	5.759	6.522	226	1.419	14.390	-

AffectNet veri setinin yeniden etiketlenmesiyle elde edilen hassas özelliklere göre diğer hassas özelliklerinin bulunma dağılımını gösteren bilgiler Tablo 7'de verilmiştir.

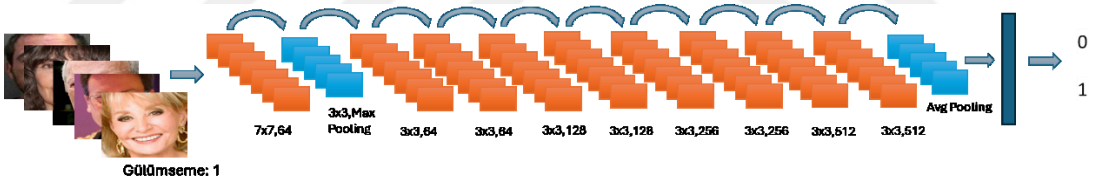
Tablo 7. AffectNet Dataseti Duygulara Göre Hassas Özellik Dağılımları

	Bıyık	***Gözlük***	***Kolye***	***Küpe***	***Şapka***	***Sakal***	***Makyaj***	***Perçem***
Bıyık	-	1.890	-	15	942	9.573	5	184
Gözlük	2.435	-	188	536	1.106	3.018	840	855
Kolye	-	31	-	606	9	-	801	511
Küpe	10	316	1.798	-	372	15	17.066	3.343
Şapka	1.125	1.003	23	494	-	1.600	822	503
Sakal	9.565	2.299	-	18	1.345	-	13	313
Makyaj	8	722	3.568	23.129	861	19	-	10.651
Perçem	263	850	2.398	5.327	554	453	11.805	-

4.2. Ayrıştırılmış Temsil Öğrenmesi ile Önyargısız Duygu Tanıma Sistemi

4.2.1. Araştırma sorusu 1: Görünüm Faktörlerinin Etkisi

Belirlenen hassas özelliklerin etkisinin araştırmak için öncelikle temel sınıflandırıcı modeli tasarlanmıştır. Temel modelimizin amacı veri kümelerimiz ile duygu tanıma modeli eğitmektir. Literatürde kullanılan ResNet mimarilerinin beş farklı versiyonu bulunmaktadır: ResNet-18, ResNet-34, ResNet-50, ResNet-101 ve ResNet-150. ResNet-18 modeli hız, performans ve veri kümesi boyutu parametrelerine göre kısıtlamalar göz önünde bulundurulmuş ve literatürde yapılan çalışmalar ile karşılaştırma yapabilmek için bu mimari modelimizde kullanılmıştır. Şekil 5, ResNet-18 mimarisini göstermektedir. AffectNet veri kümesinden "Gülümseme" etiketine göre gruplanmış yüz görsellerini almaktadır. Ancak eğitim aşamasında sonradan etiketlenerek elde edilen "Gülümseme" etiketi kullanılmıştır. Verilen mimari iki veri seti için de ortak olarak kullanılmıştır. Şekilde verilen çıktılarda "0" gülümseme yok, "1" gülümseme var anlamına gelmektedir.



Şekil 4. Temel Sınıflandırıcı ResNet-18 Mimarisi

AffectNET ve CelebA veri setleri için de öncelikle aynı parametreler kullanılarak eğitimler gerçekleştirilmiştir. 32 batch size, 30 epoch ve erken durdurma yöntemiyle tamamlanan eğitimlerde sırasıyla %87 ve %81 validasyon doğruluğuna erişilmiştir. Değerlendirilen etiket "Gülümseme" olduğundan ve temsil ediliş şekli 0-1 olarak belirtildiğinden dolayı kullanılan kayıp fonksiyonu "binary_crossentropy" olarak seçilmiştir. Modelin performansının değerlendirilmesi için epoch ve batch size sayıları test edilmiş ve en yüksek elde edilen sonuçlar Tablo 7'de paylaşılmıştır. Verilen epoch sayısı deneyde eğitilebilecek en yüksek değer olarak belirtilmiştir. Erken durma yöntemi ile 5 epoch boyunca modelde bir gelişme görülmemesi durumunda eğitim durdurulmuştur.

Tablo 7. ResNET-18 Mimarisi ile Temel Sınıflandırıcı Model Sonuçları

Çalışma	Veri Kümesi	Sınıflandırma Doğruluğu
Temel Sınıflandırıcı - ResNET-18	AffectNet	0.87
	CelebA	0.81

Eğitilen iki modelin performansının belirlediğimiz hassas özellikler üzerinde test edilmesi için özelliklerin bulunduğu (hassas etiket değeri=1) durumu için gruplanmıştır. Bu gruplarda gülümseme var/yok etiketi dengeli bir dağılım göstermektedir. Her grup için testler gerçekleştirilmiş ve sırasıyla AffectNet ve CelebA veri setleri için sonuçlar Tablo 8'de paylaşılmıştır.

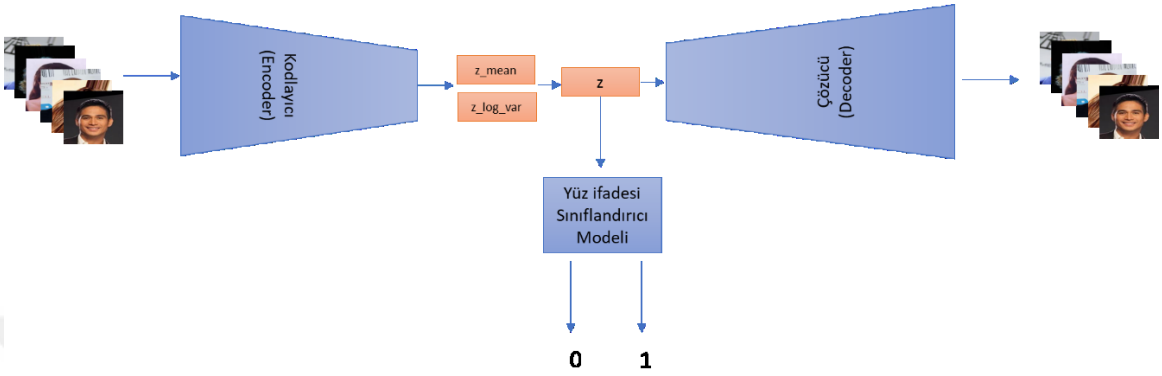
Tablo 8. Temel Sınıflandırıcı Hassas Özellik Test Sonuçları

Özellik	**CelebA**				**AffectNet**			
	Doğruluk	Kesinlik	Geri Çağırma	F1 Skoru	Doğruluk	Kesinlik	Geri Çağırma	F1 Skoru
Bıyık	0.32	0.32	1.0	0.48	0.66	0.66	1.0	0.79
Gözlük	0.40	0.40	1.0	0.57	0.47	0.47	1.0	0.64
Perçem	0.54	0.54	1.0	0.70	0.46	0.46	1.0	0.63
Kolye	0.60	0.60	1.0	0.75	0.08	0.08	1.0	0.15
Küpe	0.65	0.65	1.0	0.79	0.30	0.30	1.0	0.46
Makyaj	0.59	0.59	1.0	0.74	0.42	0.42	1.0	0.59
Şapka	0.34	0.34	1.0	0.50	0.58	0.58	1.0	0.73
Sakal	0.59	0.59	1.0	0.74	0.42	0.42	1.0	0.59

4.2.2. Araştırma sorusu 2: Ayrıştırılmış (Disentangled) Temsil Yönteminin Etkisi

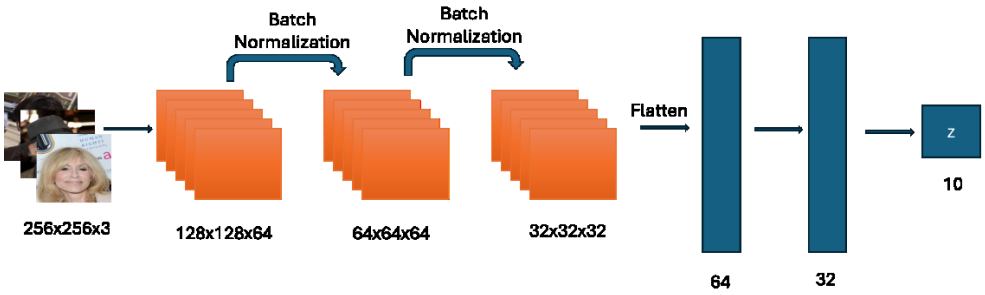
İkinci deney stratejimiz olan Beta-VAE mimarisi Şekil 5'te gösterilmektedir. Kodlayıcıdan elde edilen z temsili sınıflandırıcı modeline girdi olarak verilerek

"Gülümseme" etiketi ile eğitilmektedir. Kodlayıcı, çözücü ve sınıflandırıcı model mimarileri sırasıyla Şekil 5, Şekil 6 ve Şekil 7'de gösterilmektedir.



