T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

TÜRK İŞARET DİLİ TANIMA İÇİN ZAYIF-GÜDÜMLÜ MAKİNE ÖĞRENMESİ YÖNTEMİ

BANUÇİÇEK KANDEMİR KILINÇÇEKER
YÜKSEK LİSANS TEZİ
MATEMATİK ANABİLİM DALI

GEBZE
2022

T.C.

GEBZE TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

TÜRK İŞARET DİLİ TANIMA İÇİN ZAYIF-GÜDÜMLÜ MAKİNE ÖĞRENMESİ YÖNTEMİ

BANUÇİÇEK KANDEMİR KILINÇÇEKER

YÜKSEK LİSANS TEZİ

MATEMATİK ANABİLİM DALI

DANIŞMANI

DOÇ.DR. NURİ ÇELİK

II. DANIŞMAN

DR. ÖĞR. ÜYESİ YAKUP GENÇ

GEBZE

2022

T.R.

GEBZE TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

WEAKLY-SUPERVISED MACHINE LEARNING METHOD FOR TURKISH SIGN LANGUAGE RECOGNITION

BANUÇİÇEK KANDEMİR KILINÇÇEKER

A THESIS SUBMITTED FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS

THESIS SUPERVISOR

ASSOC.PROF.DR. NURİ ÇELİK

II. THESIS SUPERVISOR

ASSIST. PROF. DR. YAKUP GENÇ

GEBZE

2022

GEBZE TEKNİK ÜNİVERSİTESİ	**YÜKSEK LİSANS JÜRİ ONAY FORMU**

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 26/05/2022 tarih ve 2022/26 sayılı kararıyla oluşturulan jüri tarafından 15/06/2022 tarihinde tez savunma sınavı yapılan Banuçiçek Kandemir Kılınççeker 'in tez çalışması Matematik Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE
(TEZ DANIŞMANI) : Doç. Dr. Nuri ÇELİK

ÜYE
(TEZ DANIŞMANI) : Dr. Öğr. Üyesi Yakup GENÇ

ÜYE : Prof. Dr. Yusuf Sinan ALYUL

ÜYE : Dr.Öğr.Üyesi Enis KARAARSLAN

ÜYE : Dr. Öğr. Üyesi Hadi Alizadeh

ONAY

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
..…/……/…… tarih ve ……/…..... sayılı kararı.

ÖZET

Türk İşaret Dili'nin (TİD) tanınması bir bilgisayarlı görme kapsamında bir yapay zekâ problemidir. Bu problemin çözülebilmesi için Yapay Sinir Ağları (YSA) kullanılmaktadır. Bu tez kapsamında TİD'in tanınmasına yönelik YSA modeli öne sürülmektedir. Bu sınıflandırma modelinin eğitilmesi için, TİD'e özgü bir veri kümesi gereklidir. Bu veri kümesinin tüm varyasyonları içermesi gerektirmektedir. Çözülmesi gereken bu problem için gerekli olan veri kümesi birlikte düşünüldüğünde bunun pahalı bir problem olduğu aşikardır.

Bu tez kapsamında zayıf güdümlü bir makine öğrenmesi yöntemiyle veri kümesini oluşturma kısmını kolaylaştırıp modelleme yapılması hedeflenmektedir. Bu bağlamda tez 2 aşamadan oluşmaktadır. İlk olarak TİD için bir veri kümesi oluşturulmuştur. Bu veri kümesi 512 adet farklı kelime sınıfından oluşturulmuştur. Veri kümesi oluşturulurken kaynak olarak çevrim içi platformlar kullanılmıştır. Bu aşamadan sonra frekans baz alınarak farklı sınıf sayıları için dengeli ve dengesiz olmak üzere toplam 6 adet alt veri kümesi oluşturulmuştur. Oluşturulan bu veri kümelerinin sınıf sayıları sırasıyla 5, 10 ve 15'tir.

Veri kümesi oluşturulduktan sonra eğitim aşamasına geçilmiştir. Eğitim aşamasında iki farklı model kullanılmıştır. Elde edilen bu yeni veri kümesiyle basit yapılı bir UKSB tabanlı model inşa edilerek tekrar eğitim gerçekleştirilmiştir. İkinci eğitim aşamasının sonunda en yüksek doğruluğa ve en küçük kayıp değerine sahip model ağı seçilerek tahminleme yapılmıştır.

Yukarıda bahsi geçen süreçlerin ve tahmin yapma sürecinin tamamı oluşturulan 6 ayrı alt veri kümesi için tekrarlanmıştır. Eğitim süreçlerinin sonunda her iki tarz veri tipi için de başarı oranı çoğunlukla %90'ın üzerinde olsa da dengeli veri kümelerinde daha iyi sonuçlar gözlemlenmiştir. Sınıf sayısı arttırıldıkça başarı oranının arttığı gözlemlenmiştir.

Anahtar Kelimeler: Türk İşaret Dili (TİD), Derin Öğrenme, Bilgisayarlı Görme, Evrişimsel Sinir Ağları (ESA), Uzun Kısa Süreli Bellek (UKSB).

SUMMARY

Turkish Sign Language (TİD) recognition is an artificial intelligence problem within the scope of computer vision. Artificial Neural Networks (ANN) are used to solve this problem. In this thesis, the YSA model for the classification of TID is proposed. A TID-specific dataset is required to train this classification model. This dataset requires to contain all variations. Considering the dataset required for this problem to be solved, it is obvious that this is an expensive problem.

In this thesis, it is aimed to facilitate the process of obtaining the dataset with a semi-supervised machine learning method and then modeling. In this context, the thesis consists of two stages. First, it acquires a dataset, which contains 512 different word classes. It uses online platforms as a source. After this stage, a total of 6 sub-datasets, balanced and unbalanced, are formed for different class numbers based on frequency. The class numbers of these datasets are 5, 10, and 15, respectively.

Then, the training phase is initiated. Two different models are used in the training phase. With this new dataset, a simple LSTM-based model is built, and retraining is carried out. At the end of the second training phase, the model network with the highest accuracy and the smallest loss value is selected and a prediction is made.

At the end of the training process, although the success rate for both types of data is mostly above 90%, better results are observed in balanced datasets. It is observed that the success rate increased as the number of classes increased.

Keywords: Turkish Sign Language (TSL), Deep Learning, Computer Vision, Convolution Neural Network (CNN), Long Short-Term Memory (LSTM).

TEŞEKKÜR

Başta süreç boyunca bana yol gösteren desteğini hiç esirgemeyen bana akademik yolda yürüme şevki kazandıran danışman hocam Sayın Doç.Dr.Nuri ÇELİK , bilim insanı kişiliğinden, fikirlerinden çok şey öğrendiğim, öğrencisi olmaktan her zaman gurur duyacağım diğer danışman hocam Dr.Öğr.Üyesi Yakup GENÇ'e , Muğla Sıtkı Koçman Üniversitesi Bilgisayar Mühendisliği Bölümü'nde Dr.Öğr.Üyesi Enis KARAARSLAN'a laboratuvarında bulunan cihazları kullanmam doğrultusunda, tez çalışmalarım kapsamında verdiği destek için, süreç boyunca sabırla ve hevesle bana destek veren Sayın Beyza TÜRKMEN'e en içten teşekkürlerimi sunarım.

Bana her zaman inanan ve güvenen, varlığını her zaman hissettiğim Merhume Sayın Prof.Dr.Essin TURHAN'a, yoğun çalışmalarım sırasında desteğini esirgemeyen, bana olan inancını hiç kaybetmeyen ve beni her zaman cesaretlendiren hayat arkadaşım Onur KILINÇÇEKER'e teşekkürü borç bilirim.

İÇİNDEKİLER

	Sayfa
ÖZET	v
SUMMARY	vi
İÇİNDEKİLER	vii
SİMGELER VE KISALTMALAR DİZİNİ	x
ŞEKİLLER DİZİNİ	xi
TABLOLAR DİZİNİ	xiii
1. GİRİŞ	1
2.PROBLEMİN TANIMI VE İLGİLİ ÇALIŞMALAR	4
2.1 Genel Bilgiler	4
2.2 Türk İşaret Dili	5
2.3 Problemin Tanımı	6
2.4 İlgili Çalışmalar	7
3. TÜRKÇE İŞARET DİLİ İÇİN ZAYIF GÜDÜMLÜ MAKİNA ÖĞRENMESİ	12
3.1. Veri Kümesi	12
3.1.1. Girdilerin Elde Edilmesi	14
3.1.2 Girdilerin Düzenlenmesi	15
3.2 Öznitelik Çıkarımı	19
3.2.1 P1 ve P2 İçin Oluşturulan Veri Kümesi ve Öznitelik Çıkarımı	21
3.2.2 P3 ve P4 İçin Oluşturulan Veri Kümeleri ve Öznitelik Çıkarımı	23
3.2.3 P5 ve P6 İçin Oluşturulan Veri Kümesi ve Öznitelik Çıkarımı	26
3.3 Uzun-Kısa Süreli Bellek ile Eğitim	29
3.3.1 P1 ve P2 İçin Uksb Eğitimi	30
3.3.2 P3 ve P4 İçin Uksb Eğitimi	32
3.3.3 P5 ve P6 İçin Uksb Eğitimi	35

4. DENEYSEL ÇALIŞMALAR 38

4.1 P1 ve P2 İçin Karmaşıklık Matrisi 38

4.2 P3 ve P4 İçin Karmaşıklık Matrisi 40

4.3 P5 ve P6 için Karmaşıklık Matrisi 41

5.SONUÇ ve GELECEK ÇALIŞMALAR 44

REFERANSLAR 46

ÖZGEÇMİŞ 49

SİMGELER VE KISALTMALAR DİZİNİ

Simgeler ve Kısaltmalar		**Açıklamalar**
ANN	:	Artificial Neural Network
CNN	:	Convolutional Neural Network
DANN	:	Deep Artificial Neural Network
ESA	:	Evrişimsel Sinir Ağı
LSTM	:	Long-Short-Term Memory
TİD	:	Türk İşaret Dili
TRT	:	Türkiye Radyo Televizyon Kurumu
TSL	:	Turkish Sign Language
UKSB	:	Uzun-Kısa Süreli Bellek
YSA	:	Yapay Sinir Ağı

ŞEKİLLER DİZİNİ

Şekil No:		Sayfa
3.1:	Öne sürülen yaklaşımın adımları 3 aşamada tanımlanmıştır.	13
3.2:	Videolardan elde edilen görüntünün işlenmemiş hali.	14
3.3:	Videolardan elde edilen görüntünün işlenmiş hali.	15
3.4:	PyTranscriber [28] uygulamasının ara yüzü.	16
3.5:	PyTranscriber'dan alınan "srt" uzantılı dosya örneği.	17
3.6:	Google Speech to Text [30] arayüzü şekli.	18
3.7:	P1 için ESA eğitimi kayıp değeri grafiği.	21
3.8:	P1 için ESA eğitimi doğruluk fonksiyonu grafiği.	22
3.9:	P2 için frekans grafiği.	22
3.10:	P2 için ESA eğitimi doğruluk fonksiyonu grafiği.	23
3.11:	P2 için ESA eğitimi kayıp değeri grafiği.	23
3.12:	P3 için ESA eğitimi doğruluk fonksiyonu grafiği.	24
3.13:	P4 için ESA eğitimi kayıp değeri grafiği.	25
3.14:	P4 için frekans grafiği.	25
3.15:	P4 için ESA eğitimi doğruluk fonksiyonu grafiği.	26
3.16:	P4 için ESA eğitimi kayıp değeri grafiği.	26
3.17:	P5 için ESA eğitimi doğruluk fonksiyonu grafiği.	27
3.18:	P5 için ESA eğitimi kayıp değeri grafiği.	27
3.19:	P6 için frekans grafiği.	28
3.20:	P6 için ESA eğitimi doğruluk fonksiyonu grafiği.	28
3.21:	P6 için ESA eğitimi kayıp değeri grafiği.	29
3.22:	P1 için ESA eğitimi doğruluk fonksiyonu grafiği.	30
3.23:	P1 için ESA eğitimi kayıp değeri grafiği.	31
3.24:	P2 için ESA eğitimi doğruluk fonksiyonu grafiği.	32
3.25:	P2 için ESA eğitimi kayıp değeri grafiği.	32
3.26:	P3 için ESA eğitimi doğruluk fonksiyonu grafiği.	33
3.27:	P3 için ESA eğitimi kayıp değeri grafiği.	33
3.28:	P4 için ESA eğitimi kayıp değeri grafiği.	34
3.29:	P4 için ESA eğitimi doğruluk fonksiyonu grafiği.	35
3.30:	P5 için ESA eğitimi doğruluk fonksiyonu grafiği.	36

3.31: P5 için ESA eğitimi kayıp değeri grafiği. 36
3.32: P6 için ESA eğitimi doğruluk fonksiyonu grafiği. 37
3.33: P6 için ESA eğitimi kayıp değeri grafiği. 37
4.1: P1 için karmaşıklık matrisi. 39
4.2: P2 için karmaşıklık matrisi. 39
4.3: P3 için karmaşıklık matrisi. 40
4.4: P4 için karmaşıklık matrisi. 41
4.5: P5 için karmaşıklık matrisi. 42
4.6: P6 için karmaşıklık matrisi. 43

TABLOLAR DİZİNİ

Tablo No:		**Sayfa**
3.1:	ESA modelinin yapısı.	20
3.2:	Tez kapsamında ele alınan 6 adet problem.	20

1. GİRİŞ

Tüm canlıların birbiriyle iletişim kurması yaşamlarını sürdürebilmeleri açısından hayati önem taşır. Canlılar birbirileriyle iletişim kurarken ihtiyaçlarına göre, farklı yöntemler kullanmaktadır. Bu yöntemler sözlü, yazılı ya da hareketlerle olabilmektedir. İşitme kaybı olmayan insanlar birbiriyle iletişim kurabilmek için dil yetenekleri kullanırlar. İşitme problemi bulunan bireylerin dil yetenekleri zayıf olabilmektedir. Bazı durumlarda zayıf olmanın yanı sıra kullanılamaz hale gelebilmektedir çünkü işitme kaybı zamanla insanın konuşma yeteneğini de kullanılamaz kılmaktadır. Bu durumda işitme kaybı yaşayan bireyler birbirleriyle ve işitme kaybı olmayan bireylerle anlaşabilmek işaret dili kullanmaktadır. İşaret dili ifade edilmek istenen anlamı karşı tarafa iletebilmek için kullanılan görsel bir dildir [1]. Dolayısıyla ifade edilmek istenen bir anlamı karşı tarafa iletirken görsel kanal kullanılacaksa, bunun için işaret dillerine ihtiyaç duyulur. Dünyada küresel bir işaret dili mevcut değildir. Bu sebeple işaret dilleri ile ilgili yapılmış ve yapılacak olan tüm çalışmalar, o dilin kullanıldığı bölgeye özgüdür. Ayrıca bir bölgede kullanılan işaret dili o bölgede kullanılan dilin gramatikal yapısıyla birebir ilişkili değildir. İşaret dillerinde her hareket ya bir harfe ya da bir kelimeye karşılık gelmektedir. Dolayısıyla işaret dilleri ifade ediliş yöntemleri olarak ele alındığında kelime tabanlı ve harf tabanlıdır. Harf tabanlı ifade ediliş yönteminde her hareket bir harfe karşılık gelmektedir. İfade edilen bu harfler birleştirilerek kelimeleri meydana getirmektedir. Buna parmak abecesi de denmektedir [1]. Kelime tabanlı ifade ediliş yöntemi durağan değil dinamiktir. Yani sadece el hareketlerinin anlık olarak işaret oluşturmasıyla değil dinamik olarak, bir hareket akışıyla bir kelime ifade edilmektedir. Bu bağlamda işaretçinin sadece el hareketleri değil bununla birlikte mimikleri, vücut hareketleri de önem arz etmektedir.

Türk İşaret Dili, Türkiye Cumhuriyeti ve Kuzey Kıbrıs Türk Cumhuriyeti'nde kullanılan işaret dilidir [2]. Türkiye'de nüfusun yaklaşık olarak %1,1'si işitme engelli bireylerden oluşmaktadır [3]. Bu çalışmanın nihai amacı işitme engelli bireylerin, iletişim kurabilmek için kullandıkları işaret dilinin, bu dili kullanmayı bilmeyen diğer insanlar tarafından da anlık olarak anlaşılmasını sağlamaktır.

TİD'in tanınması bilgisayarlı görme problemidir. Son yıllarda bu tarz bilgisayarlı görme problemlerinin çözümü için YSA kullanılmaktadır. Bu problemin

çözümü için işaret hareketinin yakalanması gerekir. Bunun yapılabilmesi için iki yöntem mevcuttur. Bunlardan ilki için ek bir donanıma ihtiyaç duyulmaktadır [4]. İşaretçinin işaretlerini takip edebilmek için bir eldiven kullanılması gerekmektedir. Diğer yöntem kendi içerisinde 2'ye ayrılır. Eğer tanınacak olan kelime statik ise, statik hareketlerin tespiti için 2 boyutlu görüntü kümesi kullanılmalıdır. Eğer kelimeler dinamik ise işaretlerin buna uygun olarak yakalanması için kamera kullanımı gerekmektedir. Bu bağlamda TİD'in tanınabilmesi için öncelikle bir içinde TİD'i ihtiva eden bir veri kümesine ihtiyaç vardır. Bu veri kümesinin bütün çeşitliliği içermesi gerekmektedir. Tüm bu açılardan ele alındığında TİD sınıflandırması oldukça maliyetli bir problemdir.

Bu tez kapsamında amaç zayıf güdümlü bir yöntem kullanarak, problemin veri kümesi oluşturma kısmını kolaylaştırıp akabinde elde edilen veri kümesiyle modelleme yapmaktır. Bu yöntemle veri kümesi algoritmik olarak toplanıp işlemeye hazır hale getirilmektedir. Böylelikle Derin Öğrenme modelleri için gerekli olan bir adımın maliyeti ciddi bir şekilde azaltılmıştır. Veri kümesini bu şekilde oluşturulması ekonomik olmasının yanı sıra başka problemler için de uygulanabilir olması açısından büyük önem taşımaktadır.

TİD'in tanınmasına yönelik YSA modeli öne sürülmektedir. Tez kapsamında çözülmeye çalışılan problemle ilgili standart bir veri kümesi bulunmamaktadır. Dolayısıyla sınıflandırma modelinin eğitilmesi için, TİD'e özgü bir veri kümesi oluşturulmuştur. Bu veri kümesi çevrim içi platformlardan elde edilmiştir. Böylelikle gerek görüldüğü takdirde, veri kümesi dinamik olarak güncellenebilir. Oluşturulan veri kümesi kelime tabanlıdır. Veri kümesinin tamamı RGB görüntülerden oluşmaktadır. Modelin güncel hayata uyumluluğu açısından, çevrim içi platformlardan elde edilecek olan görüntülerin kullanılması çalışmayı özgün kılmaktadır. Çevrim içi platformların dinamikliği modelde kullanılacak olan veri kümesinin içeriğini çeşitlendirmektedir. Böylelikle düşük maliyetle yüksek başarı oranı elde edilmiştir.

Bunun için hem çevrim içi platformlardan içerisinde TİD mevcut olan videolar kullanılarak bu videoların Türkçe altyazı dosyaları elde edilmiştir. Bu tez çalışmasında ilk olarak TİD'e özgü, kapsamlı ve güncellenebilir bir veri kümesinin oluşturulmuştur. TİD'e ait kelimelerin tanınması ve tanınması için yapay zekâ ve makina öğrenmesi yöntemlerinden biri olan derin öğrenme yöntemi kullanılmıştır. ESA ve UKSB yöntemleri kullanarak zaman ve donanım açısından düşük maliyetli ve başarı oranı yüksek bir model kurulmuştur. ESA'nın ardından UKSB yönteminin kullanılmasının

en önemli sebebi oluşturulan veri kümesinin harf tabanlı değil kelime tabanlı olmasıdır. Yani [1] deki parmak abecesi kullanılarak ifade edilen kelimelere veri kümesinde yer verilmemiştir. Çalışma boyunca Python programlama dili kullanılmıştır. Çalışma 3 adımda ele alınmıştır. İlk olarak video görüntüleri ve bu görüntülerin altyazı dosyaları elde edilmiştir. İkinci adım olarak elde edilen bu görüntüler düzenlenerek model tarafından eğitilmeye hazır hale getirilmiştir.

Son aşamada hazırlanan veri kümesi ESA ve UKSB yöntemleri kullanılarak eğitime tabii tutulmuş ve tahminleme işlemi yapılmıştır. Tez çalışmasında hazırlanan veri kümesi hem bu çalışmaya hem de bu alanda yapılacak olan gelecek çalışmalara katkı sağlayacak niteliktedir. Çalışma sosyal açıdan, Türkiye nüfusunun yaklaşık %1,1'ini oluşturan işitme engelli bireylerle, TİD'i bilmeyen bireyler arasındaki iletişimin daha sağlıklı hale gelmesine katkı sağlayacaktır. Bilimsel açıdan, elde edilen TİD'e ait veri kümesi birçok ulusal ve uluslararası bilimsel çalışmaya ve projeye zemin hazırlamaktadır.

Bu tez çalışması 5 bölümden oluşmaktadır. İlk bölümde tez çalışmasının amacını ve özgün yanlarına yer verilmiştir. İkinci bölümde bu konuyla ilgili önceden yapılmış çalışmalar ve problemin tanımına yer almaktadır. Üçüncü bölümde oluşturulan veri kümesi, bu veri kümesiyle çözülen problemlerin öznitelik çıkarımına ve sınıflandırma işlemine yer verilmiştir. Dördüncü bölümde deneysel çalışmalara ve beşinci bölümde elde edilen sonuçlara ve gelecek çalışmalar hakkında bilgi verilmiştir.

2.PROBLEMİN TANIMI VE İLGİLİ ÇALIŞMALAR

2.1. Genel Bilgiler

Tüm veriyi doğru şekilde etiketlenmesine ihtiyaç duyulan denetimli öğrenmeden farklı olarak, zayıf güdümlü öğrenmede kesin etiketler olmadan yaklaşık ve kabaca etiketlenmiş veri aracılığıyla eğitim gerçekleştirilir [5]. Bilgisayarlı görü, videolardan ve dijital görüntülerden anlamlı bilgiler çıkarmaya çalışan bir bilim dalıdır. Bilgisayarlı görünün alt dallarını oluşturan çeşitli görevleri ve uygulamaları mevcuttur. Bunlardan bazıları hareket analizi, hareket tanıma, görüntü onarımı ve sahne yapılandırmasıdır [6]. Görüntü işleme ise dijital ortama alınan bir görüntünün nümerik hale getirilip üzerinde manuel ya da otomatik olarak bazı işlemler yapılması işlemidir. Bilgisayarlı görüyle ilgili problemlerin çözümünde ve görüntü işleme işleminde kullanılan veri görüntülerdir. Görüntünün işlenebilmesi için farklı formları vardır. Bazı durumlarda tekli görüntüler veri olarak kullanılırken bazen hareketli görsellerin kaydedilmesini sağlayan, sayısal olarak derlenmiş hareketli görüntüler dizisi olan videolar kullanılmaktadır. Tekli çerçevelerin kullanıldığı makine öğrenmesi problemlerinde, klasik bilgisayarla görme yöntemlerinin yanında derin öğrenme metotları kullanılmaya başlanmıştır. Tek bir görüntüden hareketi tanımak bilgisayarlı görme ve derin öğrenme çalışma alanlarının altında bir problemdir. Bu bağlamda problemi çözmek için temel yaklaşım ESA kullanılmasıdır. Hali hazırda kullanılan birçok ESA modeli mevcuttur [7]. Hareket tanıma probleminde temel yaklaşım çerçeve çerçeve görüntüleri işleyip birleştirme olabilir. Bunun yanı sıra belirli bir video klipe bakılarak hareket tanıma problemini çözen yaklaşımlarda mevcuttur. Bu gibi yaklaşım farklılıklarından kaynaklı farklı çözüm yöntemleri de bulunmaktadır. Yinelemeli sinir ağı mimarisi olan UKSB yöntemi bir sekanstan hareket tanıma probleminin çözümünde ESA'larla birlikte kullanılmaktadır.

2.2. Türk İşaret Dili

İşaret dilleri ifade edilmek istenen anlamı karşı tarafa iletebilmek için el, yüz ve beden hareketlerinin kullanıldığı bir iletişim yöntemidir. Global bir işaret dili mevcut değildir. Dillerin kendine özgü işaret dili bulunur [8]. Buna bağlı olarak her işaret dili kendine özgü bir dilbilgisi yapısına sahiptir. Türkiye ve Kuzey Kıbrıs Türk Cumhuriyeti'nde işitme engelli bireyler tarafından kullanılan dil TİD'dir. TİD'de toplam 29 harf bulunmaktadır. Türk Dil Kurumu'na ait TİD sözlüğünde 1986 kelime ve deyim yer almaktadır. Millî Eğitim Bakanlığı tarafından 2015 yılında yayınlanan TİD bulunmaktadır [9]. Bunun dışında bu sözlüklerin birleşimiyle oluşturulan çeşitli çevrim içi TİD sözlüğü platformları da bulunmaktadır. Bunlar İşaret Dili Sözlüğü [10] ve "İşaretçe" isimli TİD sözlüğüdür [11]. TİD içerisinde 2 farklı şekilde ifade edilebilen kelime ve ifade yapısı bulunur. Bunlardan ilki harf tabanlı kelime yapısı diğeri ise kelime tabanlıdır. Harf tabanlı kelime yapısında her kelime harf olarak hecelenir ve bu kelime yapısı statiktir. Kelime tabanlı ifade şeklinde hareket dinamiktir. Bir sekans oluşturur. TİD'de ağırlıklı olarak dinamik tabanlı kelimeler mevcuttur.

Türkiye'de nüfusun yaklaşık %1,1'i işitme engelli bireylerden oluşmaktadır [3]. Bu bireylerin, topluluğun tüm üyeleriyle düzgün ve doğru bir şekilde iletişim kurabilmesi sosyolojik olarak büyük önem taşımaktadır. Ancak işitme engelli bireylerin kullandığı işaret dili, uzmanlar dışında toplumda pek bilinmemektedir. Bu, engelli bireylerin topluma entegrasyonu için büyük bir sorun teşkil etmektedir. Bunu önlemek için Türkiye Radyo Televizyon Kurumu (TRT) ve bazı televizyon kanalları işitme engelli kişilere haber hizmeti vermektedir. Ancak toplum üyeleri arasında iletişimi sağlayacak etkin bir çözüm Türkiye'de ve diğer ülkelerde sunulmamaktadır. Bu sorunu çözmek için makina öğrenmesi yöntemlerini kullanarak önerilen bilimsel çalışmalar gelecek için umut vadetmektedir. Ancak bu çalışmalar işaret dili açısından harflerin tespiti ve dönüşümü ile sınırlıdır. Bu açıdan insanlar arası iletişimi sağlamak için etkili bir yöntem önermek büyük önem taşımaktadır.

2.3. Problemin Tanımı

TİD'in sınıflandırması bilgisayarlı görme kapsamında ele alınan bir problemdir. Bu problemi çözmek için temel yaklaşımlar mevcuttur. Bu yaklaşımlar hem veri kümesinin oluşturulması aşamasında hem de modelleme aşamasında farklılık göstermektedir. Modelleme aşamasına geçilebilmesi için bir veri kümesine ihtiyaç duyulmaktadır. Bu veri kümesinin elde edilmesinde kullanılan farklı metotlar bulunmaktadır. Bazı çalışmalarda veri kümesinin oluşturulması için ekstra donanıma ihtiyaç duyulmaktadır. Bu donanım bazen işareti belirleyebilmek için harici bir eldiven bazen de işaretçiyi dijital bir ortamda kayıt altına almak için bir kayıt cihazı olabilmektedir [22]. Bunun dışında işaretçilerin ve ortamın birbirinden farklılık göstermesi modelin başarısı açısından önem arz etmektedir.

Bu tezin nihai amacı zayıf güdümlü bir yöntem kullanarak günümüzde işitme engeli olan bireylerin, kendi aralarında anlaşabilmek için kullandıkları işaret dilinin, diğer insanlar tarafından anlık olarak anlaşılmasını sağlamaktır. Bu kapsamda, TİD içeren videodan anlık olarak, altyazı metinleri üretilerek, işitme engeli olan insanlarla duyabilen insanların arasındaki iletişim engelini azaltmak amaçlanmaktadır. Elde edilen alt yazı metinlerinde geçen kelimeler videolardan elde edilen çerçevelerle eşleştirilmeye çalışılmaktadır. Bu eşleştirme işleminde oluşan zaman kaymalarını önlemek için elde edilen zaman bilgisinden yola çıkılarak, kelimeye karşılık gelen çerçeveler belirlenmiştir. Ancak yakalanan bu kelime çerçeveleri, kelimenin geçtiği yerde doğrudan geçmemiş olabilir. Belirli bir çerçeve öncesinde veya sonrasında geçmiş olabilir. Bu gibi durumlardan kaynaklanan hataları tolere edebilmek için kelimeler ortalama olarak 1 saniye (25 çerçeve) olacak şekilde belirlenmiştir.

Bu tezin hedefleri arasında TİD'e ait kapsamlı ve kullanılabilir bir veri kümesinin oluşturulması ve farklı derin öğrenme metotları kullanılarak modelin başarı oranını arttırılması, bunun için ESA ve UKSB yöntemleri uygulanması bulunmaktadır. Bu veri kümesi oluşturulurken zayıf güdümlü bir yöntem kullanılarak modelleme için gerekli olan bu kısmın maliyeti azaltılması hedeflenmiştir.

2.4. İlgili Çalışmalar

Tez çalışması kapsamında TİD'in tanıması için bir yöntem öne sürülmektedir. Çalışmada TİD'e özgü bir veri kümesi oluşturulmuştur. Bu veri kümesi kullanılarak TİD'e ait kelimeler tanınarak bir YSA modeli geliştirilmiştir. Bu model sayesinde günlük hayattan alınan TİD'e ait kelimeler otomatik olarak tanınabilecektir. Bu sayede TİD'i konuşan ve konuşamayan bireyler arasındaki iletişim sorununun çözülmesi hedeflenmektedir.

Dünyada global bir işaret dili mevcut değildir. Her dilin kendine özgü işaret dili bulunur. Böylece bir işaret dili için elde edilen veri kümesi ve bu veri kümesi için kullanılan yöntem kendine özgüdür. Lokasyonlara göre değişkenlik gösteren işaret dili sınıflandırma problemi için hali hazırda mevcut olan veri kümeleri mevcuttur. Bunların en kapsamlılarından biri American Sign Language [12]dir. Bundan dolayı TİD'i tanıma için mevcut, güncel literatür aşağıda verilmektedir. Yalçınkaya vd. [13], "Hareket Geçmişi Görüntüsü" yöntemi kullanılarak TİD'i tanıma için bir yaklaşım öne sürmüşlerdir. Bu yaklaşımda "En Yakın Komşuluk" algoritması sınıflandırma işlemi için kullanılmıştır. Öne sürülen yaklaşım ile %95 oranında başarı elde edilmiştir. TİD'deki veri eksikliğinden dolayı Yalçınkaya vd. [13], çalışmalarında eğitim ve test aşamasında kendi oluşturdukları veri kümesini kullanmışlardır. Bu veri seti sadece "acımak", "akşam"," arkadaş"," bateri"," direksiyon"," fayda"," makas" ve "sevmek "kelimeleri için sınıflandırma işlemi yapmışlardır. Bu çalışmanın, TİD'in 2000 sözcük ve kavramdan oluştuğu göz önünde bulundurulduğunda çok dar kapsamda olduğu görülmektedir [14].

Kın [15], çalışmasında TİD tanıma için halihazırda eğitilmiş bir ESA modeli kullanılarak transfer öğrenme metodu kullanmıştır. Çalışmada 29 harf ve 3 karakterden oluşan, toplamda 32 sınıflı bir veri seti eğitilmiştir. Her bir sınıf için 1500 adet görüntü mevcut olduğu için derin öğrenme metodu kullanılabilmiştir. Böylece %90 başarı oranı elde edilmiştir. [13] de verilen çalışmaya kıyasla, bu çalışmada sınıf sayısında ciddi bir artış mevcuttur. Fakat TİD'in günlük kullanımında, yalnızca güncel TİD sözlüğünde mevcut olmayan kavramlar ve özel kelimeler için harflerle iletişim kurulduğundan, bu çalışma günlük hayatta uygulanabilirliği açısından zayıftır.

Beşer vd. [16], TİD'e ait tüm rakamların tanınması için kapsül ağı kullanarak bir yöntem öne sürmüşlerdir. 218 farklı katılımcıdan 10'ar adet örnek alarak veri kümesi

oluşturulmuştur. Literatürde oldukça yeni olan kapsül ağlarının kullanımı çalışmayı özgün yapmaktadır. Ancak benzer şekilde sadece rakamları kullanarak yapılan sınıflandırma TİD'in günlük hayattaki kullanımı düşünüldüğünde, uygulama ve kullanılabilirlik açısından zayıftır.