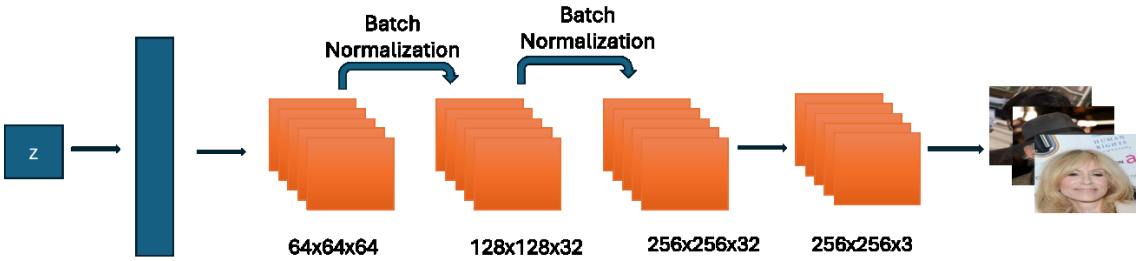
Şekil 5. Beta-VAE Ayrıştırılmış Temsil Model Mimarisi

Şekil 6'da verilen kodlayıcı mimarisinde üç adet konvolüsyonel katman bulunmaktadır. RGB formatında alınan görseller bu katmanlardan geçirilerek kodlayıcı içerisinde özelliklerin ayrıştırılması sağlanmaktadır. Çıktı olarak elde edilen z_mean ve z_log_var ile z temsili elde edilebilmektedir. Bu temsil girdi olarak verilen görselin özelliklerini barındırmaktadır. Bu sayede ayrı bir sınıflandırıcı ağı, bu z temsilini girdi olarak alarak eğitim gerçekleştirilebilmektedir.



Şekil 6. Kodlayıcı Mimarisi

Çözücü mimarisi Şekil 7'de gösterilmiştir. Sıkıştırılan görsel temsilinin ilk girdi ile aynı boyuta getirilebilmesi için katmanlar içermektedir. Bu katmanlarda transpose konvolüsyonel katmanları eklenerek oluşturulmaktadır. Bu sayede girdi olarak aldığı z temsili sonucunda çözücü ilk görseli oluşturmaya çalışmaktadır.



Şekil 7. Çözücü Mimarisi

Yeniden yapılandırmanın başarı durumunun takip edilebilmesi için kayıp hesaplaması yapılmaktadır. Bu kayıp sonucuna göre elde edilen çözücü çıktısının girdi ile farkına bakılmaktadır.

Bu aşamada belirlenmesi kritik öneme sahip olan beta parametresi için ayrı ayrı deneyler yapılmıştır. Deneylerin sonucunda en iyi performansa ulaşan değerler belirlenmiştir. Beta-VAE mimarisiyle eğitilen model sonuçları Tablo 9'da verilmiştir. Bu modelde hassas özellik etiketleri ayrıca bulunmamaktadır. Hedef etiketimiz olan "Gülümseme" 0 ve 1 olarak gösterilmektedir. Belirlenen hassas özelliklerle ilgili hazırlanan veri kümesi dağılımlarına göre bu modeller ayrı ayrı eğitilmiş ve sonuçları Tablo 9'da paylaşılmıştır.

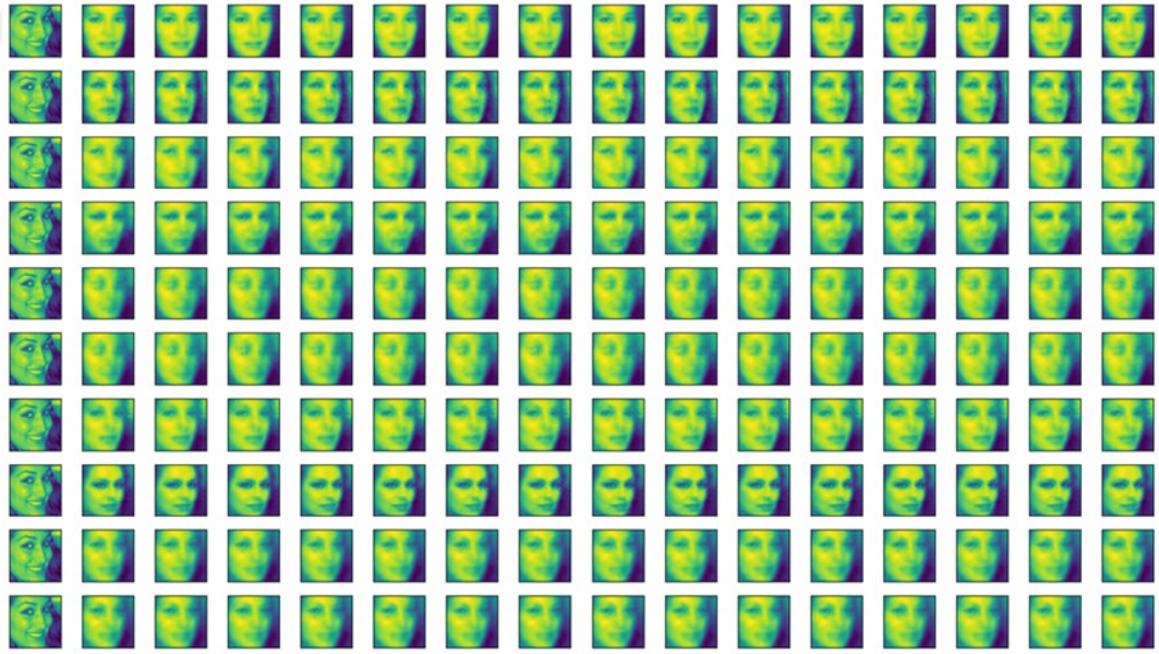
Tablo 9. B-VAE ile Eğitilen Duygu Tanıma Model Sonuçları

Veri Kümesi	*Sınıflandırma Doğruluğu*
CelebA	0.7926
AffectNet	0.7415

CelebA veri seti için verilen sınıflandırma doğruluğu 64x64 boyutlu veri ile beta=10 modelinde elde edilmiştir. AffectNet veri seti içinse beta=1 ve 64x64 boyutlu görseller ile eğitilmiş model ile elde edilmiştir. Bu parametrelerin belirlenebilmesi için yapılan deneyler ve sonuçları araştırma sorusu 3'te detaylı bir şekilde anlatılmıştır.

Beta-VAE modelimizde kodlayıcının oluşturduğu gizli temsil görselleştirilmiş ve odaklandığı özellikler tespit edilmeye çalışmıştır. Şekil 8'de veri setinden rastgele seçilen yüzün gizli temsilleri görselleştirilmiştir. Bu şekilde her bir satırın temsil ettiği

özellik yapılan diğer çalışmalarda manuel olarak karşılaştırma yapılarak ya da göz ile belirlenmeye çalışılmıştır. Ancak bu noktada elde edilen temsillerin hangi özelliğe ait olduğu belirgin bir doğrulukla tespit edilememektedir. Çünkü bazı özelliklerin temsillere dahil edilip edilmediği insan gözü ile ayırt edilememektedir. Ancak görselleştirme ve test adımlarıyla birlikte özellik tahmini yapmak mümkün olabilmektedir. Bu şekilde temsillerin hassas özelliklere ait bilgileri içermediğinden de emin olunamamaktadır.