Haberdar vd. [17], Türkçe İşaret Dili tanımayı sağlayan video tabanlı bir sistemi öne sürmüşlerdir. Bu kapsamda Saklı Markov Modellerini kullanarak %95,7 başarı oranı elde edilmiştir. 22 tek el 28 çift el olmak üzere toplam 50 ifadeden oluşan bir veri kümesi hazırlamışlardır. Bu 50 ifade için toplam 750 örnek kullanılmıştır. Bunun 500 adedi eğitim, 250 adedi test aşamasında kullanılmıştır. 50 ifade için 750 adet veri sayısı oldukça azdır. Bu kadar az örnekleme rağmen eğitim süresinin çok zaman aldığı belirtilmiştir.

Demircioğlu vd. [18] çalışmalarında TİD'in gerçek zamanlı sınıflandırılmasıyla ilgili bir sistem önermişlerdir. Çalışmada Leap Motion cihazı kullanılarak temel el hareketleri elde edilmiştir. Bu elde edilen hareketler 18 sınıfta kümelenmiştir. Bu veri seti çevrim dışı olarak Rastgele Orman (Random Forest) ve Çok Katmanlı Algılayıcı (Multi Layer Perceptron) yöntemleri kullanılarak eğitim ve test işlemi gerçekleştirilmiştir. Elde edilen sonuçlar kıyaslamalı olarak sunulmuştur. Rastgele Orman metodunun başarı oranı %93,59 iken Çok Katmanlı Algılayıcı metodunun başarı oranı %96,67 olarak elde edilmiştir. Başarı oranları oldukça yüksek olsa da TİD'in kelime çeşitliliği düşünüldüğünde, 18 sınıftan oluşan veri seti bir dezavantaj olarak görülebilir.

Memiş ve Albayrak [19] Kinect Duyarga cihazı kullanarak elde ettikleri TİD verisinin sınıflandırılması için bir yöntem öne sürmüşlerdir. Bu yöntemi Manhattan Uzaklığı metriği ile K- En Yakın Komşu sınıflandırıcı kullanılarak oluşturmuşlardır. Çalışma sonucunda %90 başarı oranı elde edilmiştir. Bu başarı oranını etkileyen en önemli faktör veriyi elde ettikten sonra uygulanan görüntü işleme yöntemleridir. TİD'e ait 3 ayrı kategoriden oluşan 111 kelimelik, toplamda 1002 video görüntüsünden meydana gelen bir veri kümesi kullanılmıştır.

İncelenen çalışmalarda tespit edildiği üzere TİD verisi çok büyük öneme sahiptir. Ayrıca oluşturulan bu veri setinin herkesin kullanımına açık olması bu konuyla ilgili çalışmaların devamlılığı için fayda sağlayacaktır. Bu kapsamda, Camgöz vd. [20] tarafından oluşturulan veri kümesi bu ihtiyacı giderecektir. Bu veri kümesi toplam 3 alandan elde edilmiştir. Bu alanlardan birincisi sağlık ile ilgili olup 496 ifadeyi ihtiva etmektedir. İkincisi bankacılık ve finans alanında olup 171 ifadeyi

içermektedir. Üçüncüsü ise günlük hayatta sıklıkla kullanılan ifadelerle ilgili olup 188 ifadeyi içermektedir. Bu veri seti Microsoft Kinect v2 cihazı kullanılarak elde edilmiştir. Veri tabanı toplamda iki milyona yakın kelime içermektedir.

BosphorusSign22k olarak adlandırılan bu veri setini kullanan, Kındırıoğlu vd. [21] %81,37, Özdemir vd. [22] %88,53 ve Gökçe vd. [23] %94,94 başarı oranı elde etmiştir.

Sincan ve Keleş [24], TİD için 43 farklı kişiden, 226 farklı işaret için toplam 38,336 video görüntüsü oluşturmuşlardır. Bu veri seti Microsoft Kinect v2 ile elde etmişlerdir. Görüntüler hem iç hem dış ortamda kaydedilmiş olup, çok çeşitli arka planlar içermektedir. Bunun dışında görüntüler elde edilirken mekânsal konumlar ve işaretçilerin duruş pozisyonları farklılık göstermiştir. Bu tip veri kümesi oluşturulurken, birçok çalışmada arka planın benzerlik göstermesi, işaretçilerin çok çeşitlilik göstermemesi gibi faktörler gözetilirken, Sincan ve Keleş [24], bu veri kümesinde gerçek hayatla uyumlu olması açısından bu faktörleri göz ardı etmişlerdir. Böylelikle gerçek hayata daha yakın bir veri kümesi oluşturmuşlardır. Sincan ve Keleş [24], ayrıca bu veri kümesini ESA ve UKSB yöntemlerini kullanarak deneysel çalışmalar gerçekleştirmişlerdir. Bu deneysel çalışmalarda seçilen en iyi yaklaşım için %62,02 başarı oranı elde edilmiştir. Ayrıca Sincan ve Keleş [24] tarafından kullanılan yöntem Montalbano isimli İtalyan İşaret Dili'ni içeren veri kümesi üzerinde %96,11 başarı oranına ulaşmıştır.

Türkmen [25] tez çalışması kapsamında 2 farklı problemle yer vermiştir. Bunlardan ikincisinin çözüm yöntemi bu tez çalışmasıyla benzerlik göstermektedir. İkinci problem yer alan el yıkama işlemi sırasında alınan el görüntülerinin hijyen durumuna göre sınıflandırılmasıdır. Toplam 11 adet sınıf sayısı bulunmaktadır. Çözüm için RGB videolardan elde edilen görüntülerle öncelikle ESA tabanlı model kullanılıp öznitelik çıkarımı yapılırken devamında UKSB yöntemiyle el görüntülerinin zamansal etkileşimini yakalanmaya çalışılmıştır. Başarı oranı önerilen farklı yöntemler için doğrulama veri kümesi için %80 ile %91 aralığında ve test kümesi için %77 ile %86 aralığında değişmektedir.

Güney ve Erkuş [26] 3 farklı kişiden 3.200 adet görüntü ile 32 sınıflı bir veri kümesi oluşturmuşlardır. Daha sonra veri artırımı metoduyla bu adeti 19.200'e artırmıştırlar. Dolayısıyla sınıf sayısı başına frekans 100'den 600'e ulaşmıştır. Birçok farklı metot deneyerek yüksek başarı oranına ulaşılan bu çalışma, TİD'de bulunan

sadece statik kelime gruplarından oluşan veri seti içindir. Oysa TİD'e statik kelimelerin yanı sıra çoğunlukla durağan olmayan kelimeler yer almaktadır.

Yukarıda verilen çalışmalar ve veri setleri göz önünde bulundurulduğunda, mevcut tez çalışmasının özgün değerleri aşağıdaki gibi sıralanabilir:

- Veri Kümesi

Tez kapsamında oluşturulacak veri seti çevrim içi platformlarda mevcut olan videolardan, otomatik olarak elde edilecektir. Bu haliyle Camgöz vd. [20], Sincan ve Keleş [24] çalışmalarında kullanılan Microsoft Kinect gibi harici donanıma ihtiyaç yoktur.

Ayrıca Camgöz vd. [20], Sincan ve Keleş [24] veri kümelerini oluşturmak için fiziksel olarak kişilere ihtiyaç duyuyorken, mevcut çalışmada böyle bir ihtiyaç yoktur. Çünkü oluşturulacak olan veri kümesi çevrim içi platformlardan elde edilecektir. Bunun anlamı: ihtiyaç duyulduğu taktirde veri kümesinin güncellenebilir olacağıdır. Bundan dolayı oluşturulacak veri kümesi statik değil, dinamiktir. Yani çevrim içi platformlardan güncel videolar elde edildikçe, veri setini otomatik olarak güncellemek mümkün hale gelecektir.

Bunlara ek olarak, oluşturulacak olan veri kümesi, çeşitli çevrim içi platformlardan elde edilebileceği için, arka plan, işaretçiler vb. değişkenlerin sürekli farklılaşması, çalışmanın gerçek hayata uygulanabilirliği açısından özgünlük içermektedir.

- Yöntem

Çalışmada videolardan elde edilen RGB görüntüler kullanılmıştır. Bunun için otomatik olarak videolardan elde edilen RGB görüntülerden bölütleme (segmentation) işlemi ile Türk İşaret Dili'ne ait görüntü bölütleri OpenCV [27] kütüphanesi kullanılarak elde edilmiştir.

Çalışmanın seyrine göre iki farklı yaklaşım yöntemi izlenmiştir. Bunlardan birincisi ESA, ikincisi UKSB yönteminin kullanılmasıdır. ESA'nın zaman serisi verilerinde sınırlı çalışma yeteneği nedeniyle zamansal veriyi daha iyi işleyebilmek için UKSB tabanlı model kullanılmıştır. Öne sürülen yaklaşım Sincan ve Keleş [24]'in çalışmasına benzerlik göstermektedir. Burada temel farklılık kullanılan veri kümesi ve

bu veri kümesine uyumlu hale getirilen farklı metotlardır. Sincan ve Keleş [24] çalışmasında spesifik bir cihaz kullanarak görüntüleri elde etmiş ve veri kümesini oluşturmuştur. Dolayısıyla [24]'te oluşturulan veri kümesi güçlü bir etiketlemeye sahiptir.

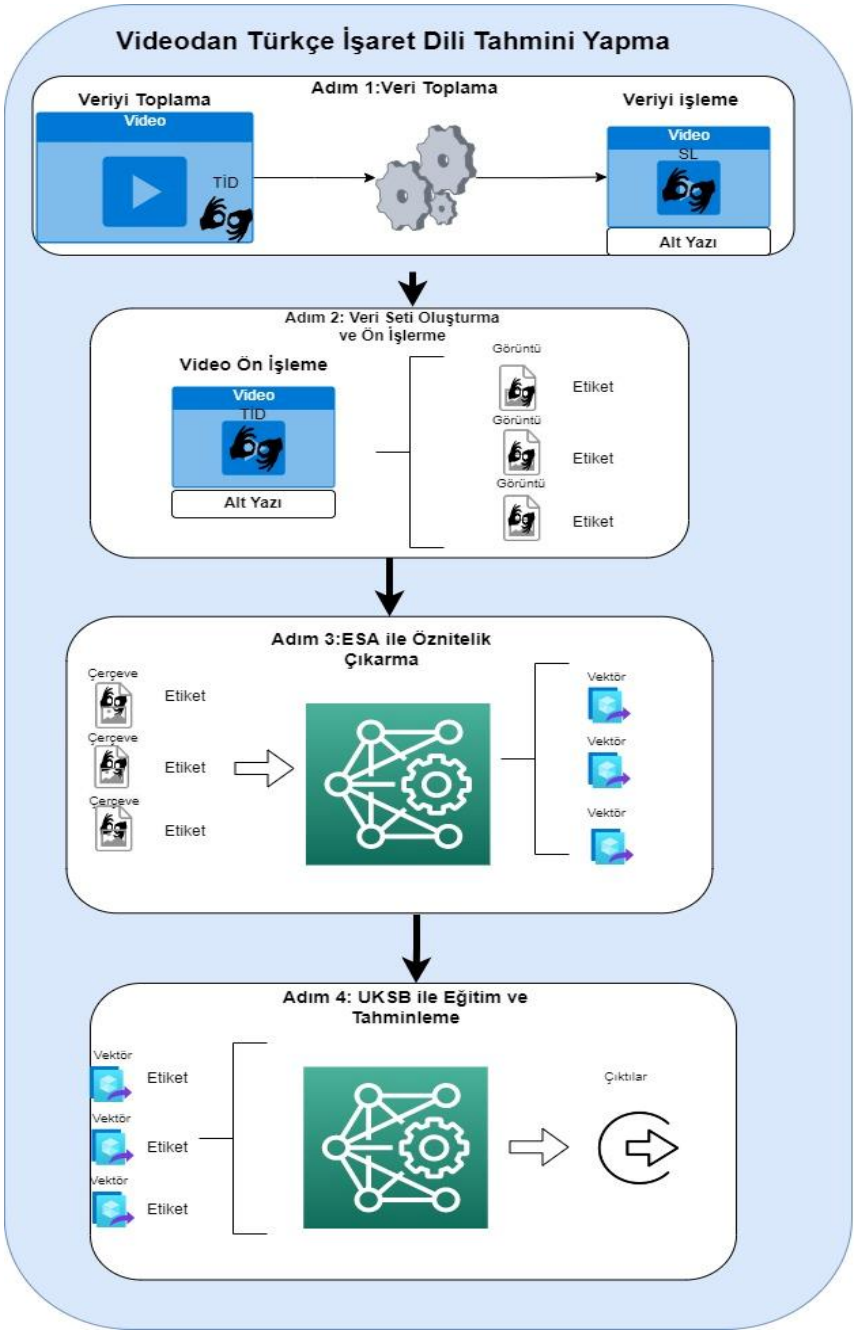
3. TÜRKÇE İŞARET DİLİ İÇİN ZAYIF GÜDÜMLÜ MAKİNA ÖĞRENMESİ

Bu bölümde çalışma sürecinde kullanılacak olan veri kümesinin oluşturulmasına, elde edilen veri kümesinden öznitelik çıkarımına ve eğitim sürecine yer verilmiştir. Bu tez çalışmasında TİD'in tanınmasına yönelik iki aşamalı YSA modeli öne sürülmektedir. Bu sınıflandırma modelinin eğitilmesi için, zayıf güdümlü bir yöntem kullanarak TİD'e özgü bir veri kümesi oluşturulmuştur. Bu veri kümesi çevrim içi platformlardan elde edilmiştir. Böylelikle gerek görüldüğü takdirde, veri kümesi dinamik olarak güncellenebilmektedir. Modelin güncel hayata uyumluluğu açısından, çevrim içi platformlardan elde edilecek olan görüntülerin kullanılması çalışmayı özgün kılmaktadır. Veri kümesi oluşturulurken kullanılan videolarda hem ses hem de işaret dili görüntüsü bulunmaktadır. Öncelikle bu videoların alt yazı dosyaları oluşturulmuştur. Ardından bu alt yazı dosyalarında geçen kelimelerden TİD'e uygun olanlar seçilmiştir. Seçilen bu kelimelerin geçtiği bölümler çeşitli uygulamalar (bakınız bölüm 3.2) yardımıyla videoların içinden alınarak veri kümesi elde edilmiştir. Çalışmada öne sürülen modellerden ilki ESA'dır. ESA verilen girdilerden öznitelik vektörlerinin çıkarılmasını sağlanmıştır. Öne sürülen ikinci modelde elde edilen bu öznitelik vektörleriyle UKSB kullanılarak veri kümesi ikinci eğitim sürecinden geçirilmiş ve bunun sonucunda sınıflandırma yapılmıştır. Çalışmanın genel iş akış şeması Şekil 3.1'de gösterilmiştir.

3.1. Veri Kümesi

Çalışmanın bu bölümünde veri toplama ve dönüştürme süreci anlatılmaktadır. Eğitim sürecinde kullanılacak olan veri kümesi çevrim içi platformlarda mevcut olan video görüntülerinden zayıf güdümlü bir yöntem kullanarak otomatik olarak elde edilmiştir. Çalışmada veri olarak çevrim içi platformlardan alınan, içerisinde TİD'i ihtiva eden videolar kullanılmıştır. Bu videolar hem Türkçe ses, video ve TİD içerir. Veri kümesinin oluşturulması için harici bir donanıma ihtiyaç yoktur. Ayrıca otomatik olarak elde edildiğinden dolayı statik değil dinamiktir. Çevrim içi platformlardan güncel videolar elde edildikçe, veri kümesini otomatik olarak güncellemek mümkündür. Bunlara ek olarak, oluşturulan veri kümesi, çeşitli çevrim içi

platformlardan elde edilebileceği için, arka plan, işaretçiler vb. değişkenlerin sürekli farklılaşması, çalışmanın gerçek hayata uygulanabilirliği açısından özgünlük içermektedir. Çevrim içi platformların dinamikliği modelde kullanılacak olan veri kümesinin içeriğini çeşitlendirmiştir.



Şekil 3.1: Öne sürülen yaklaşımın temel adımları 3 aşamada tanımlanmıştır.

Modelin güncel hayata uyumluluğu açısından, çevrim içi platformlardan elde edilecek olan görüntülerin kullanılması çalışmayı özgün kılmaktadır. Böylelikle düşük maliyetle yüksek başarı oranı elde edilmiştir.