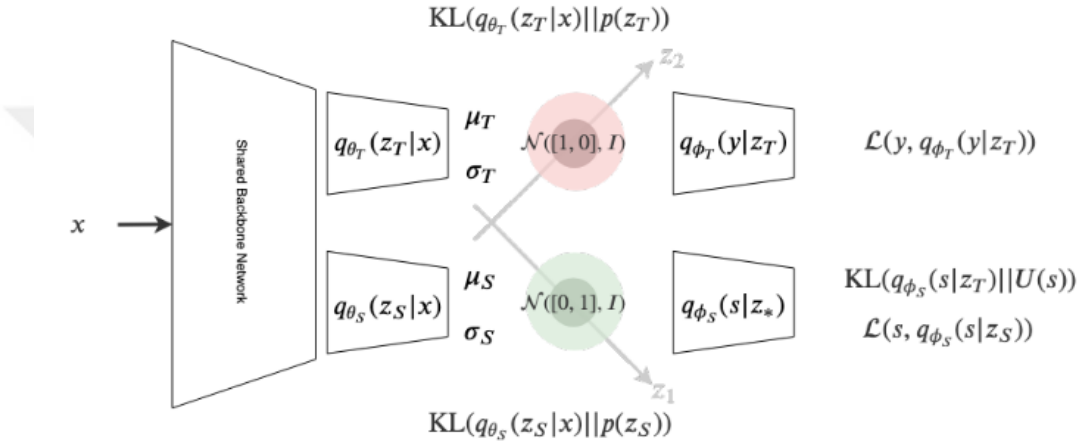


Şekil 8. Veri setinden rastgele seçilen bir görüntünün kodlayıcı çıktısının görselleştirilmesi

Ayrıştırmaların doğru bir şekilde yapılmış olunduğundan emin olmak için Sarhan ve arkadaşları [18], ortogonalite hesabı ile belirlenen hassas ve hedef özelliklerin temsillerinden mümkün olduğunca birbirlerinden ayrıştırılmasını amaçlamışlardır. Şekil 9'da bu çalışmada önerilen model tasarımı verilmektedir.

Veri kümesi girdisi X olarak gösterilmektedir. Ortak ağ ise girdiyi işleyerek iki farklı düşük boyutlu temsil oluşturmaktadır. Temsillerden z_t hedef, z_s hassas özellikleri temsil etmektedir. μ ortalama, σ ise standart sapma olarak belirtilir. İki temsilin de normal dağılıma uyması beklenmektedir. Bunun için iki dağılım arasındaki fark KL kaybı ile ölçülmektedir. z_t ve z_s'nin ayrı ayrı standart normal dağılıma olan yakınlığı

hesaplanırken aynı zamanda z_t'nin hassas özelliklerin bilgisini içerip içermediği de hesaplanmaktadır. L ile belirtilen kayıp hesaplamasında ise z_t ve hedef özellik kullanılarak tahminleme arasındaki fark takip edilmektedir. İkinci olarak verilen L ise hassas özelliklerin hassas temsiller için yapılan tahmin arasındaki farkı ölçmektedir.



Şekil 9. Sarhan ve Arkadaşları tarafından önerilen yapı [33]

Bu mimari ile belirlenen veri setlerimiz ve hassas özelliklerimize göre bir dizi deney gerçekleştirilmiştir. Bu deney sonuçlarında hassas özellik doğruluğunun düşük olması beklenmektedir. Bu doğruluk bize elde edilen temsilin hassas etiketler ile verdiği performansı göstermektedir. Hedef etiket doğruluğu ise gülümseme sınıfı yüz ifade sınıflandırma modelimizin performansını göstermektedir. Hesaplama gücünün ve zamanını dikkate alarak her model iyileştirmeleri gözlemlenerek en fazla 100 epoch olacak şekilde eğitilmiştir. Tasarlanan deneyler CelebA ve AffectNet veri setleri için gerçekleştirilmiş, sonuçları Tablo 10'da gösterilmektedir.

Tablo 10. Ayrıştırılmış Temsil Yöntemi ile Duygu Tanıma Eğitim Sonuçları

	CelebA			AffectNET		
Hassas Özellik	Örnek Sayısı	Hassas Etiket Doğruluğu	Hedef Etiket Doğruluğu	Örnek Sayısı	Hassas Etiket Doğruluğu	Hedef Etiket Doğruluğu
Bıyık	10798	0.85	0.85	19840	0.78	0.83
Gözlük	21200	0.90	0.79	33460	0.89	0.83
Kolye	36176	0.96	0.82	1664	0.79	0.83
Küpe	52308	0.94	0.82	41008	0.82	0.81
Şapka	13368	0.73	0.88	14164	0.81	0.80
Sakal	47532	0.20	0.24	20645	0.86	0.82
Makyaj	127564	0.52	0.81	112336	0.69	0.70
Perçem	56000	0.76	0.80	44296	0.81	0.82

Belirlenen hassas özellikler kendi içerisinde fiziksel faktörler (sakal, bıyık, saç modeli), aksesuarlar (gözlük, küpe, kolye, şapka), kozmetik faktörler (makyaj) şeklinde gruplanabilmektedir. Sonuçlar bu çerçevede yorumlandığında CelebA veri seti için fiziksel özelliklerin ortalama hassas etiket doğruluğu %60,3 olarak tespit edilirken AffectNet veri seti için %81,6 doğruluk oranı elde edilmiştir. Aksesuar grubu içinse CelebA ve AffectNet veri setleri için sırasıyla %88,25 ve %82,75 hassas etiket doğruluğu elde edilmiştir. Kozmetik faktör olarak ele alınan makyaj etiketi için ise %52 ve %69 doğruluk gözlemlenmiştir.

4.2.3. Araştırma sorusu 3: Algoritma Seçimi ve Optimizasyonu

Ayrıştırılmış temsil öğrenimi ile temel sınıflandırıcı test sonuçları karşılaştırıldığında ayrıştırılmış temsil algoritmasının öne çıktığı görülmektedir. Ancak bu algoritmanın performansı model mimarilerine ve parametrelere bağlıdır. Beta-VAE modeli için beta değeri kritik öneme sahiptir. Yapılan bu çalışmada farklı beta değerleriyle deneyler gerçekleştirilmiştir. Beta değerinin 1'e eşit olması durumu VAE ile aynı görevi yaptığı anlamına gelmektedir. 1'den büyük olması durumunda ayrıştırılmanın etkisini gözlemlemek amacıyla iki farklı beta değeri seçilmiştir. Bu beta değerleri belirlenirken ara değerler ve 100,150 gibi yüksek değerler de çalışmaya dahil edilmiştir. Ancak bu derece büyük beta değerlerinin doğruluk ve ayrıştırma oranın olumlu bir etkisi olmadığı gözlemlendiği için tabloda verilen sonuçlar karşılaştırılmıştır. Hesaplama süresi ve gücünden dolayı daha büyük beta değerleri denenmemiştir. Tablo 11'de deney sonuçları paylaşılmaktadır.

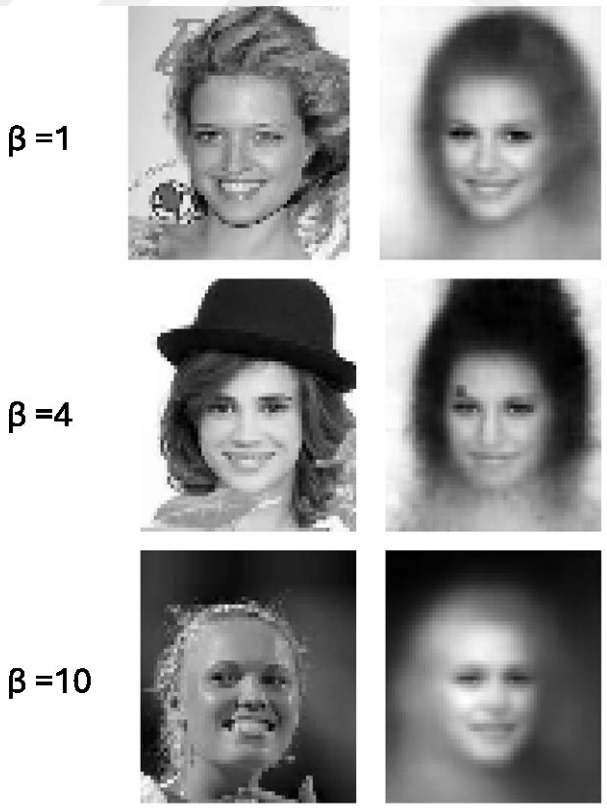
Tablo 11. 64x64 Boyutlu CelebA Veri Setinin Beta Değerlerine Göre Eğitim Sonuçları

Beta Değeri	*Toplam Kayıp*	*KL-Kaybı*	*Yeniden yapılandırma Kaybı*	*Sınıflandırma Doğruluğu*
1	*2239.71*	*23.73*	*2215.98*	*79.15*
4	*2296.56*	*15.49*	*2234.58*	*79.14*
10	*2376.49*	*9.74*	*2278.99*	*79.26*

Beta değeri 1 olarak seçildiğinde KL kaybına göre modelin gizli alanı düzgün bir şekilde düzenlemeye çalıştığı görülmektedir. Yeniden yapılandırma kaybının ise karşılaştırmalara göre daha düşük olduğu gözlemlenmektedir. Beta 4 olduğunda ise KL kaybı düşmüş, yeniden yapılandırma kaybı ise artmaktadır. Ancak eğitim doğruluğunda belirgin bir fark görülmemektedir. Beta değeri 10 olarak model eğitildiğinde ise KL kaybı daha da düşmüş ve yeniden yapılandırma kaybı beta 4'e göre daha da artmıştır. Model doğruluğunda bir değişim gözlemlenmemiştir.

Beta değerinin artmasıyla birlikte KL kaybı azalırken yeniden yapılandırma kaybının arttığı söylenebilmektedir. Düşük beta değerlerinde gizli temsil daha iyi düzenlenmektedir.

Şekil 10'da 64 boyutlu görseller ile eğitilen Beta-VAE modellerinin beta değerine solda girdiler sağda ise yeniden oluşturulan yüzler görülmektedir. Beta 1 için görseller ele alındığında yüz hatlarının benzerliği görülmektedir. Ancak gülümseme şekli saç detayları bulanık görünmektedir. Beta 4 iken girdi olarak belirlenen görselde bulunan şapkanın çıktıda da belirgin bir şekilde oluştuğu görülmektedir. Ancak yine saç detayları bulanıktır. Perçem için kaş üzerinde noktasal olarak bir bölge oluşturulmuştur. Son olarak beta 10 için saç rengi ve yüz duruşu için genel bir temsil elde edildiği söylenebilmektedir. Ancak gülümseme şekli ve detaylar burada da net bir şekilde görülmemektedir.



Şekil 10. 64x64 Boyutlu CelebA Veri Setinin Beta Değerlerine Göre Beta-VAE Çıktıları

Görsellerdeki bulanıklık ele alındığında daha fazla detayın öğrenilebilmesi için boyut artırımı ele alınmıştır. 128x128 boyutlu görseller ile deneyler tekrarlanmıştır. Bu sonuçlar Tablo 12'de verilmektedir.

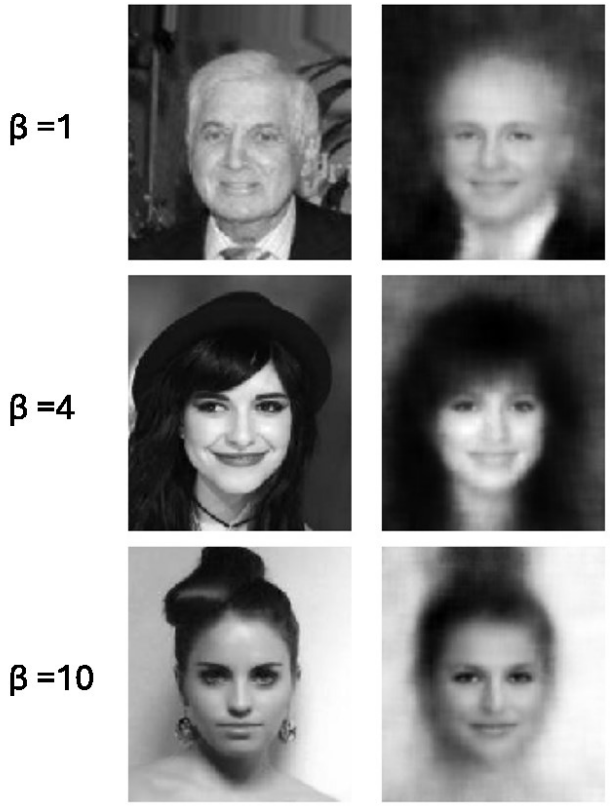
Tablo 12.128x128 Boyutlu CelebA Veri Setinin Beta Değerlerine Göre Eğitim Sonuçları

Beta Değeri	*Toplam Kayıp*	*KL-Kaybı*	*Yeniden yapılandırma Kaybı*	*Eğitim Doğruluğu*
1	8844.25	33.03	8811.21	79.24
4	8960.06	23.57	8865.75	79.23
10	*9085.72*	*18.35*	*8902.21*	*79.17*

Tabloya göre beta 1 deneyinde KL kaybı daha yüksek ve yeniden yapılandırma kaybı daha düşüktür. Eğitim doğruluğu da 64x64 boyutlu deneylere yakın olarak gözlemlenmiştir. Beta 4 değerinde ise KL kaybı azalmıştır. Gizli temsil düzenlemesi azalırken yeniden yapılandırma performansı da düşmüştür. Beta 10'da ise hem KL kaybı düşmüş yeniden yapılandırma kaybı artmıştır. KL kaybını azalırken yeniden yapılandırma performansı da azalmaktadır. Bu modelin gizli temsil düzenlemesiyle yeniden yapılandırma performansı arasında bir denge olabileceğini göstermektedir. Düşük beta değerlerinde gizli temsil daha iyi düzenlenirken, yüksek beta da yeniden yapılandırma performansını düşürmektedir.

Şekil 11'de 128x128 boyutlu girdi görselleri ve yeniden yapılandırılmış görüntüler verilmektedir. Beta 1 durumunda duruş ve yüzün genel hatları yeniden yapılandırılmış görselde öne çıkmaktadır. Aynı zamanda gömlek yakaları ve ceketin de temsilde yer aldığı görülmektedir. Boyun bölgesindeki değişimler belirgin bir biçimde yansıtılmaktadır. Ancak saç şekli gibi detaylar bulanıktır. Beta 4'te şapka takan bir kadın görseli bulunmaktadır. Saç rengine yakın renkte görünün şapka yapılandırılmış

görselde az da olsa belirgin haldedir. Saç şeklinde perçem büyük oranda temsil edilmiştir. Beta 10 değeri ile elde edilen son satırdaki görselde ise küpelerin temsilde detaylandırılmadığı ancak saç şeklinin belirgin olarak yansıtıldığı görülmektedir.



Şekil 11. 128x128 Boyutlu CelebA Veri Setinin Beta Değerlerine Göre Beta-VAE Çıktıları

Tablo 13'te boyutlara göre yapılan deneylerde özellik bazlı kümeler ile beta değerlerine göre gerçekleştirilen test sonuçları verilmektedir. Bu sonuçlara göre yeniden yapılandırma görsel sonuçlarında görülenin tersine test doğruluklarında 64 ve 128 boyut arasında beta değerlerine göre bir fark görülmemektedir.

Tablo 13. CelebA Veri Setinin Beta Değerlerine ve Hassas Özelliklere Göre Test Doğrulukları

Özellik	**64x64**			**128x128**		
	Beta Değeri			Beta Değeri		
	1	4	10	1	4	10
Bıyık	0.67	0.67	0.67	0.67	0.67	0.67
Gözlük	0.59	0.59	0.59	0.59	0.59	0.59
Perçem	0.45	0.45	0.45	0.45	0.45	0.45
Kolye	0.39	0.39	0.39	0.39	0.39	0.39
Küpe	0.34	0.34	0.34	0.34	0.34	0.34
Makyaj	0.40	0.40	0.40	0.40	0.40	0.40
Şapka	0.65	0.65	0.65	0.65	0.65	0.65
Sakal	0.40	0.40	0.40	0.40	0.40	0.40

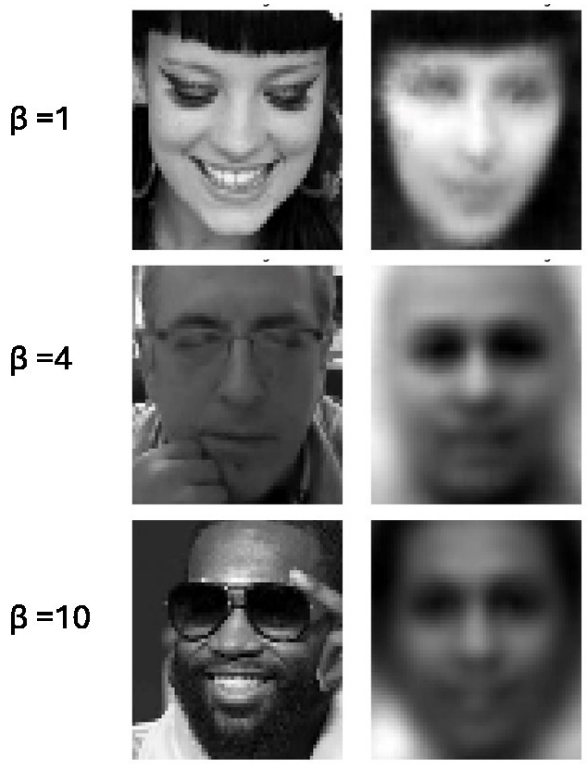
CelebA için gerçekleştirilen deneyler sırasıyla AffectNet için de tamamlanmıştır. Tablo 14'te 64x64 boyutları için beta değerlerine göre Beta-VAE performansı ve sınıflandırıcı doğruluğu paylaşılmıştır. Beta 1'de sınıflandırıcı doğruluğu diğer iki beta değerine göre en iyi performansı göstermektedir. Gizli uzayda düzenleme performansı da öne çıkmaktadır. Beta 4 değerine gelindiğinde sınıflandırıcı doğruluğunun düştüğü gözlemlenmektedir. Beta 10'da ise gizli temsil düzenlemesi azalarak yeniden yapılandırma performansı da düşmüştür. Sınıflandırma doğruluğu da en düşük bu değerde elde edilmiştir.