3.1.1. Girdilerin Elde Edilmesi

Girdiler TRT Haber Merkezi'nin hazırlayıp sunduğu işitme engelliler haber bülteninde alınmıştır. Toplam 110 adet video kullanılmıştır. Bu videolar öncellikle indirilmiştir. Her bir videoda hem ses hem de haber akışı hem de TİD bulunmaktadır. Videolar herhangi bir işleme tabii tutulmadan önce lokal bir makineye indirilmiştir. Eğitimde kullanılacak olan videolar haber programından alındığı için her videonun başında ve sonunda haberin jeneriği bulunmaktadır. Çalışma da kullanılacak videolar da ise sadece konuşmanın geçtiği kısımların olması yeterlidir. Dolayısıyla indirilen videolar konuşmanın başladığı ve sonlandığı zamana göre kısaltılmıştır. Ayrıca videoların içerisinde hem işaretçi hem de haber içeriğinin görüntüsü bulunmaktadır. Bu durum eğitim sürecinde karışıklığa sebep olacağı için, çoklu kısaltma işlemi tamamlandıktan sonra görüntüler, video görüntülerinin alanı işaret dilinin bulunduğu yeri kapsayacak şekilde kesilmiştir. Her videoda işaretçinin bulunduğu lokasyon ve konuşmanın olduğu bölüm farklıdır. Bundan dolayı bu işlemler her bir video için ayrı ayrı uygulanmıştır. Bu işlemler sonucunda her bir video hem zaman bakımından kısaltılarak hem de görüntü bakımından ilgili kısım alınarak işlenmiştir. Şekil 3.1. ve Şekil 3.2.'de video görüntülerinin işlenmiş ve işlenmemiş hali gösterilmiştir.

Şekil 3.2: Videolardan elde edilen görüntünün işlenmemiş hali.

Şekil 3.3: Videolardan elde edilen görüntünün işlenmiş hali.

Eğitimde kullanılacak olan videoların saniyelik görüntü sayısı (fps) birbirinden farklıdır. Bu fark eğitim sürecini sekteye uğratmaması için çevrim içi platformlar kullanılarak videoların saniyelik görüntü sayıları 25'e sabitlenmiştir. Videolardan elde edilecek olan tüm görüntülerin boyutları 224x224 piksel boyutuna sabitlenmiştir. Görüntülerin tamamı 3 kanallıdır. Özetlemek gerekirse girdi olarak kullanılacak olan videolar saniyelik görüntüsü 25'tir ve bu görüntüler 224x224x3 boyutundadır. Videolarda iki farklı işaretçi bulunmaktadır. Arka plan görüntüsü her videoda değişkenlik göstermektedir. Bu çalışma kapsamında ham videolardan büyük ve dağınık bir veri kümesi elde edildiği için çalışmanın başında eğitilecek sınıf sayısı ve örnek sayısı belirlenmemiştir. Veri kümesi elde edildikten sonra eğitim sürecine uygun olan sınıflar seçilerek eğitime dahil edilmiştir. Dolayısıyla veri kümesi giderek büyümüş ve bundan dolayı veri kümesi büyüdükçe eğitim süresi daha fazla zaman almıştır.

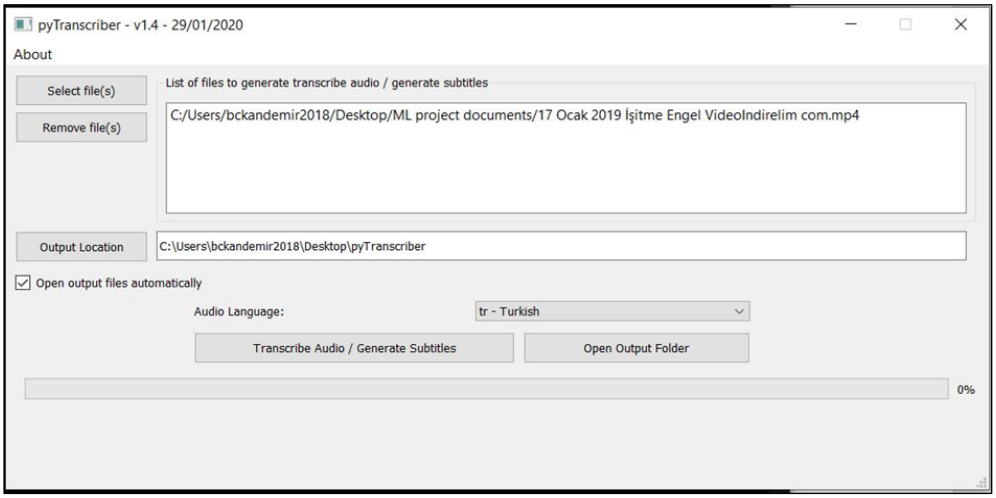
3.1.2. Girdilerin Düzenlenmesi

Video görüntüler elde edilmesinden sonra bu videoların içerisinde TİD'e uygun kelimelerin elde edilmesi gerekmektedir. TİD'e uygun kelimelerin elde edilmesi eğitim süreci için oldukça önem arz etmektedir. Bundan dolayı en uygun ve güvenilir yöntem aranmıştır. TİD'de sıklıkla kullanılan 510 adet kelime belirlenmiştir. Amaç bölüm 3.1.1.'de bahsi geçen video görüntülerin içerisinde işaretçinin bu kelimeleri gösterdiği bölümlerin alınmasıdır. Bu kısımda kelimelerin hazırlanmasıyla ilgili iki

farklı uygulama sonuçları incelenmiştir. İlk uygulama doğruluk oranı yüksek sonuçlar vermemiştir. Bu sebeple ikinci bir yöntem arayışına girilmiştir. Kullanılan uygulamaların her ikisi de ses kaydını ya da içinde ses bulunan videoyu yazıya çevirme yöntemi içerir. Bu sesten yazı elde etme ile ilgili uygulamaların doğruluk oranları birçok değişkene bağlıdır. Bunlardan bazıları arka plandaki gürültüsü, dönüştürülecek sesin kalitesi ve netliğidir. Denenen her iki uygulamada da doğruluk oranları zaman zaman %99'a ulaşmıştır. Bununla birlikte bu oran, uygulamadan elde edilen tüm çıktılar için geçerli değildir. Doğruluk oranın yüksekliğinin devamlılık sağlamaması, eğitim sürecinde kullanılacak olan verinin üretiminde problemler ortaya çıkarmaktadır.

3.1.2.1. Pytranscriber Uygulamasının Kullanımı

PyTranscriber [28] açık kaynak kodlu, ses ve video dosyaları için otomatik transkripsiyon oluşturmak için kullanılan bir uygulamadır. Bölüm 3.1.1.'de bahsi geçen videoların alt yazı dosyalarını oluşturmak için bu uygulama kullanılmıştır. Bu uygulamaya alt yazsının oluşturulması istenen video verildiğinde, Şekil 3.3'te de görüldüğü üzere çıktı olarak "srt" ve "txt" uzantılı dosyalar oluşturmaktadır.



Şekil 3.4: PyTranscriber [28] uygulamasının ara yüzü.

Burada amaç videolarda elde edilen karelere uygun etiketler oluşturmaktır. Bunun için öncelikle 25 saniyelik görüntü sayısına sahip her bir video teker teker PyTranscriber [28] uygulamasına verilmiştir. Her bir video için elde edilen "srt" ve

“txt” uzantılı dosyalar üzerinde çeşitli dönüşümler yapılmıştır. Bunlardan ilki “txt” dosyasındaki her bir kelimenin kökünün elde edilmesidir.

Kelimeler, Zemberek [29] isimli açık kaynak kodlu Türkçe doğal dil işleme kütüphanesinde işlenerek kökleri elde edilmiştir. Videoların içerisinde en sık geçen önceden belirlenmiş olan 510 kelime kök halinde aranmıştır. İkinci olarak Şekil 3.4’te görüldüğü üzere elde edilen “srt” uzantılı dosyalar başlangıç ve bitiş süreleri göz önünde bulundurularak her bir kelimenin video içerisinde geçen zaman aralığı bulunmaya çalışılmıştır.

Şekil 3.4’te görüldüğü üzere uygulama srt uzantılı dosya oluştururken, her kelime için videoda geçen zaman aralığını hesaplamamıştır. Kelimeler keyfi bir biçimde gruplara ayrılmıştır. Fakat videodan elde edilen srt dosyasında her satırın başlangıç ve bitiş zamanı belli olduğu için bu şekilde kelimelerin video hangi zaman aralığında geçtiği tahmin edilmeye çalışılmıştır. Bazı kelimeler için doğru sonuçlar elde edilmesine rağmen birçok kelime de hesaplanan zaman aralığı ile video kelimenin bulunduğu an aynı sonucu vermediği için bu uygulamayı kullanılmaya devam edilmemiştir.

```
1
00:00:00,256 --> 00:00:04,608
Bu kez işitme engelliler için hazırladığımız haber bülteni ile karşınızdayız

2
00:00:04,864 --> 00:00:06,912
Dengeleme Disiplin ve

3
00:00:07,424 --> 00:00:09,728
Yeni ekonomi programının temelleri

4
00:00:09,984 --> 00:00:12,288
Hazine ve maliye bakanı Berat Albayrak

5
00:00:12,544 --> 00:00:13,312
İstanbul'da

6
00:00:13,568 --> 00:00:15,360
3 yıllık yeni ekonomi programını

7
00:00:15,616 --> 00:00:16,640
Büyüme
```

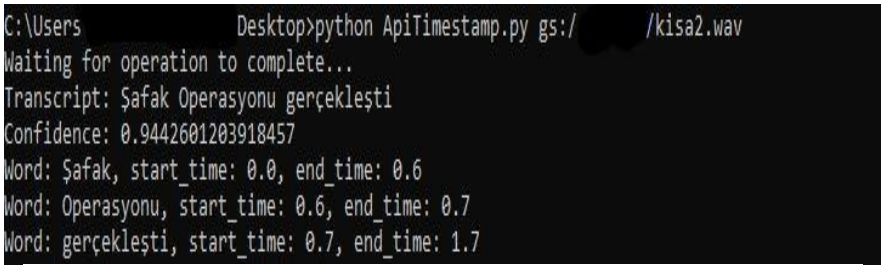
Şekil 3.5: PyTranscriber’dan alınan “srt” uzantılı dosya örneği.

3.1.2.2. Google API Speech To Text Uygulamasının Kullanımı

Google’ın “Speech to Text” [30] uygulaması ses dosyalarını metin dosyalarına dönüştürmektedir. Uygulamanın birçok varyasyonu bulunmaktadır. Bu tez çalışması

kapsamında ses dosyaları uygulamaya verilerek hem metin hem de her bir kelimenin ses dosyasında geçtiği zaman bilgisi elde edilmiştir. Bu uygulamanın PyTranscriber [28] uygulamasından farkı her kelime için ayrı ayrı zaman bilgisine ulaşılabilmesidir. Bunun dışında Şekil 3.5'te de görüldüğü üzere uygulama her ses dosyası için dönüştürme işlemini tamamladığında, işlemle ilgili doğruluk oranı bilgisini de vermektedir. Her iki uygulama için bu oran karşılaştırıldığında Google Speech to Text [30] uygulamasının PyTranscriber [28] uygulamasına göre daha yüksek doğruluk oranına sahip olduğu görülmüştür. Videolar uygulamaya verilmeden önce çevrim içi bir platform kullanılarak ses dosyalarına dönüştürülmüştür.

Her bir videonun ses dosyası için uygulamadan elde edilen çıktılar da Türkçe olduğu için karakter hataları ortaya çıkmıştır. Bu hatalar düzenlendikten sonra elde edilen metin dosyalarındaki kelimelerin her biri Zemberek'e [29] kütüphanesine verilerek kelimelerin kökü elde edilmiştir. Bu aşamadan sonra elde edilen kök halinde olan her kelime ve kelimenin video içerisinde başlangıç ve bitiş zaman bilgisi sıralı bir şekilde kayıt altına alınmıştır. Kelimelerin süresi birbirinden farklıdır. Bundan dolayı son olarak istenilen kelimeleri videolara ayırmadan önce her kelimenin video süresi ortalama bir saniyeye eşitlenmiştir. Bu işlemden sonra 111 videodan, önceden belirlenen 510 kelimenin içinde bulunduğu her bir bölüm alınarak bir saniyelik video klipler halinde kaydedilmiştir. Mevcut durumda beş yüz on adet kelime için oluşturulmuş klasörler ve bu klasörlerin içerisinde sayıları birbirinden farklı toplam 56.867 adet video klip elde edilmiştir.



Şekil 3.6: Google Speech to Text [30] arayüzü şekli.

Kelimeler ile bunlara karşılık gelen çerçevelerin eşleştirilmesi işleminde oluşan zaman kaymalarını önlemek amacıyla gerçekleştirilen adımlar bölüm 2.3'te bahsedilmektedir.

3.2. Öznitelik Çıkarımı

Bu tez kapsamında ESA'nın zaman serisi verilerindeki sınırlı çalışma yeteneği nedeniyle zamansal veriyi daha iyi işleyebilmek için UKSB tabanlı model kullanılmıştır. Çalışmanın bu bölümünde ESA kullanılarak eğitilecek olan sınıfların belirlenmesi ve belirlenen sınıflarından aynı sınıf sayısına sahip hem dengeli hem dengesiz veri setleriyle öznitelik çıkarımına yer verilmiştir. Elde edilen her biri 25 saniyelik görüntü sayısına sahip ve bir saniyelik olan 56.867 adet video klipin her sınıfı için frekansı hesaplanmıştır. Eğitim sürecine dahil edilecek olan kelimelerin seçimine hem frekans sayısına hem de kelimenin yapısına bakılarak karar verilmiştir. Örnek vermek gerekirse en sık geçen kelimeler arasında "ve" bağlacı yer almaktadır. Fakat çoğu zaman işaretçi bu bağlaca anlatımında yer vermemiştir. Dolayısıyla sınıf seçimi belirlenirken frekansla birlikte kelimenin yapısı da bu çalışma için önem arz etmektedir. Bu bağlamda alınacak olan kelime sınıfları belirlendikten sonra her sınıfa ait video klipler sıralı bir biçimde çerçevelere ayrılmıştır. ESA görüntülerden öznitelik çıkarmak için kullanılabilen bir sinir ağı yapısıdır. Çalışmanın bu bölümünde ESA'nın uzamsal bilgileri işleme yeteneğinden dolayı, oluşturulan görüntülerden ESA kullanılarak öznitelik çıkarılmıştır. ESA modelinde DenseNet [31] mimarisine benzer bir yapı kullanılmıştır. Oluşturulan veri düzgün bir biçimde sınıflara ayrılarak klasörlere ayrıldığı için, bu veri yapısına uygun bir jeneratör yazılarak etiketler oluşturulmuş ve eğitilen ESA tabanlı modelden en ideal eğitim ağı seçilerek UKSB modeline verilmeye uygun biçimde vektörler üretilmiştir. ESA modelinin yapısı ve kullanılan katmanlar Tablo:3.1'de gösterilmiştir.

Her bir eğitim süreci için sınıf sayısı arttırılmıştır. Üç ayrı sınıf sayısıyla eğitim yapılmıştır. Bu üç sınıf için de hem dengeli hem dengesiz veri kümesi eğitime tabii tutulmuştur. Dengeli veri kümesi oluşturulurken seçilen kelime sınıflarının frekansları hesaplanmış ve en küçük frekans sayısına sahip kelime sınıfı baz alınarak diğer sınıflar düzenlenmiştir. Bu veri kaybına sebep olmuştur. Çalışmada bu aşamadan sonra aynı sınıf sayısını için ikinci veri kümesi oluşturulup eğitime tabii tutulmuştur. Oluşturulan bu ikinci veri kümesi düzensiz bir veri kümesidir. Bu veri kümesi hazırlanırken sınıflar arasında en yüksek frekansa sayısına sahip olan sınıfın frekansı, en düşük frekans sayısına sahip sınıfın iki katından fazla olmayacak şekilde düzenlenmiştir. Sınıf sayıları her bir eğitim sürecinde 5'er arttırılmıştır. Seçilecek olan kelimeler

belirlenirken hem frekansına hem de kelimenin cümle içerisinde kullanımı göz önünde bulundurulmuştur. Buna bağlı olarak eğitim sürecine dahil olan görüntü sayısı, oluşturulan her veri kümesi için artış göstermiştir. Bu şekilde oluşturulan toplam 6 adet veri kümesi için 6 adet problem bulunmaktadır. Oluşturulan bu veri kümelerine ait problem, sınıf ve frekans bilgisi Tablo: 3.2'de gösterilmiştir. Tüm problemlerin eğitim sürecinde kullanılan makina Dual Intel Xeon Processor E5-2640 v4 (10C, 2.4GHz,3.4GHz Turbo,2133MHz, 25MB,90W,8GB NVDIA Quadro M4000) 'dir. Öznitelik çıkarıma işlemi sonunda videolardan elde edilen her bir çerçeve için sınıf sayısı kadar boyuta sahip vektörler elde edilmiştir.

Tablo 3.1: ESA modelinin yapısında DenseNet [31] mimarisine benzer bir yapı.