Tablo 14. 64x64 Boyutlu AffectNet Veri Setinin Beta Değerlerine Göre Eğitim Sonuçları

Beta Değeri	*Toplam Kayıp*	*KL-Kaybı*	*Yeniden yapılandırma Kaybı*	*Sınıflandırıcı Doğruluğu*
1	2444.19	21.08	2423.11	74.15
4	2496.04	12.67	2445.32	73.34
10	2555.96	7.27	2483.27	72.97

Şekil 12'de karşılaştırmalı olarak girdi ve yeniden yapılandırılan 64x64 görseller verilmiştir. İlk satırda beta 1 değeri yorumlandığında yüzün konumlandırılması, perçem yeniden yapılandırılmış görselde de görünmektedir. Ancak göz makyajı, gülümseme, küpe gibi detaylar temsil edilememiştir. Beta 4'te verilen gözlüklü erkek görseli ise göz bölgesinde koyuluk olarak temsil edilmiştir. Dudaklar benzer şekilde temsil ediliyor olarak görünse de detaylar net değildir. Son beta değerimizde ise yine gözlüklü gülen bir erkek görseli bulunmaktadır. Burada da göz bölgesinde bir karartı

görülmektedir. Tüm beta değerleri için genel yüz yapısının öğrenildiği gözlemlenmektedir. Ancak detaylar konusunda modeller zorlanmaktadır.



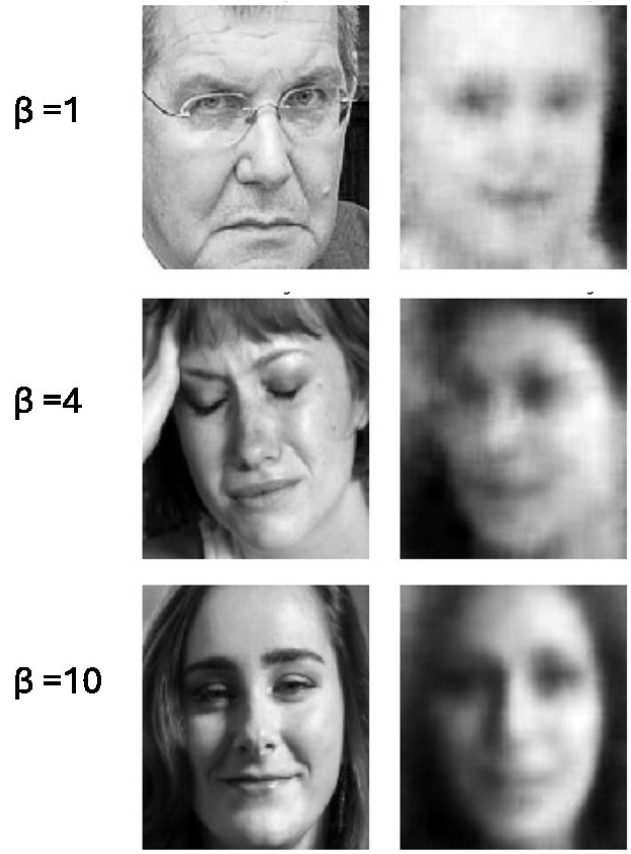
Şekil 12.64x64 Boyutlu AffectNet Veri Setinin Beta Değerlerine Göre Beta-VAE Çıktıları

128x128 boyutlu görseller ile eğitilen model sonuçları Tablo 15'te paylaşılmıştır. Diğer tablolarda elde edilen sonuçlara benzer bir şekilde beta 1 ile daha iyi bir temsil düzenlemesi ve yeniden yapılandırma performansı elde edilmiştir. Yine bu değerde en yüksek sınıflandırma doğruluğuna ulaşılmıştır. Beta 4 değerinde yeniden yapılandırma ve düzenleme arasında dengede bir sonuç gözlemlenirken beta 10'da temsil düzenlemesi ve yeniden yapılandırma performansı azalmıştır.

Tablo 15.128x128 Boyutlu AffectNet Veri Setinin Beta Değerlerine Göre Eğitim Sonuçları

Beta Değeri	*Toplam Kayıp*	*KL-Kaybı*	*Yeniden yapılandırma Kaybı*	*Eğitim Doğruluğu*
1	9679.46	30.06	9649.42	74.03
4	9773.26	21.21	9688.40	73.85
10	9884.52	15.78	9726.68	73.78

Beta değerlerine göre yeniden yapılandırılmış 128x128 boyutlu görsellerin çıktıları Şekil 13'te verilmektedir. Beta 1 için her ne kadar yüz konumu korunmuş olsa da yaşlı olarak verilen girdi görselinin hiçbir detayı yapılandırılmış görselde bulunmamaktadır. Beta 4 için seçilen örnekte eli başında olan ve gülümseme olmayan bir kadın görseli bulunmaktadır. Bu görsel yeniden yapılandırıldığında yüz konumlandırılması benzerlik gösterirken perçem gibi detaylar görülememektedir. Beta 10' da ise yine bulanık ve detaysız bir temsil elde edilmiştir.



Şekil 13. 128x128 Boyutlu AffectNet Veri Setinin Beta Değerlerine Göre Beta-VAE Çıktıları

Tablo 16'ya göre bıyık, gözlük perçem, kolye, küpe, makyaj ve sakal özellikleri için de beta ve çözünürlüğe göre belirgin farklar gözlemlenmemiştir. Küpe doğruluklarında 64x64 çözünürlükte beta değişiklinde yaklaşık %1-2'lik değişimler gözlemlenirken kolye için beta 10 değerinde %0.02'lik bir düşüş görülmektedir.

Tablo 16. AffectNet Veri Setinin Beta Değerlerine ve Hassas Özelliklere Göre Test Doğrulukları

Özellik	64x64			128x128		
	Beta Değeri			Beta Değeri		
	1	4	10	1	4	10
Bıyık	0.62	0.62	0.64	0.61	0.62	0.62
Gözlük	0.47	0.47	0.46	0.47	0.47	0.47
Perçem	0.45	0.45	0.45	0.45	0.45	0.45
Kolye	0.10	0.10	0.08	0.10	0.10	0.10
Küpe	0.33	0.33	0.31	0.32	0.33	0.32
Makyaj	0.42	0.42	0.42	0.42	0.42	0.42
Şapka	0.56	0.56	0.56	0.56	0.56	0.56
Sakal	0.42	0.42	0.42	0.42	0.42	0.42

5. SONUÇLAR

CelebA ve AffectNET veri kümeleri için duygu etiketi olarak sadece "Gülümseme (Smiling)" sınıfı dikkate alınmıştır. Bu sınıfa göre elimizdeki yüz görselleri gruplanmış ve ResNET-18 mimarisi ile oluşturulan modelimiz eğitilmiştir. Veri kümesi 0,2 oranında validasyon ve eğitim setlerine ayrıştırılmıştır. Adam optimizerı ve "categorical_crossentropy" kayıp fonksiyonu ile model eğitilmiştir. Eğitim erken durma yöntemi olarak en fazla 30 epoch olacak şekilde başlatılmıştır. 10 epoch sonucunda sırasıyla %81 ve %87 validasyon doğruluğuna ulaşılmıştır. Bu modeller ile hassas özelliklere göre gruplanmış veri setleri üzerinde testler yapılmış ve doğruluk,kesinlik, geri çağırma ve F1 skora göre sonuçlar paylaşılmıştır. Tabloda CelebA ve AffectNet için hassas özellik sonuçlarında farklar gözlemlenmektedir. Bıyık için CelebA veri setinde düşük doğruluk ve kesinlik sonuçları elde edilirken AffectNet veri seti ile arasında yaklaşık olarak %34'lük bir fark bulunmaktadır. Gözlük özelliği ele alındığında %7'lik bir fark gözlemlenmektedir. Aynı zamanda AffectNet daha iyi bir F1 skoruna sahiptir. Perçem özelliği için CelebA ile eğitilen ve test edilen modelin öne çıktığı görüşmektedir. Kolye hassas özelliği için AffectNet çok düşük bir doğruluk ve kesinlik değerine sahiptir. CelebA veri seti %52'lik bir fark ile daha iyi bir performans göstermektedir. Küpe için de benzer bir durum gözlemlenmiştir. %35 fark ile CelebA'nın performansı öne çıkmaktadır. Makyajda da CelebA öne çıkarken, Şapka özelliği için %24 fark ile AffectNet daha iyi sonuçlar vermektedir. Son olarak sakal için de yakın doğruluk ve kesinlik sonuçları elde edilmesine rağmen %17 fark ile AffectNet performansı geride kalmıştır. Modellerin eğitimden sonra elde edilen test sonuçları eğitim doğruluğuna yakın olsa da özellik bazlı bakıldığında büyük oranda fark görülmektedir. Amacımız bu farkların azalması dolayısıyla önyargıyı azaltarak duygu sınıflandırma modeli elde etmektir.

Ayrıştırılmış temsil yöntemi ile eğitilen modellerde CelebA ve AffectNET datasetlerinde "Gülümseme" etiketi için belirlenen her hassas özelliğin elde edilebileceği en yüksek sayı dikkate alınarak eşit dağılıma sahip olarak hazırlanan veri kümeleri kullanılarak eğitimler gerçekleştirilmiştir. Bu eğitimler sırasında aynı parametreler kullanılmıştır. Her bir hedef ve hassas özellik 0 ve 1 olarak temsil

edilecek şekilde etiket düzenlemeleri yapılmıştır. Bu deneylerde Beta-VAE için 3 farklı beta değeri (1,4,10) ve iki farklı boyut (64,128) değerlendirilmeye alınmıştır. Bu sonuçlara göre boyuta ve beta değerine göre sınıflandırıcı doğruluğu üzerinde belirgin bir fark gözlemlenmemiştir. Ancak Tablo 11 ve Tablo 14 karşılaştırıldığında iki veri seti için de beta değeri arttıkça toplam kaybın ve yeniden yapılandırma kaybının arttığı çıkarılmaktadır. Ancak AffectNet veri seti için toplam kaybın ve yeniden yapılandırma kaybının CelebA'ya göre daha fazla olduğu görülmektedir. Her iki veri seti için de KL kaybı beta değeri arttıkça azalmaktadır. CelebA sonuçlarında KL kaybı AffectNet'e göre daha yüksektir. Bu nedenle bu modellerin gizli uzayı düzenleme konusunda daha iyi performans gösterdiği söylenebilmektedir. Tablo 12 ve Tablo 15'e göre de benzer yorumları yapmak mümkündür. Beta değerleri ile toplam kayıp ve yeniden yapılandırma kaybı arasında yine doğru orantı görülmektedir. Bu boyut için de AffectNet veri setinin yeniden yapılandırma ve toplam kaybı CelebA'ya göre daha fazladır. Bu AffectNet'in daha fazla çeşitlilik içeren veri örneklerine sahip olduğu şeklinde de yorumlanabilmektedir. Son olarak CelebA veri seti ile eğitilen modellerin doğruluğu beta değerine göre yakın sonuçlar verirken AffectNet sonuçlarında beta değeri arttıkça eğitim doğruluğunun azaldığı gözlemlenmektedir. Tablo 13 ve 16'ya bakıldığında genel olarak beta değeri ve çözünürlük değişimi bıyık ve kolye gibi özelliklerde küçük farklar görülürken diğer özellikler için aynı sonuçlar elde edilmiştir. Bıyık, gözlük, kolye, şapka gibi özelliklerin tespitinde CelebA ile eğitilen modeller öne çıkmaktadır.

Tablo 8, 13 ve 16 karşılaştırıldığında Beta-VAE'nin bıyık, gözlük, şapka gibi hassas özellikleri için hedef etiketin daha yüksek performans ile tespit edildiği gözlemlenmiştir. Bu özelliklerde daha adil ve önyargısız sonuçlar elde edilebileceği söylenebilmektedir.

Kodlayıcıdan elde edilen temsil görselleştirildiğinde ilgili özelliklerin ayrıştırılma doğruluğunun göz ile kontrol edilmesi mümkün değildir. Hangi temsilin hangi özelliği temsil ettiği ancak yorumlanabilmektedir. Bu nedenle deneyler hassas özelliklerin kayıp hesaplamasına dahil edilerek temsilden çıkarılmasına yönelik tasarlanan model

kullanılarak bir dizi deney ile eğitimler ve testler gerçekleştirilmiştir. Kodlayıcı mimarisinde kullanılan Resnet18 ile elde edilen sonuçlara göre sırasıyla hassas özellik bıyık için %85, gözlük için %79, kolye için %79, küpe için %82, şapka için %88, sakal için %24, makyaj için %81 ve perçem için %80 olacak şekilde CelebA veri kümesiyle validasyon doğruluklarına ulaşılmıştır. Aynı hassas ve hedef etiket göz önüne alınarak AffectNet ile yapılan çalışmada ise bıyık için %78, gözlük için %89, kolye için %79, küpe için %82, şapka için %81, sakal için %86, makyaj için %69 ve perçem için %81 validasyon doğruluğuna ulaşılmıştır. Bu sonuçlar detayları ile Tablo 10'da paylaşılmıştır.

Yüzde belirlenen hassas özelliklerin ayrıştırılma zorluklarından dolayı özelliklere göre veri seti önişleme adımları değişiklik göstermektedir. Örneğin gözlük, sakal, bıyık, makyaj gibi özellikler doğrudan yüz üzerinde olduğundan dolayı sadece yüz bölgesi kesilmiştir. Ancak küpe, şapka, kolye gibi yüzü destekleyen bölgelerde bulunan aksesuarların tespiti için daha geniş bir kesme işlemine ihtiyaç duyulmaktadır. Bu farktan dolayı deneylerde öncelikle veri setlerinin kesilmemiş halleri kullanılarak bu aksesuar bilgilerine ulaşılmaya çalışılmıştır. Deneylerde kullanılan AffectNet veri seti doğrudan yüz sınırları dikkate alınarak kesme işleminden geçirilerek paylaşılmıştır. CelebA veri seti ise ilgili kişinin boyun, baş üstü gibi kısımları daha geniş bir açıdan aldığı için iki farklı önişleme yöntemi ile de test edilmiş kabul edilebilmektedir. Çalışmamızda bulunan kolye, şapka gibi yüz bölgesinden daha geniş kısımlarda tespit edilebilecek hassas özellikler için bu durum tartışılmıştır. Her ne kadar bu özellikler AffectNet veri setinde kısıtlı açıda bulunsa da modellerin performansı bu deneylerde daha yüksek çıkmıştır.

6. TARTIŞMA

Çalışmamızın başında oluşturduğumuz araştırma sorularımız ve hipotezlerimiz elde ettiğimiz sonuçlara göre değerlendirilmiştir. Hipotez 1'de belirtildiği üzere aksesuarların varlığı, yüz ifadesi algılama sisteminde etkisinin araştırılması için temel sınıflandırıcı model doğruluğumuz ile her hassas özelliğe sahip olan ve olmayan grup üzerinde testler gerçekleştirilmiştir. Validasyon doğruluğu yüksek görünse de hassas gruplar üzerinde değerlendirilen metrikler arasında farklar olduğu gözlemlenmiştir. Bu farklar eğitilen modelin hassas özellikler karşısında yaklaşık sonuçları vermediğini ve dolayısıyla önyargının olduğunu göstermektedir. Ele aldığımız problemde bu önyargının azaltılmasına yönelik yapılan çalışmalarla sonuçlar karşılaştırılmıştır.

Beta-VAE modeli ile özellik ayrıştırmasının model performansı üzerindeki etkisi izlenmiştir. Ayrıştırılmış temsili etkileyen beta değeri, eğitim epoch sayısı gibi parametreler bulunmaktadır. Çeşitli değerlere göre yapılan çalışmalarda belirlenen epoch sayısı için görsel boyutunun doğruluk değeri üzerinde belirgin bir fark ortaya çıkarmamış olsa da görselleştirme çalışmaları sonucunda belirlenen hassas özelliklerin beta=1'e göre beta=4 deneyinde daha doğru bir şekilde ifade edildiği görülmektedir. Beta=10 çalışmasında ise iki veri seti için de temsillerin beta 4'e göre daha genel temsil edildiği gözlemlenmiştir. Ancak hala test doğrulukları arasında belirgin farklar görülmektedir. Bu deneylerde elde edilen temsillerin hassas özellikleri içermediğinden emin olunamamaktadır ve görselleştirmeler sonucunda manuel bir şekilde tespit edilmesi mümkün değildir. Bu nedenle Hipotez 2'den yola çıkılarak hazırlanan deneylerimizde belirlenen hassas özelliklerin hedef özellik vektörüne sızması Sarhan ve arkadaşlarının [18] önerdiği yöntemin uyarlanması ile önlenmesi amaçlanmıştır.