Layers	Output Size	Dense-Net	
Convolution	112 x 112	7 x 7 conv, stride 2	
Pooling	56 x 56	3 x 3 max pool, stride 2	
Dense Block (1)	56 x 56	1 x 1 conv 3 x 3 conv	→ X 2
Transition Layer (1)	56 x 56	1 x 1 conv	
	28 x 28	2 x 2 average pool, stride 2	
Dense Block (2)	56 x 56	1 x 1 conv 3 x 3 conv	→ X 3
Transition Layer (2)	28 x 28	1 x 1 conv	
	14 x 14	2 x 2 average pool, stride 2	
Dense Block (3)	14 x 14	1 x 1 conv 3 x 3 conv	→ X 4
Transition Layer (3)	14 x 14	1 x 1 conv	
	7 x 7	2 x 2 average pool, stride 2	
Dense Block (4)	7 x 7	1 x 1 conv 3 x 3 conv	→ X 3

Tablo 3.2: Tez kapsamında ele alınan 6 adet problem mevcuttur.

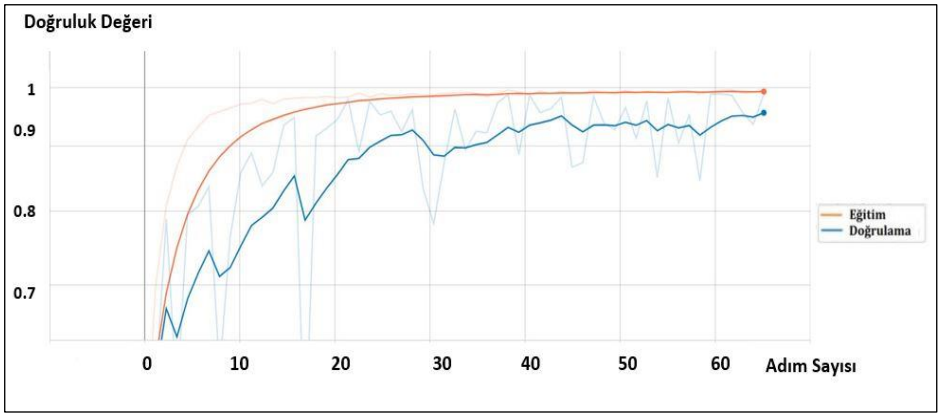
Problemin Adı	Sınıf Sayısı	Veri Tipi	Frekans
P1	5	Dengeli	41.005
P2	5	Dengesiz	58.966
P3	10	Dengeli	64.100
P4	10	Dengesiz	96.283
P5	15	Dengeli	71.235
P6	15	Dengesiz	115.530

3.2.1. P1 ve P2 İçin Oluşturulan Veri Kümesi ve Öznitelik Çıkarımı

Çalışmanın bu bölümünde 111 adet videonun içerisinde geçen 5 farklı kelime için sınıflar oluşturulmuştur. Bu sınıflar sırasıyla gözaltı, örgüt, çalışmak, operasyon ve Türkiye kelimeleridir.

3.2.1.1. P1 İçin Veri Kümesi ve Öznitelik Çıkarımı

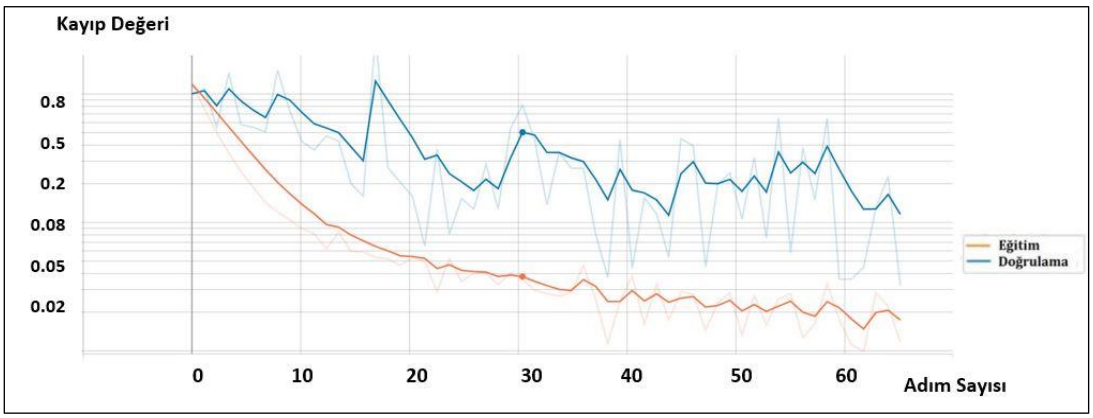
Beş sınıf için toplam 41.005 adet görüntü bulunmaktadır. En küçük frekansa sahip sınıf sayısı göz önüne alınarak diğer sınıfların frekansları düzenlenmiştir. 41.005 görüntünün %70'i eğitim kümesi, %15'i doğrulama kümesi olarak, geriye kalan %15'i ise test kümesi olarak ayrılmıştır.



Şekil 3.7: P1 için ESA eğitimi doğruluk fonksiyonu grafiği edilmiştir.

Sınıf başına düşen görüntü sayısı 8.201 adettir. Eğitim süreci yaklaşık olarak 13 saat sürmüştür. Eğitim sürecinin sonlarına doğru Şekil 3.7'de görüldüğü üzere doğruluk oranı yaklaşık olarak eğitim kümesi için %99'e, doğrulama kümesi için %93'e ulaşmıştır. Kayıp değerinin değeri ise Şekil 3.8'de görüldüğü üzere eğitim süreci ilerledikçe eğitim kümesi için %01, doğrulama kümesi için %09'a kadar düşmüştür. Eğitim sürecinin sonunda videolardan elde edilen her bir çerçeve için 5 boyutlu vektörler elde

Eğitim tamamlandıktan sonra en yüksek doğruluğa ve en düşük kayıp değerine sahip modeli seçilerek UKSB modeline verilmeye uygun biçimde vektörler

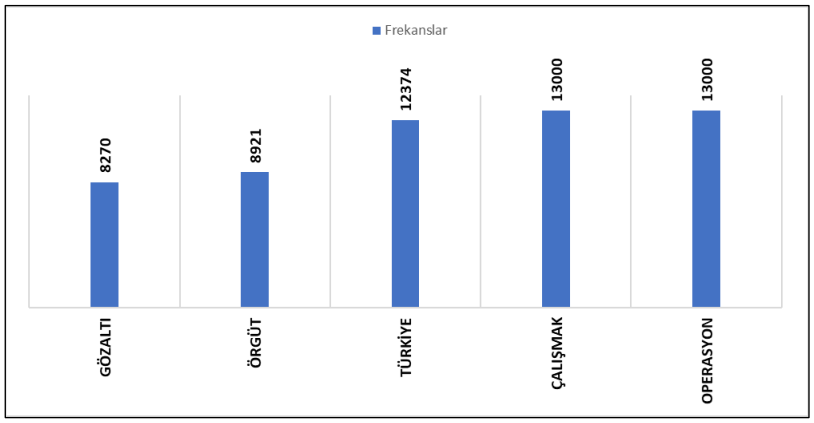


Şekil 3.8: P1 için ESA eğitimi kayıp değeri grafiği.

üretilmiştir.

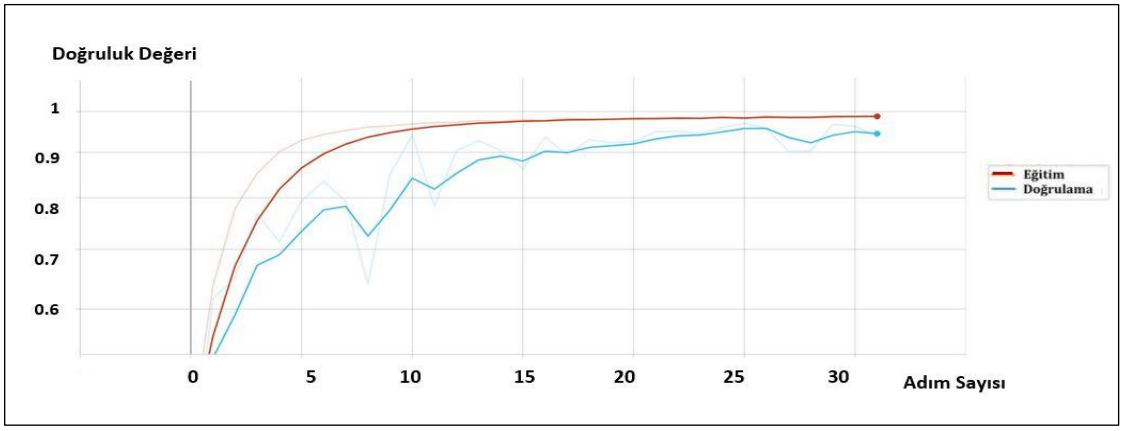
3.2.1.2. P2 İçin Veri Kümesi ve Öznitelik Çıkarımı

Beş sınıf için toplam 58.966 adet görüntü bulunmaktadır. En küçük frekansa sahip sınıf sayısı göz önüne alınarak diğer sınıfların frekansları düzenlenmiştir. 58.966 görüntünün %70'i eğitim kümesi, %15'i doğrulama kümesi olarak, geriye kalan %15'i ise test kümesi olarak ayrılmıştır. Sınıf başına düşen görüntü sayısı sabit değildir. Maksimum sınıf frekansı minimum frekansa sahip sınıfın iki katından fazla olmayacak biçimde ayarlanmıştır ve Şekil 3.9'da gösterilmiştir.



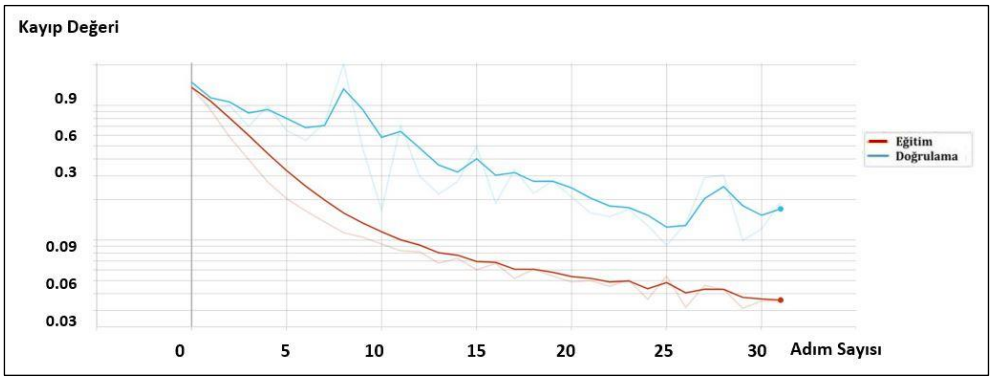
Şekil 3.9: P2 için frekans grafiği.

Eğitim süreci yaklaşık olarak 10 saat sürmüştür. Eğitim sürecinin sonlarına doğru Şekil 3.10'da görüldüğü üzere doğruluk oranı yaklaşık olarak eğitim kümesi için %98'e, doğrulama kümesi için %93'e ulaşmıştır. Kayıp değerinin değeri ise Şekil 3.11'de görüldüğü üzere eğitim süreci ilerledikçe eğitim kümesi için %03, doğrulama kümesi için %18'a kadar düşmüştür. Eğitim sürecinin sonunda videolardan elde edilen her bir çerçeve için 5 boyutlu vektörler elde edilmiştir.



Şekil 3.10: P2 için ESA eğitimi doğruluk fonksiyonu grafiği.

Eğitim tamamlandıktan sonra en yüksek doğruluğa ve en düşük kayıp değerine sahip modeli seçilerek UKSB modeline verilmeye uygun biçimde vektörler üretilmiştir.



Şekil 3.11: P2 için ESA eğitimi kayıp değeri grafiği.

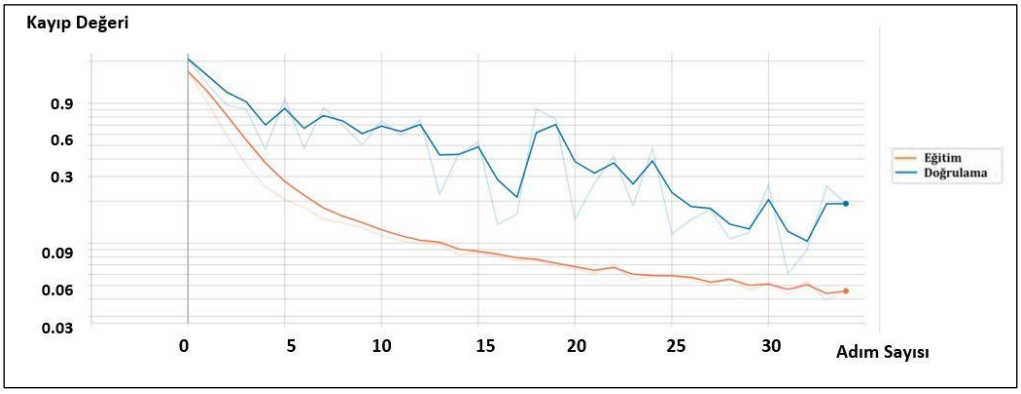
3.2.2. P3 ve P4 İçin Oluşturulan Veri Kümeleri ve Öznitelik Çıkarımı

Çalışmanın bu bölümünde 111 adet videonun içerisinde geçen 10 adet farklı kelime için sınıflar oluşturulmuştur. Bu sınıflar sırasıyla gözaltı, örgüt, çalışmak, operasyon, Türkiye, açıklamak, bakan, araç, başlamak, engel kelimeleridir. Türkçe

İşaret Dili'nde gösterim şekli birbiriyle aynı olduğu için kelimeler gruplandırılarak tek bir sınıfa toplanmıştır. Bu kelime grupları araç-araba-otomobil, bakan-başkan-bakanlık şeklindedir.

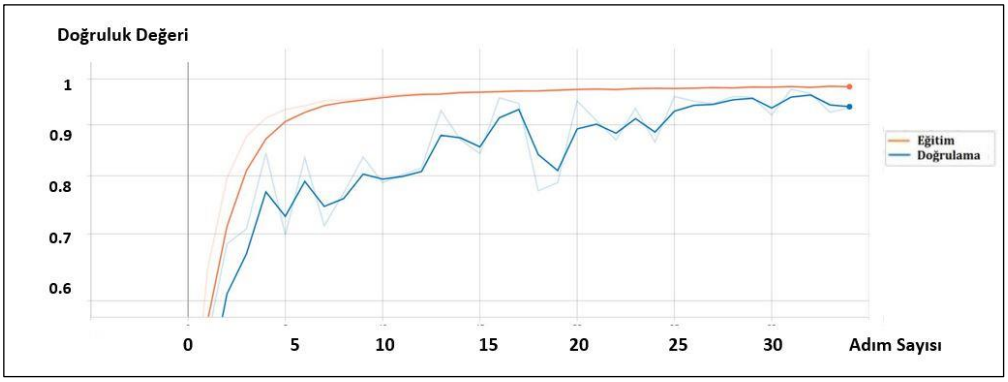
3.2.2.1. P3 İçin Veri Kümesi ve Öznitelik Çıkarımı

Beş sınıf için toplam 64.100 adet görüntü bulunmaktadır. En küçük frekansa sahip sınıf sayısı göz önüne alınarak diğer sınıfların frekansları düzenlenmiştir. 64.100 görüntünün %70'i eğitim kümesi, %15'i doğrulama kümesi olarak, geriye kalan %15'i ise test kümesi olarak ayrılmıştır. Sınıf başına düşen görüntü sayısı 6.410 adettir. Oluşturulan bu veri kümesinde sınıf sayısı artmış olsa da dengeli bir veri kümesi oluşturulduğu için sınıf başına düşen örneklem sayısı 6.410 görüntüye düşmüştür. Eğitim süreci yaklaşık olarak 13 saat sürmüştür. Eğitim sürecinin sonlarına doğru Şekil 3.12'de görüldüğü üzere doğruluk oranı yaklaşık olarak eğitim kümesi için %98'e, doğrulama kümesi için %93'e ulaşmıştır. Kayıp değerinin değeri ise Şekil 3.13'te görüldüğü üzere eğitim süreci ilerledikçe eğitim kümesi için %04, doğrulama kümesi için %19'a kadar düşmüştür. Eğitim sürecinin sonunda videolardan elde edilen her bir çerçeve için 10 boyutlu vektörler elde edilmiştir.



Şekil 3.12: P3 için ESA eğitimi doğruluk fonksiyonu grafiği.

Eğitim tamamlandıktan sonra en yüksek doğruluğa ve en düşük kayıp değerine sahip modeli seçilerek UKSB modeline verilmeye uygun biçimde vektörler üretilmiştir.

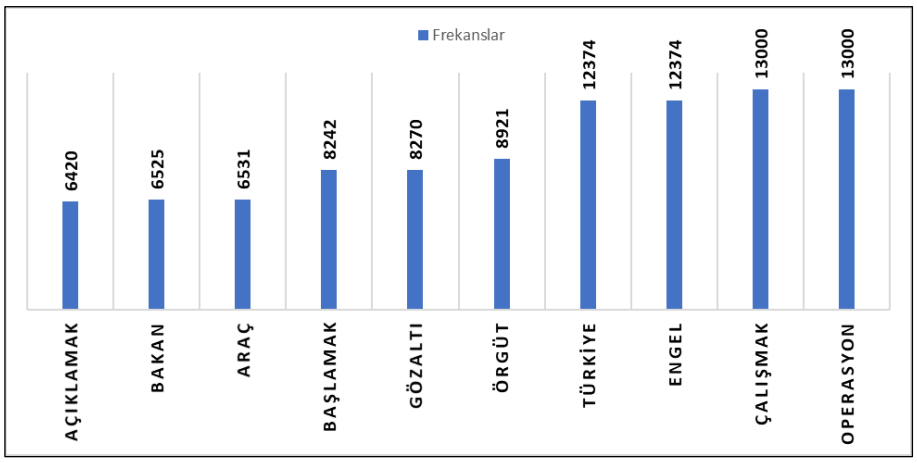


Şekil 3.13: P4 için ESA eğitimi kayıp değeri grafiği.

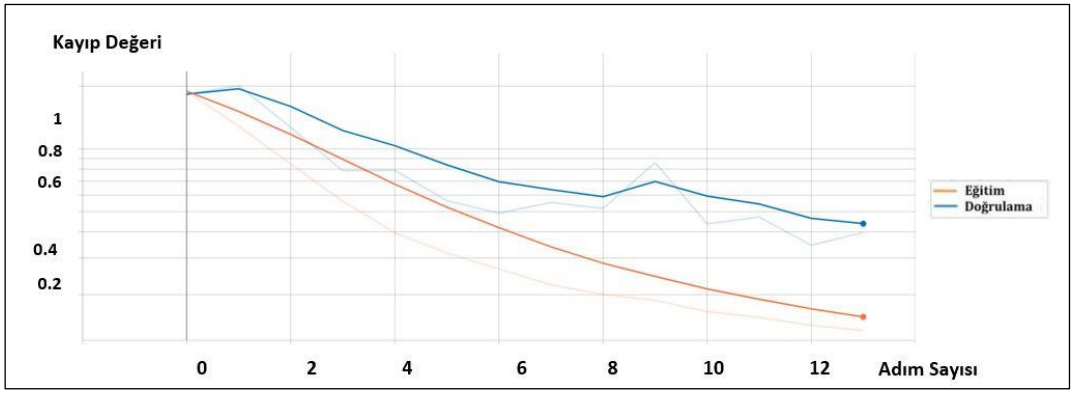
3.2.2.2. P4 İçin Veri Kümesi ve Öznitelik Çıkarımı

Beş sınıf için toplam 96.283 adet görüntü bulunmaktadır. En küçük frekansa sahip sınıf sayısı göz önüne alınarak diğer sınıfların frekansları düzenlenmiştir. 96.283 görüntünün %70'i eğitim kümesi, %15'i doğrulama kümesi olarak, geriye kalan %15'i ise test kümesi olarak ayrılmıştır. Sınıf başına düşen görüntü sayısı sabit değildir. Maksimum sınıf frekansı minimum frekansa sahip sınıfın iki katından fazla olmayacak biçimde ayarlanmıştır ve Şekil 3.14'te gösterilmiştir.

Eğitim süreci yaklaşık olarak 7 saat sürmüştür. Eğitim sürecinin sonlarına doğru Şekil 3.15'te görüldüğü üzere doğruluk oranı yaklaşık olarak eğitim kümesi için %95'e, doğrulama kümesi için %86'e ulaşmıştır. Kayıp değerinin değeri ise Şekil 3.16'da görüldüğü üzere eğitim süreci ilerledikçe eğitim kümesi için %13, doğrulama kümesi için %39'a kadar düşmüştür. Eğitim sürecinin sonunda videolardan elde edilen her bir çerçeve için 10 boyutlu vektörler elde edilmiştir.

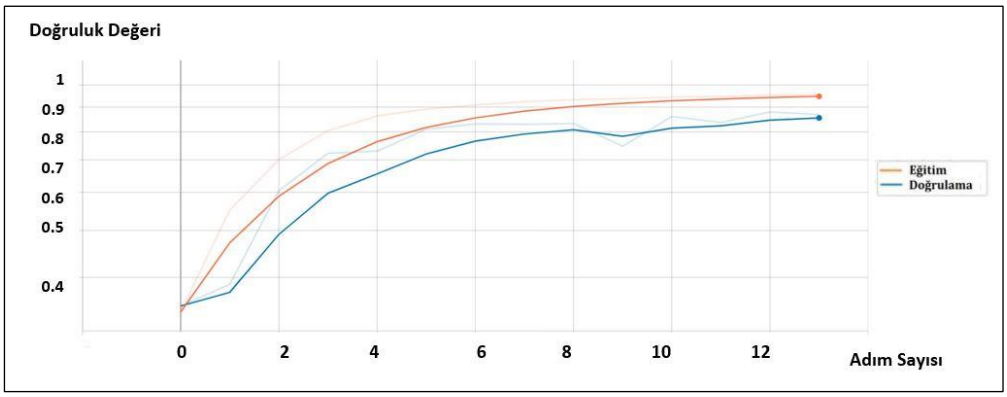


Şekil 3.14: P4 için frekans grafiği.



Şekil 3.15: P4 için ESA eğitimi kayıp değeri grafiği.

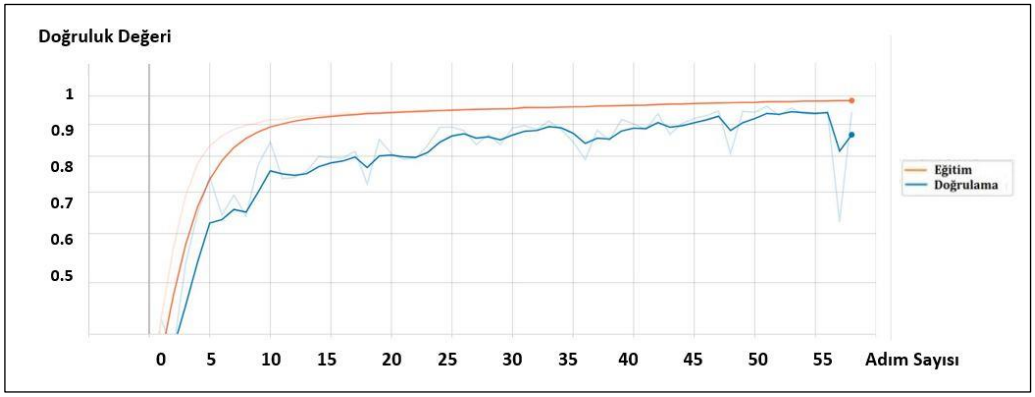
Eğitim tamamlandıktan sonra en yüksek doğruluğa ve en düşük kayıp değerine sahip modeli seçilerek UKSB modeline verilmeye uygun biçimde vektörler üretilmiştir.



Şekil 3.16: P4 için ESA eğitimi kayıp değeri grafiği.

3.2.3. P5 ve P6 İçin Oluşturulan Veri Kümesi ve Öznitelik Çıkarımı

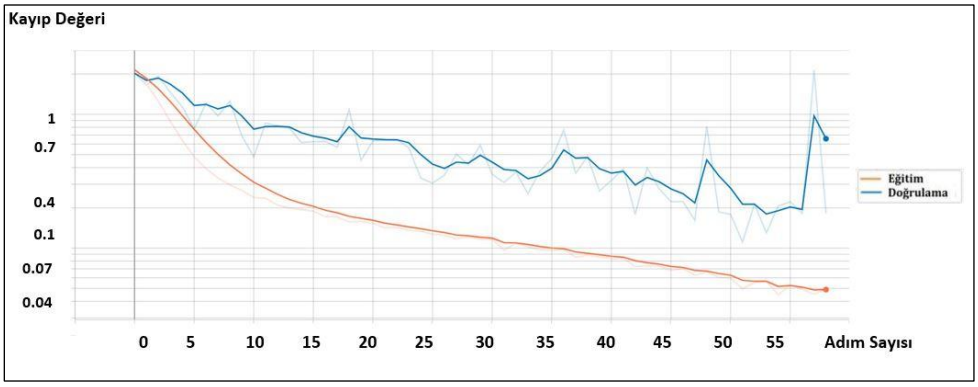
Çalışmanın bu bölümünde 111 adet videonun içerisinde geçen 10 adet farklı kelime için sınıflar oluşturulmuştur. Bu sınıflar sırasıyla gözaltı, örgüt, çalışmak, operasyon, Türkiye, açıklamak, bakan, araç, başlamak, engel, almak, haber, hazırlamak, söylemek, yakalamak kelimeleridir. Seçilen kelime sınıflarından bazılarının Türkçe İşaret Dili'nde gösterim şekli birbiriyle aynı olduğu için kelimeler gruplandırılarak tek bir sınıfa toplanmıştır. Bu kelime grupları araç-araba-otomobil, hazırlık-hazırlamak, söylemek- konuşmak, almak- yakalamak, bakan-başkan-bakanlık şeklindedir.



Şekil 3.17: P5 için ESA eğitimi doğruluk fonksiyonu grafiği.

3.2.3.1. P5 İçin Veri Kümesi ve Öznitelik Çıkarımı

Beş sınıf için toplam 71.235 adet görüntü bulunmaktadır. En küçük frekansa sahip sınıf sayısı göz önüne alınarak diğer sınıfların frekansları düzenlenmiştir. 71.235 görüntünün %70'i eğitim kümesi, %15'i doğrulama kümesi olarak, geriye kalan %15'i ise test kümesi olarak ayrılmıştır. Sınıf başına düşen görüntü sayısı 4.748 adettir. Oluşturulan bu veri kümesinde sınıf sayısı artmış olsa da dengeli bir veri kümesi oluşturulduğu için sınıf başına düşen örneklem sayısı 4.748 görüntüye düşmüştür. Eğitim süreci yaklaşık olarak 23 saat sürmüştür. Eğitim sürecinin sonlarına doğru Şekil 3.17'de görüldüğü üzere doğruluk oranı yaklaşık olarak eğitim kümesi için %98'e, doğrulama kümesi için %94'e ulaşmıştır. Kayıp değerinin değeri ise Şekil 3.18'de görüldüğü üzere eğitim süreci ilerledikçe eğitim kümesi için %04, doğrulama kümesi için %18'a kadar düşmüştür. Eğitim sürecinin sonunda videolardan elde edilen her bir çerçeve için 15 boyutlu vektörler elde edilmiştir.

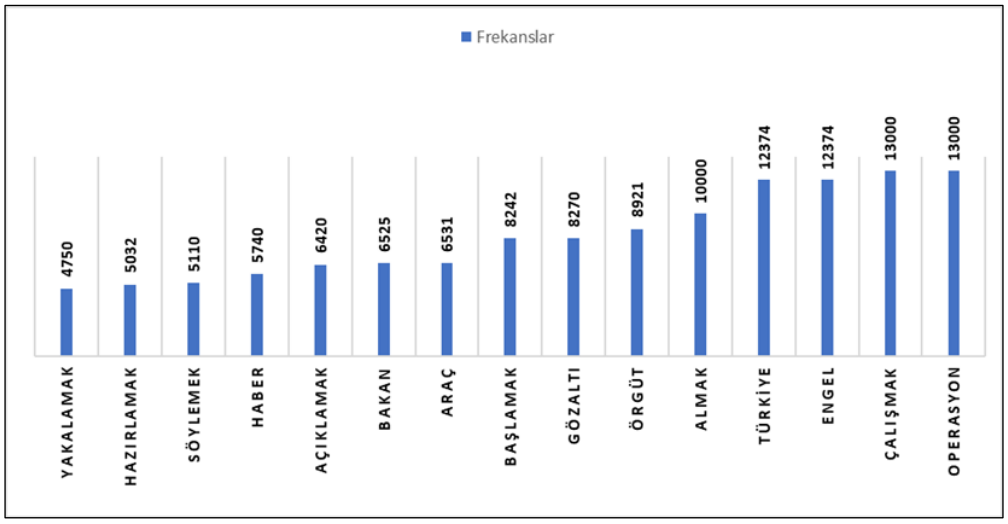


Şekil 3.18: P5 için ESA eğitimi kayıp değeri grafiği.

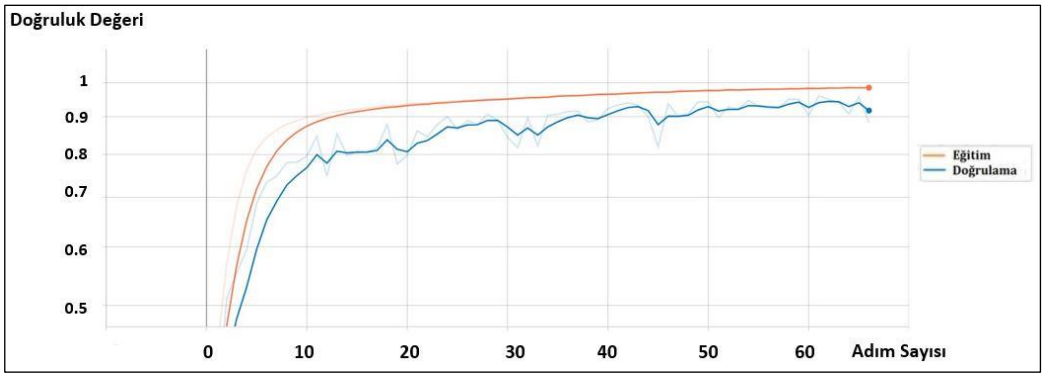
Eğitim tamamlandıktan sonra en yüksek doğruluğa ve en düşük kayıp değerine sahip modeli seçilerek UKSB modeline verilmeye uygun biçimde vektörler üretilmiştir.

3.2.3.2. P6 İçin Veri Kümesi ve Öznitelik Çıkarımı

Beş sınıf için toplam 115.530 adet görüntü bulunmaktadır. En küçük frekansa sahip sınıf sayısı göz önüne alınarak diğer sınıfların frekansları düzenlenmiştir. 115.530 görüntünün %70'i eğitim kümesi, %15'i doğrulama kümesi olarak, geriye kalan %15'i ise test kümesi olarak ayrılmıştır. Sınıf başına düşen görüntü sayısı sabit değildir. Maksimum sınıf frekansı minimum frekansa sahip sınıfın iki katından fazla olmayacak biçimde ayarlanmıştır ve Şekil 3.19'da gösterilmiştir.



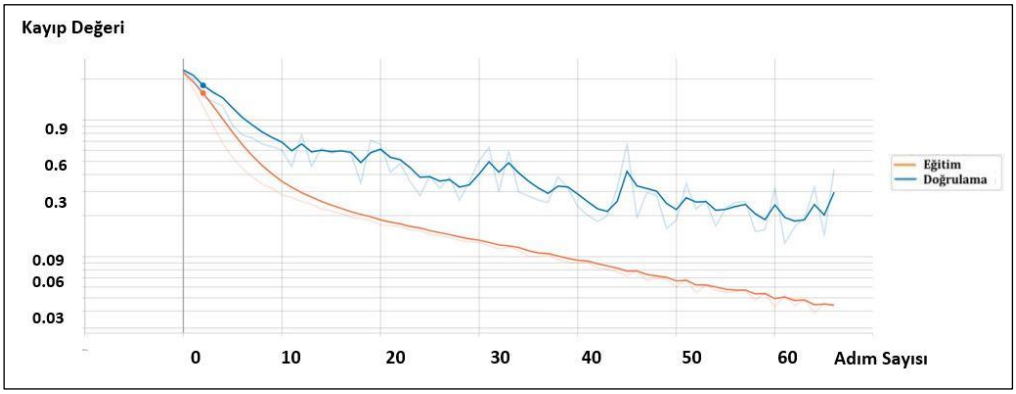
Şekil 3.19: P6 için frekans grafiği.



Şekil 3.20: P6 için ESA eğitimi doğruluk fonksiyonu grafiği.

Eğitim süreci yaklaşık olarak 43 saat sürmüştür. Eğitim sürecinin sonlarına doğru Şekil 3.20'de görüldüğü üzere doğruluk oranı yaklaşık olarak eğitim kümesi için %98'e, doğrulama kümesi için %88'e ulaşmıştır. Kayıp değerinin değeri ise Şekil 3.21'te görüldüğü üzere eğitim süreci ilerledikçe eğitim kümesi için %04, doğrulama kümesi için %20'a kadar düşmüştür. Eğitim sürecinin sonunda videolardan elde edilen her bir çerçeve için 15 boyutlu vektörler elde edilmiştir.