Sırasıyla hassas özellikler için elde edilen en yüksek sonuçlar olan %88, %85, %82, %80'lik validasyon doğrulukları ile bu konuda başarı gösterdiği tespit edilmiştir.

Hipotez 3'ü araştırabilmek için çalışmamızda belirlenen üç farklı yöntem olan temel sınıflandırıcı, Beta-VAE ve ortogonal ayrıştırma yöntemleri kullanılmıştır. Bunların sonucunda çeşitli parametreler ile deneyler sürdürülmüş ve sonuçlar paylaşılmıştır. Bu sonuçlara göre kullanılan çapraz kayıp fonksiyonu ile eğitilen Beta-VAE yöntemi ile

ayrıştırılmış temsil kullanımı temel sınıflandırıcıya göre öne çıkarken en iyi performans ortogonalite hesabı dahil edildiğinde elde edilmiştir.

Çalışma sonucunda belirtilen hassas özelliklerimizin duygu tanıma sistemlerini doğrudan etkilediği gözlemlenerek, bu etkinin azaltılmasına yönelik deney sonuçları paylaşılmıştır. Elde edilen en iyi validasyon doğruluğu %88 ile şapka hassas özelliğini ayrıştırmak amacıyla CelebA veri seti kullanılarak eğitilen ortogonalite hesabı yapılan model ile elde edilmiştir.

Gözlük, kolye, küpe, bıyık gibi doğrudan yüz ve boyun bölgesinde bulunan aksesuarlar ve özelliklerin yüksek doğrulukta tespit edildiği gözlemlenirken, yoğun makyaj özelliğinin daha düşük bir doğruluk oranına sahip olduğu görülmektedir. Bu fark, belirlenen hassas özelliklerin tespit edilmesinde ve ayrıştırılmasında yaşanan zorluk derecesinden kaynaklanabilir. Yoğun makyaj, doğrudan yüz özellikleri üzerini de kapsayıcı olduğundan dolayı duygu etiketi ile ayrıştırma performansını ve dolayısıyla sınıflandırıcının validasyon doğruluğunu etkilemektedir.

7. ÇALIŞMANIN SINIRLARI VE GELECEK PLANI

Çalışmamızda elimizdeki veri kümelerinin belirttiğimiz hassas özellik dağılımına göre alt kümelere ayrılması veri sayısını önemli ölçüde etkilemektedir. Sınıfların dengeli bir şekilde oluşturulması ve modelin eğitilmesi için yapılan bölümlemede elimizde olan kısıtlı sayıda görsel kullanılmıştır. Bu da kullanılan yöntemlerdeki hedef doğruluk sonucunun beklenildiğinden daha düşük çıkmasına sebep gösterilebilmektedir. Çalışmanın ilerleyen aşamalarında odaklanılan hassas özelliklere göre örneklerin üretici ağlar tarafından arttırılarak model eğitimlerine devam edilmesi planlanmaktadır. GAN ile örneklerin çoğaltılması adımı bu çalışmamızda hesaplama kısıtlamaları değerlendirilerek dahil edilmemiştir.

Veri seti dışında model mimarisi ve parametreler de doğruluk üzerinde etkilidir. Parametreler arttıkça ve model daha karmaşık hale geldikçe hesaplama için gerekli donanım ve süre de artmaktadır. Bu yüzden elimizdeki güncel donanım ile daha karmaşık model mimarileri eklenememiştir.

Seçilen hassas özelliklerin yüz üzerinde detaylı konumlarda olduğu göze alındığında çözünürlüğün etkisi öne çıkmaktadır. Görsel boyutları en fazla 128x128 olacak şekilde planlanmıştır. İlerleyen çalışmalarda daha yüksek boyutlu ve GAN ile çözünürlük artırımı yapılarak yüksek donanımlı cihazlar ile daha yüksek boyutlu görseller ile eğitim yapılmasının sonuçları olumlu olarak etkileyeceği düşünülmektedir.

Literatürde bulunan kimlik ve duygu ayrıştırması çalışmaları incelenmiş ve kullanılan yöntemler bu çalışmada deneylere eklenmiştir. Ancak bu çalışma doğrudan kimlik bilgilerini ayrıştırmayı kapsamamaktadır. Bu sebeple çalışmanın gelecek aşamalarında kimlik bilgileri de bu çalışmaya dahil edilerek kimlik-duygu- aksesuar üçlemesi ile bir ayrıştırma çalışması yapılabilecektir. Bu anlamda görsellerden ayrıştırılan özelliklerin testi ile daha saf bir duygu tanıması modeli elde edilebileceği ön görülmektedir. Bu kapsamda yeni bir deney kümesi planlanması ve tartışılması düşünülmektedir.

AffectNet veri setinde bulunan sekiz farklı duygu kırılımında deneyler tasarlanamamıştır. Bunun sebebi seçilen hassas özellik kırılımında görsel sayısı dikkate alındığında eğitim ve test için yeterli görülmemesidir. Bu deneylerde hem

duygu sınıfları hem de hassas özellik sınıfları için dengeli veri setlerinin oluşması oldukça önemlidir. Bu nedenle çalışmanın ilerleyen aşamalarında veri artırımı yöntemleri bu veri seti için belirlenen kırılımlarda yapılarak deneyler tasarlanabilecektir.

8. KAYNAKLAR

[1] Fischer, A. (2016). Social functions of emotion. Retrieved from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=0b5caae2af7276da904e7e4ec8985544cc502327#page=228

[2] Susskind, J. M., Littlewort, G., Bartlett, M. S., Movellan, J., & Anderson, A. K. (2007). Human and computer recognition of facial expressions of emotion. *Neuropsychologia*, *45*(1), 152–162. https://doi.org/10.1016/j.neuropsychologia.2006.05.001

[3] Sha, M., Yang, W., Wei, F., Lu, Z., Chen, M., Ma, C., Zhang, L., & Shi, H. (2024). Speech Emotion Recognition Using RA-Gmlp Model on Time–Frequency Domain Features Extracted by TFCM. Electronics, 13(3), 588–588. https://doi.org/10.3390/electronics13030588

[4] Teixeira, F. L., Salviano Pinto Soares, J.L. Pio Abreu, Oliveira, P. M., & Teixeira, J. P. (2024). Comparative Analysis of Windows for Speech Emotion Recognition Using CNN. Communications in Computer and Information Science, 233–248. https://doi.org/10.1007/978-3-031-53025-8_17

[5] Patlar Akbulut, F. (2022). Hybrid deep convolutional model-based emotion recognition using multiple physiological signals. Computer Methods in Biomechanics and Biomedical Engineering, 1–13. https://doi.org/10.1080/10255842.2022.2032682

[6] Domínguez-Jiménez, J. A., Campo-Landines, K. C., Martínez-Santos, J. C., Delahoz, E. J., & Contreras-Ortiz, S. H. (2020). A machine learning model

for emotion recognition from physiological signals. Biomedical Signal Processing and Control, 55, 101646. https://doi.org/10.1016/j.bspc.2019.101646

[7] Zhang, X., Liu, J., Shen, J., Li, S., Hou, K., Hu, B., Gao, J., Zhang, T., & Hu, B. (2021). Emotion Recognition From Multimodal Physiological Signals Using a Regularized Deep Fusion of Kernel Machine. IEEE Transactions on Cybernetics, 51(9), 4386–4399. https://doi.org/10.1109/TCYB.2020.2987575

[8] Doma, V., & Pirouz, M. (2020). A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals. Journal of Big Data, 7(1). https://doi.org/10.1186/s40537-020-00289-7

[9] Awais, M., Raza, M., Singh, N., Bashir, K., Manzoor, U., Islam, S. U., & Rodrigues, J. J. P. C. (2021). LSTM-Based Emotion Detection Using Physiological Signals: IoT Framework for Healthcare and Distance Learning in COVID-19. IEEE Internet of Things Journal, 8(23), 16863–16871. https://doi.org/10.1109/JIOT.2020.3044031

[10] Singh, S., & Nasoz, F. (2020, January 1). Facial Expression Recognition with Convolutional Neural Networks. IEEE Xplore. https://doi.org/10.1109/CCWC47524.2020.9031283

[11] Zhang, X., Fu, X., Qi, G., & Zhang, N. (2023). A multi-scale feature fusion convolutional neural network for facial expression recognition. *Expert Systems*, *41*(4). https://doi.org/10.1111/exsy.13517