Eğitim tamamlandıktan sonra en yüksek doğruluğa ve en düşük kayıp değerine sahip modeli seçilerek UKSB modeline verilmeye uygun biçimde vektörler üretilmiştir.



Şekil 3.21: P6 için ESA eğitimi kayıp değeri grafiği.

3.3. Uzun-Kısa Süreli Bellek ile Eğitim

Bu tez kapsamında ESA'nın zaman serisi verilerindeki sınırlı çalışma yeteneği nedeniyle zamansal veriyi daha iyi işleyebilmek için Uzun-Kısa Süreli Bellek (UKSB) tabanlı model kullanılmıştır. Çalışmanın bu bölümünde veri kümesi olarak, ESA kullanılarak öznitelik çıkarma işlemiyle elde edilen, her bir çerçeve için sınıf sayısı kadar boyutu olan vektörler kullanılmıştır. Eğitilecek olan sınıflar ve bu sınıfların frekansları ESA modelinin eğitim sürecinde belirlenmiştir. Aynı sınıf sayısına sahip hem dengeli hem dengesiz veri kümesiyle ayrı eğitim yapılmıştır. UKSB modeli zamansal bilgiyi işleyebilen yapay bir yinelemeli sinir ağı yapısıdır. Öznitelik çıkarıma işlemiyle elde edilen veri düzgün bir biçimde sınıflara ayrılarak klasörlere ayrıldığı için, bu veri yapısına uygun bir jeneratör yazılarak etiketler oluşturulmuştur. Tüm problemlerin eğitim sürecinde kullanılan makina Dual Intel Xeon Processor E5-2640

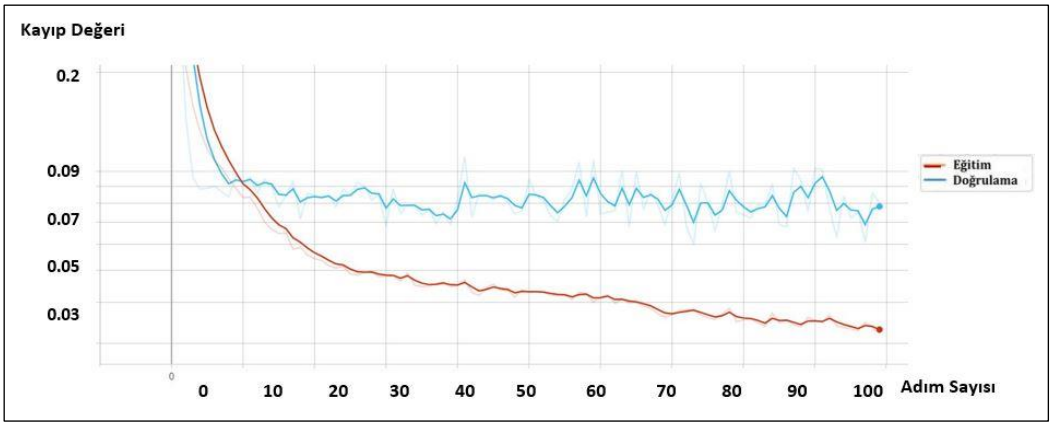
v4 (10C, 2.4GHz,3.4GHz Turbo,2133MHz, 25MB,90W,8GB NVDIA Quadro M4000) ‘dir. UKSB modelinde 3 katmanlı basit bir mimari kullanılmıştır. Yapılmış olan 6 adet eğitimin her biri için aynı model mimarisi kullanılmıştır.

3.3.1. P1 ve P2 İçin Uksb Eğitimi

Çalışmanın bu bölümünde 111 adet videonun içerisinde geçen 5 farklı kelime için oluşturulan veri kümesiyle yapılan ESA modelinin eğitimi sonucu elde edilen yeni veri kümesi kullanılarak 5 sınıf için UKSB modeli ile hem dengeli hem dengesiz veri kümesi için eğitim yapılmıştır. Bu sınıflar sırasıyla gözaltı, örgüt, çalışmak, operasyon ve Türkiye kelimeleridir.

3.3.1.1. P1 İçin Uksb Eğitimi

Beş sınıf için toplam 41.005 adet görüntü bulunmaktadır. En küçük frekansa sahip sınıf sayısı göz önüne alınarak diğer sınıfların frekansları düzenlenmiştir. 41.005 görüntünün %70’i eğitim kümesi, %15’i doğrulama kümesi olarak, geriye kalan %15’i ise test kümesi olarak ayrılmıştır. Sınıf başına düşen görüntü sayısı 8.201 adettir. Eğitim süreci yaklaşık olarak yarım saat sürmüştür.

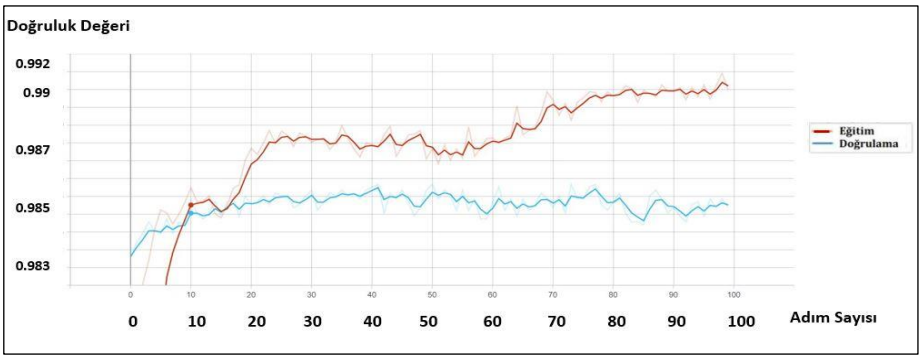


Şekil 3.22: P1 için UKSB eğitimi doğruluk fonksiyonu grafiği.

Eğitim sürecinin sonlarına doğru Şekil 3.22’de görüldüğü üzere doğruluk oranı yaklaşık olarak eğitim kümesi için %99’e, doğrulama kümesi için %98’e ulaşmıştır.

Kayıp değerinin değeri ise Şekil 3.23'te görüldüğü üzere eğitim süreci ilerledikçe eğitim kümesi için %03, doğrulama kümesi için %08'a kadar düşmüştür.

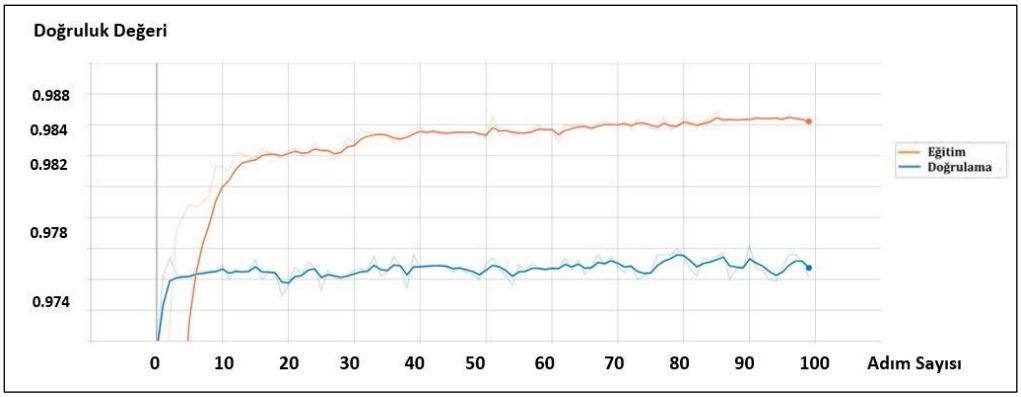
Eğitim tamamlandıktan sonra en yüksek doğruluğa ve en düşük kayıp değerine sahip modeli seçilerek tahminleme yapılmıştır.



Şekil 3.23: P1 için UKSB eğitimi kayıp değeri grafiği.

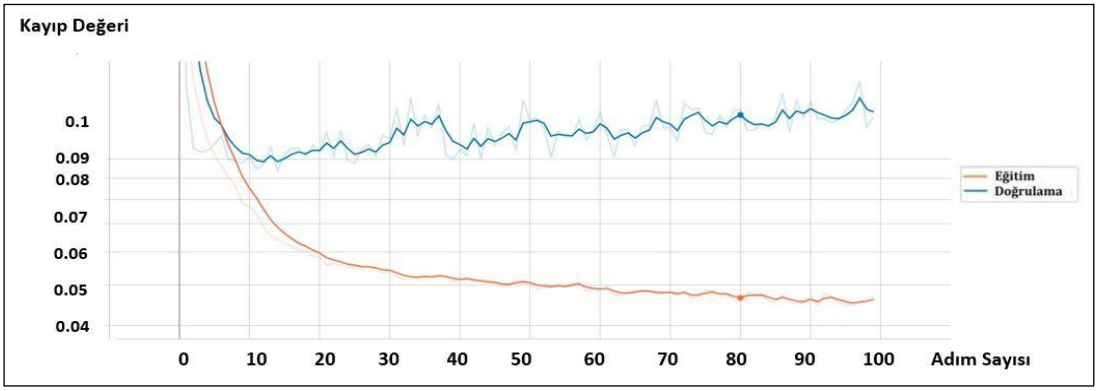
3.3.1.2. P2 İçin Uksb Eğitimi

Beş sınıf için toplam 58.966 adet görüntü bulunmaktadır. En küçük frekansa sahip sınıf sayısı göz önüne alınarak diğer sınıfların frekansları düzenlenmiştir. 58.966 görüntünün %70'i eğitim kümesi, %15'i doğrulama kümesi olarak, geriye kalan %15'i ise test kümesi olarak ayrılmıştır. Sınıf başına düşen görüntü sayısı sabit değildir. ESA modeli eğitilirken belirlendiği üzere, maksimum sınıf frekansı minimum frekansa sahip sınıfın iki katından fazla olmayacak biçimde ayarlanmıştır. Eğitim süreci yaklaşık olarak yarım saat sürmüştür. Eğitim sürecinin sonlarına doğru Şekil 3.24'te görüldüğü üzere doğruluk oranı yaklaşık olarak eğitim kümesi için %98'e, doğrulama kümesi için %97'e ulaşmıştır. Kayıp değeri ise Şekil 3.25'te görüldüğü üzere eğitim süreci ilerledikçe eğitim kümesi için %04, doğrulama kümesi için %12'a kadar düşmüştür.



Şekil 3.24: P2 için UKSB eğitimi doğruluk fonksiyonu grafiği.

Eğitim tamamlandıktan sonra en yüksek doğruluğa ve en düşük kayıp değerine sahip modeli seçilerek tahminleme yapılmıştır.



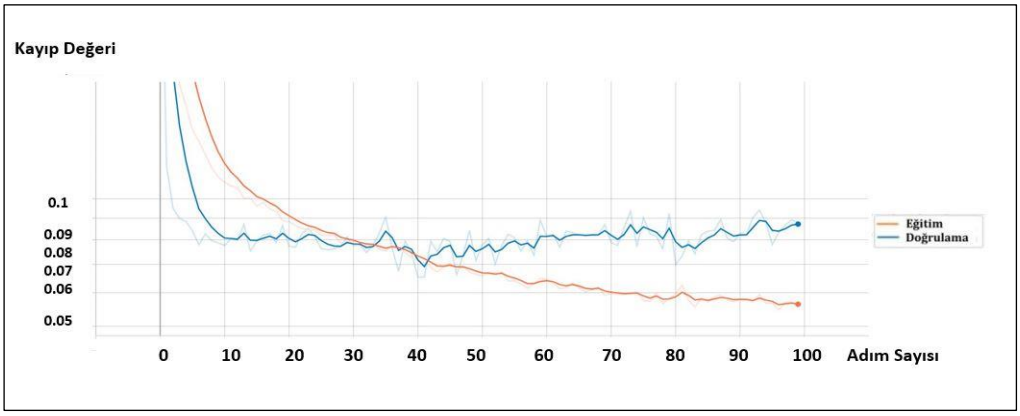
Şekil 3.25: P2 için UKSB eğitimi kayıp değeri grafiği.

3.3.2. P3 ve P4 İçin Uksb Eğitimi

Çalışmanın bu bölümünde 111 adet videonun içerisinde geçen 10 farklı kelime için oluşturulan veri kümesiyle yapılan ESA modelinin eğitimi sonucu elde edilen yeni veri kümesi kullanılarak 10 sınıf için UKSB modeli ile hem dengeli hem dengesiz veri kümesi için eğitim yapılmıştır. Bu sınıflar sırasıyla gözaltı, örgüt, çalışmak, operasyon, Türkiye, açıklamak, bakan, araç, başlamak, engel kelimeleridir.

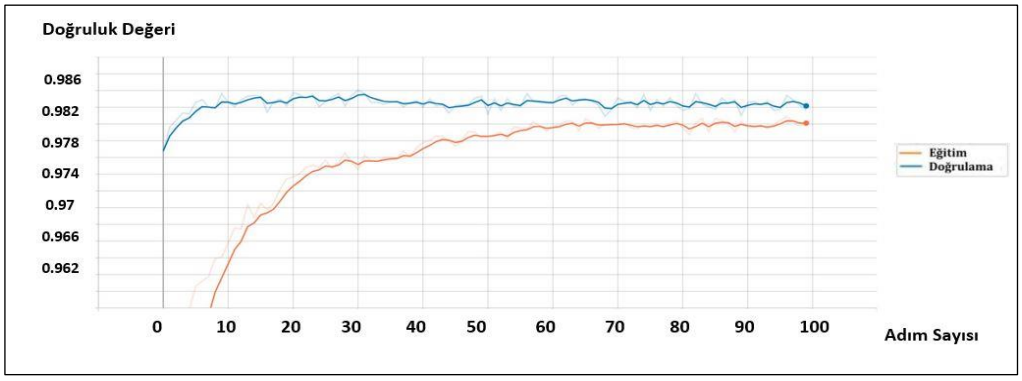
3.3.2.1. P3 İçin Uksb Eğitimi

Beş sınıf için toplam 64.100 adet görüntü bulunmaktadır. En küçük frekansa sahip sınıf sayısı göz önüne alınarak diğer sınıfların frekansları düzenlenmiştir. 64.100 görüntünün %70'i eğitim kümesi, %15'i doğrulama kümesi olarak, geriye kalan %15'i ise test kümesi olarak ayrılmıştır. Sınıf başına düşen görüntü sayısı 6.410 adettir.



Şekil 3.26: P3 için UKSB eğitimi kayıp değeri grafiği.

Eğitim süreci yaklaşık olarak bir saat sürmüştür. Eğitim sürecinin sonlarına doğru Şekil 3.26'da görüldüğü üzere doğruluk oranı yaklaşık olarak eğitim kümesi için %98'e, doğrulama kümesi için %97'e ulaşmıştır. Kayıp değerinin değeri ise Şekil 3.27'de görüldüğü üzere eğitim süreci ilerledikçe eğitim kümesi için %05, doğrulama kümesi için %08'a kadar düşmüştür.



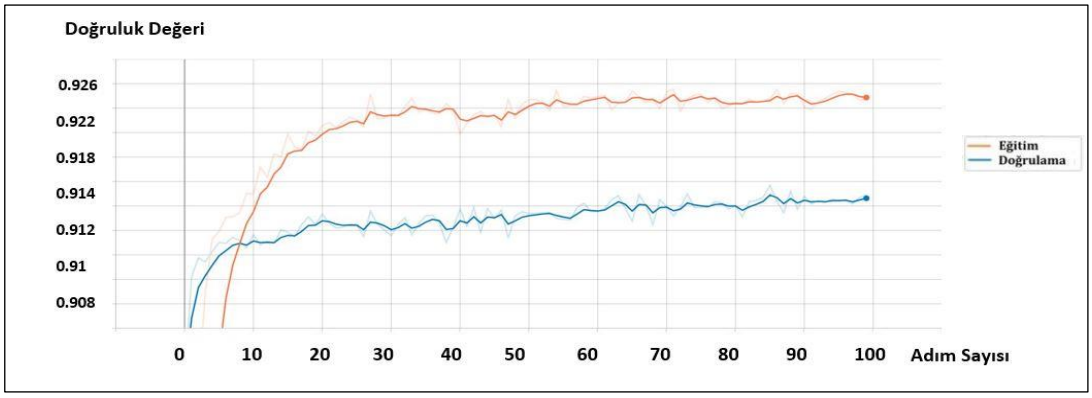
Şekil 3.27: P3 için UKSB eğitimi doğruluk fonksiyonu grafiği.

Eğitim tamamlandıktan sonra en yüksek doğruluğa ve en düşük kayıp değerine sahip modeli seçilerek tahminleme yapılmıştır.