[12] Li, J., Jin, K., Zhou, D., Kubota, N., & Ju, Z. (2020). Attention mechanism-based CNN for facial expression recognition. Neurocomputing, 411, 340–350. https://doi.org/10.1016/j.neucom.2020.06.014

[13] Wu, Y. (2023). Facial Expression Recognition in Classroom Environment Based on Attention Mechanism. *Frontiers in Artificial Intelligence and Applications*. https://doi.org/10.3233/faia230873

[14] Abdullah, M., Ahmad, M., & Han, D. (2020, January 1). Facial Expression Recognition in Videos: An CNN-LSTM based Model for Video Classification. IEEE Xplore. https://doi.org/10.1109/ICEIC49074.2020.9051332

[15] Zhang, S., Pan, X., Cui, Y., Zhao, X., & Liu, L. (2019). Learning Affective Video Features for Facial Expression Recognition via Hybrid Deep Learning. 7, 32297–32304. https://doi.org/10.1109/access.2019.2901521

[16] Wang, X., Chen, H., Tang, S., Wu, Z., & Zhu, W. (2022). Disentangled Representation Learning. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.2211.11695

[17] Zhang, Y., Huang, J., & Yang, J. (2024). Disentangled Variational Autoencoder for Social Recommendation. *Neural Processing Letters/Neural Processing Letters*, *56*(3). https://doi.org/10.1007/s11063-024-11607-y

[18] Zhang, B., Xie, H., Gao, Z., & Wang, Y. (2024). Choose What You Need: Disentangled Representation Learning for Scene Text Recognition, Removal and Editing. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2405.04377

[19] Liang, S., Zhou, Z., Guo, Y., Gao, X., Zhang, J., & Bao, H. (2022). Facial landmark disentangled network with variational autoencoder. *Applied Mathematics/Applied Mathematics. A Journal of Chinese Universities/Gao-Xiao Yingyong Shuxue Xuebao*, *37*(2), 290–305. https://doi.org/10.1007/s11766-022-4589-0

[20] Liu, X., Sanchez, P., Thermos, S., O'Neil, A. Q., & Tsaftaris, S. A. (2022). Learning disentangled representations in the imaging domain. Medical Image Analysis, 80, 102516. https://doi.org/10.1016/j.media.2022.102516

[21] Doersch, C. (2016). Tutorial on Variational Autoencoders. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1606.05908

[22] Mitra, B., Craswell, N., & Delft, B. (2018). An Introduction to Neural Information Retrieval. *An Introduction to Neural Information Retrieval*, *xx,No.xx*, 1–18. https://doi.org/10.1561/XXXXXXXXX

[23] Burgess, C. P., Higgins, I., Pal, A., Matthey, L., Watters, N., Desjardins, G., & Lerchner, A. (2018). Understanding disentangling in $β$-VAE. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.1804.03599

[24] Zhou, L., Deng, W., & Wu, X. (2020). Unsupervised anomaly localization using VAE and beta-VAE. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2005.10686

[25] Furkan Ulger, Seniha Esen Yuksel, & Yilmaz, A. (2021). Anomaly Detection for Solder Joints Using β-VAE. *IEEE Transactions on*

Components, Packaging, and Manufacturing Technology, *11*(12), 2214–2221. https://doi.org/10.1109/tcpmt.2021.3121265

[26] Crespi, L., Loiacono, D., & Chiti, A. (2021). Chest X-Rays Image Classification from $\beta{-}$ Variational Autoencoders Latent Features. *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*. https://doi.org/10.1109/ssci50451.2021.9660190

[27] Borges, R., & Stefanidis, K. (2022). Feature-blind fairness in collaborative filtering recommender systems. *Knowledge and Information Systems*. https://doi.org/10.1007/s10115-022-01656-x

[28] Quan, T., Zhu, F., Ling, X., & Liu, Q. (2022). Learning fair representations by separating the relevance of potential information. *Information Processing and Management*, *59*(6), 103103–103103. https://doi.org/10.1016/j.ipm.2022.103103

[29] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014). Generative Adversarial Networks. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.1406.2661

[30] Zhu, Q., Gao, L., Song, H., & Mao, Q. (2021). Learning to disentangle emotion factors for facial expression recognition in the wild. *International Journal of Intelligent Systems*, *36*(6), 2511–2527. https://doi.org/10.1002/int.22391

[31] Jiang, J., & Deng, W. (2022). Disentangling Identity and Pose for Facial Expression Recognition. IEEE Transactions on Affective Computing, 1–11. https://doi.org/10.1109/taffc.2022.3197761

[32] Xu, T., White, J., Kalkan, S., & Gunes, H. (2020). Investigating Bias and Fairness in Facial Expression Recognition. Computer Vision – ECCV 2020 Workshops, 506–523. https://doi.org/10.1007/978-3-030-65414-6_35

[33] Mhd Hasan Sarhan, Nassir Navab, Eslami, A., & Shadi Albarqouni. (2020). Fairness by Learning Orthogonal Disentangled Representations. Lecture Notes in Computer Science, 746–761. https://doi.org/10.1007/978-3-030-58526-6_44

[34] Ali, K., & Hughes, C. E. (2019). Facial Expression Recognition Using Disentangled Adversarial Learning. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1909.13135

[35] Tian, H., Zhu, T., & Zhou, W. (2022). Fairness and privacy preservation for facial images: GAN-based methods. *Computers & Security*, *122*, 102902. https://doi.org/10.1016/j.cose.2022.102902

[36] Hwang, S., Park, S., Kim, D., Do, M., & Byun, H. (2020). FairFaceGAN: Fairness-aware Facial Image-to-Image Translation. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2012.00282

[37] Zhang, F., Zhang, T., Mao, Q., & Xu, C. (2018). *Joint Pose and Expression Modeling for Facial Expression Recognition*. Openaccess.thecvf.com.

https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_Joint_Pose_and_CVPR_2018_paper.html

[38] Xie, S., Hu, H., & Chen, Y. (2021). Facial Expression Recognition With Two-Branch Disentangled Generative Adversarial Network. IEEE Transactions on Circuits and Systems for Video Technology, 31(6), 2359–2371. https://doi.org/10.1109/tcsvt.2020.3024201

[39] Rifai, S., Bengio, Y., Courville, A., Vincent, P., & Mirza, M. (2012). Disentangling Factors of Variation for Facial Expression Recognition. Computer Vision – ECCV 2012, 808–822. https://doi.org/10.1007/978-3-642-33783-3_58

[40] Ali, K., & Hughes, C. E. (2021). *Facial Expression Recognition By Using a Disentangled Identity-Invariant Expression Representation*. https://doi.org/10.1109/icpr48806.2021.9412172

[41] Zhang, W., Ji, X., Chen, K., Ding, Y., & Fan, C. (2021). Learning a Facial Expression Embedding Disentangled From Identity. Retrieved June 6, 2024, from openaccess.thecvf.com website: https://openaccess.thecvf.com/content/CVPR2021/html/Zhang_Learning_a_Facial_Expression_Embedding_Disentangled_From_Identity_CVPR_2021_paper.html

[42] Bai, M., Xie, W., & Shen, L. (2019). *Disentangled Feature Based Adversarial Learning for Facial Expression Recognition*. https://doi.org/10.1109/icip.2019.8802941

[43] Susskind, J. M., Littlewort, G., Bartlett, M. S., Movellan, J., & Anderson, A. K. (2007). Human and computer recognition of facial

expressions of emotion. *Neuropsychologia*, *45*(1), 152–162. https://doi.org/10.1016/j.neuropsychologia.2006.05.001

[44] Xu, T., White, J., Kalkan, S., & Gunes, H. (2020). Investigating Bias and Fairness in Facial Expression Recognition. Computer Vision – ECCV 2020 Workshops, 506–523. https://doi.org/10.1007/978-3-030-65414-6_35

[45] weblineindia/AIML-Human-Attributes-Detection-with-Facial-Feature-Extraction. (2024, January 28). GitHub. https://github.com/weblineindia/AIML-Human-Attributes-Detection-with-Facial-Feature-Extraction

[46] Locatello, F., Abbati, G., Rainforth, T., Bauer, S., Bernhard Schölkopf, & Olivier Bachem. (2019). On the Fairness of Disentangled Representations. ArXiv (Cornell University).

[47] Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2019). AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. IEEE Transactions on Affective Computing, 10(1), 18–31. https://doi.org/10.1109/TAFFC.2017.2740923

[48] Liu, Z., Luo, P., Wang, X., & Tang, X. (2014). Deep Learning Face Attributes in the Wild. https://doi.org/10.48550/arxiv.1411.7766