3.3.2.2. P4 İçin Uksb Eğitimi

Beş sınıf için toplam 96.283 adet görüntü bulunmaktadır. En küçük frekansa sahip sınıf sayısı göz önüne alınarak diğer sınıfların frekansları düzenlenmiştir. 96.283 görüntünün %70'i eğitim kümesi, %15'i doğrulama kümesi olarak, geriye kalan %15'i ise test kümesi olarak ayrılmıştır. Sınıf başına düşen görüntü sayısı sabit değildir. ESA modeli eğitilirken belirlendiği üzere, maksimum sınıf frekansı minimum frekansa sahip sınıfın iki katından fazla olmayacak biçimde ayarlanmıştır. Eğitim süreci yaklaşık olarak bir saat sürmüştür. Eğitim sürecinin sonlarına doğru Şekil 3.28'de görüldüğü üzere doğruluk oranı eğitim kümesi için yaklaşık olarak %92'e, doğrulama kümesi için %91'e ulaşmıştır.

Kayıp değerinin değeri ise Şekil 3.29'da görüldüğü üzere eğitim süreci ilerledikçe eğitim kümesi için %25, doğrulama kümesi için %29'a kadar düşmüştür. Eğitim tamamlandıktan sonra en yüksek doğruluğa ve en düşük kayıp değerine sahip modeli seçilerek tahminleme yapılmıştır.

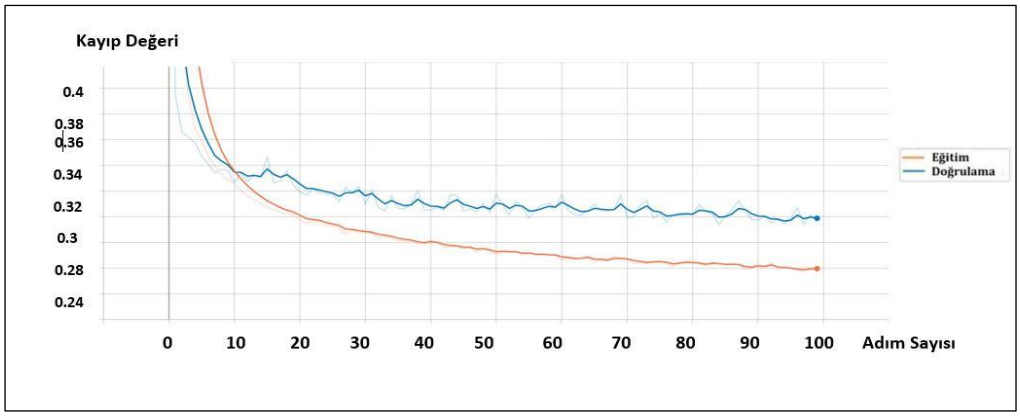


Şekil 3.28: P4 için UKSB eğitimi doğruluk fonksiyonu grafiği.

Kayıp değerinin değeri ise Şekil 3.29'da görüldüğü üzere eğitim süreci ilerledikçe eğitim kümesi için %25, doğrulama kümesi için %29'a kadar düşmüştür. Eğitim tamamlandıktan sonra en yüksek doğruluğa ve en düşük kayıp değerine sahip modeli seçilerek tahminleme yapılmıştır.

3.3.3. P5 ve P6 İçin Uksb Eğitimi

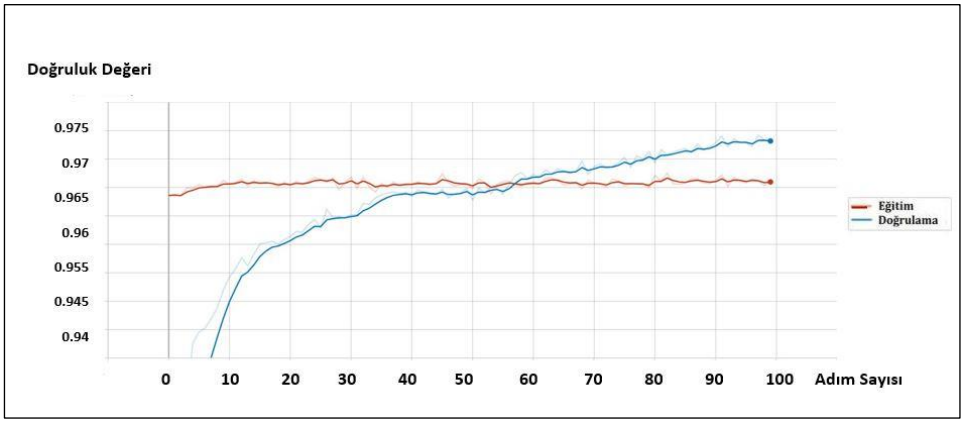
Çalışmanın bu bölümünde 111 adet videonun içerisinde geçen 15 farklı kelime için oluşturulan veri kümesiyle yapılan ESA modelinin eğitimi sonucu elde edilen yeni veri kümesi kullanılarak 15 sınıf için UKSB modeli ile hem dengeli hem dengesiz veri kümesi için eğitim yapılmıştır. Bu sınıflar sırasıyla gözaltı, örgüt, çalışmak, operasyon, Türkiye, açıklamak, bakan, araç, başlamak, engel, almak, haber, hazırlamak, söylemek, yakalamak kelimeleridir.



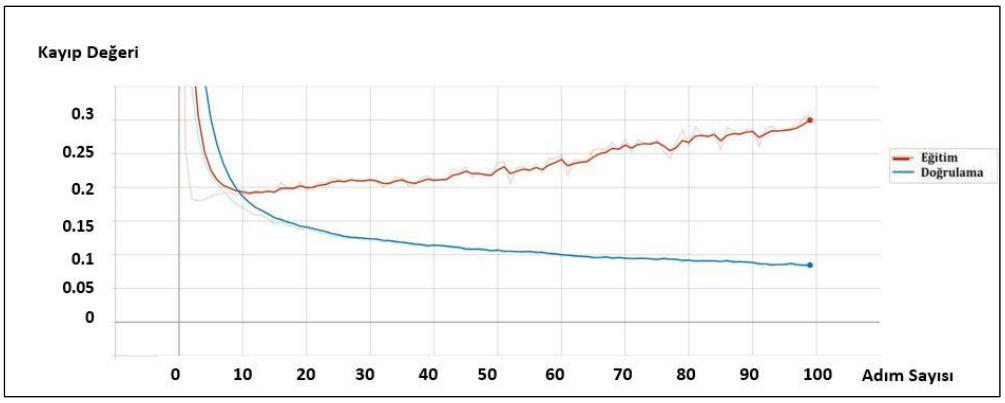
Şekil 3.29: P4 için UKSB eğitimi kayıp değeri grafiği.

3.3.3.1. P5 İçin Uksb Eğitimi

Beş sınıf için toplam 71.235 adet görüntü bulunmaktadır. En küçük frekansa sahip sınıf sayısı göz önüne alınarak diğer sınıfların frekansları düzenlenmiştir. 71.235 görüntünün %70'i eğitim kümesi, %15'i doğrulama kümesi olarak, geriye kalan %15'i ise test kümesi olarak ayrılmıştır. Sınıf başına düşen görüntü sayısı 4.748 adettir. Eğitim süreci yaklaşık olarak 1 saat sürmüştür. Eğitim sürecinin sonlarına doğru Şekil 3.30'da görüldüğü üzere doğruluk oranı yaklaşık olarak eğitim kümesi için %97'e, doğrulama kümesi için %96'e ulaşmıştır. Kayıp değerinin değeri ise Şekil 3.31'de görüldüğü üzere eğitim süreci ilerledikçe eğitim kümesi için %13, doğrulama kümesi için %24'a kadar düşmüştür.



Şekil 3.30: P5 için UKSB eğitimi doğruluk fonksiyonu grafiği.

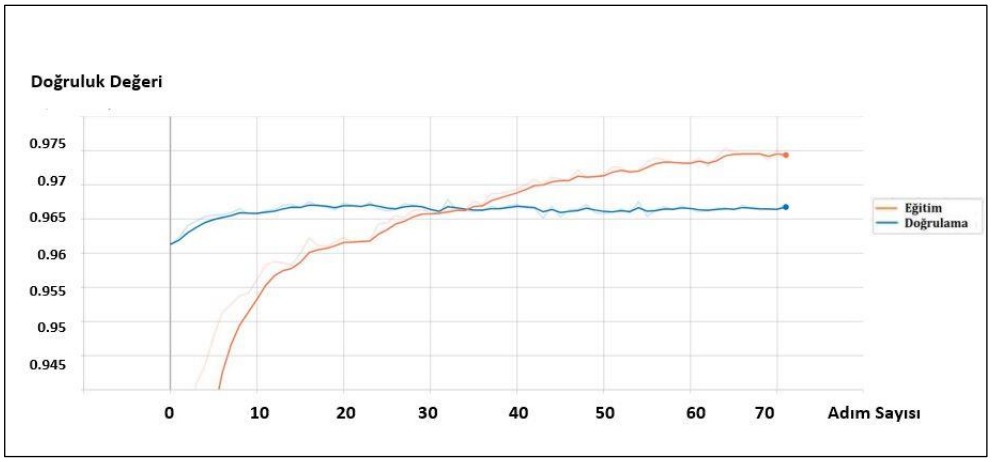


Şekil 3.31: P5 için UKSB eğitimi kayıp değeri grafiği.

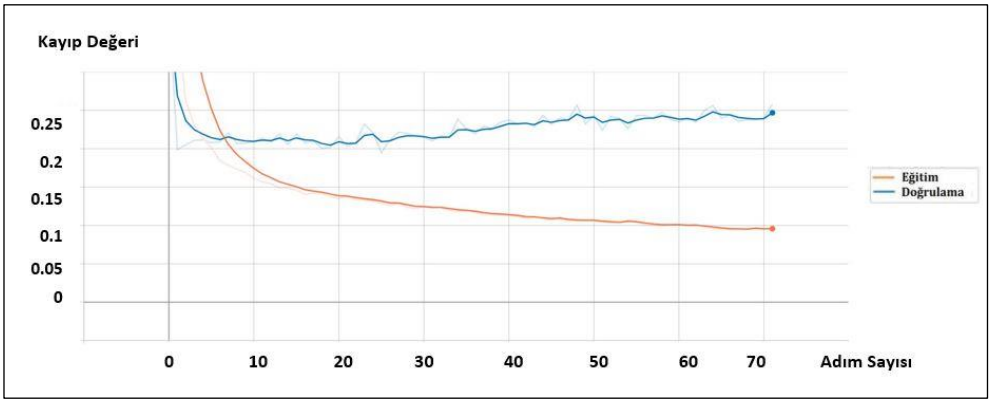
3.3.3.2. P6 İçin Uksb Eğitimi

Beş sınıf için toplam 115.530 adet görüntü bulunmaktadır. En küçük frekansa sahip sınıf sayısı göz önüne alınarak diğer sınıfların frekansları düzenlenmiştir. 115.530 görüntünün %70'i eğitim kümesi, %15'i doğrulama kümesi olarak, geriye kalan %15'i ise test kümesi olarak ayrılmıştır. Sınıf başına düşen görüntü sayısı sabit değildir. ESA modeli eğitilirken belirlendiği üzere, maksimum sınıf frekansı minimum frekansa sahip sınıfın iki katından fazla olmayacak biçimde ayarlanmıştır.

Eğitim süreci yaklaşık olarak bir saat sürmüştür. Eğitim sürecinin sonlarına doğru Şekil 3.32de görüldüğü üzere doğruluk oranı eğitim kümesi için yaklaşık olarak %97'e, doğrulama kümesi için %96'e ulaşmıştır. Kayıp değerinin değeri ise Şekil 3.33'te görüldüğü üzere eğitim süreci ilerledikçe eğitim kümesi için %09, doğrulama kümesi için %25'a kadar düşmüştür.



Şekil 3.32: P6 için UKSB eğitimi doğruluk fonksiyonu grafiği.



Şekil 3.33: P6 için UKSB eğitimi kayıp değeri grafiği.

4. DENEYSEL ÇALIŞMALAR

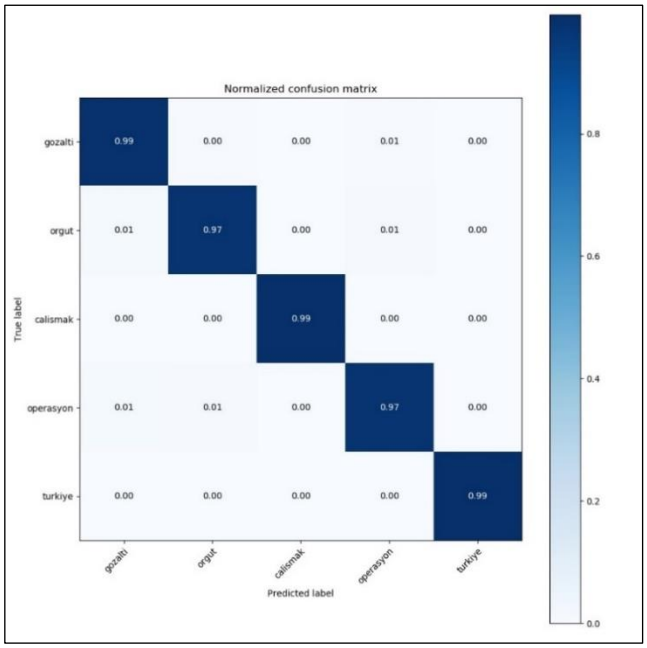
Modellerin başarısını değerlendirmek için bazı yöntemler mevcuttur. Bu bölümde 3 farklı eğitim kümesiyle yapılan toplam 6 adet eğitimin sonuçlarına yer verilmiştir. Bu değerlendirmeler her bir model için bölüm 3.1'de oluşturulan veri setlerinin test bölümleri kullanılarak elde edilmiştir. Değerlendirme aşamasında modele göre girdi, veri kümesinin test için ayrılmış kısmındaki çerçeveler çıktısı ise mevcut çerçevenin sınıfıdır. Değerlendirme için karmaşıklık matrisi kullanılmıştır. Karmaşıklık matrisinde köşegen elemanları doğru tahminleri, matrisin diğer elemanları ise yanlış tahminleri içerir. Matrisin diyagonal elemanlarının dışındaki elemanlarının sıfırdan farklı olması sınıflandırmada bir karışıklık olduğu anlamı taşır. Böylelikle karışıklığın bulunduğu değerin sütun ve satır kesişimlerine bakılarak eğitilen model hakkında yorum yapma imkânı elde edilir. Aşağıda oluşturulan 3 farklı sınıf sayısı için oluşturulan 6 adet karmaşıklık matrisinin karşılaştırılması verilmiştir.

4.1. P1 ve P2 İçin Karmaşıklık Matrisi

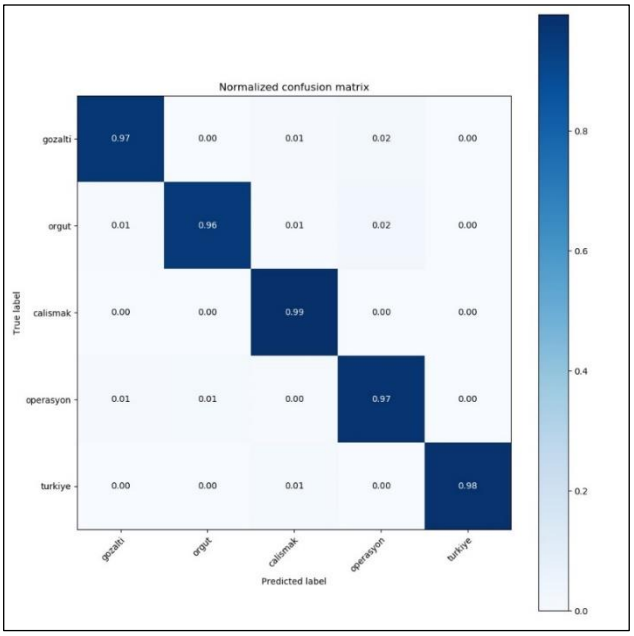
111 adet videonun içerisinde bulunan 5 farklı kelime sınıfından oluşan veri kümesinin %15'i test aşamasında kullanılmak üzere ayrılmıştır. Dengeli veri kümesi için 6.150 adet veri karmaşıklık matrisinde kullanılmıştır. Dengesiz veri kümesinin karmaşıklık matrisinde kullanılan veri adeti 8.845'tir. Her iki durum için de model bu veriyle daha önce hiç karşılaşmamıştır. İki durum için de hesaplanan karmaşıklık matrisinin normalize edilmiş hali aşağıda verilmiştir.

Her iki karmaşıklık matrisinde diyagonal elemanlara bakıldığında oranın çok yüksek olduğu görülmektedir. Buna rağmen aralarında bazı farklılıklar tespit edilmiştir. Örneğin gözaltı sınıfının başarı oranı %99'dan %97'ye düşmüştür. Bunun sebebi düzensiz veri kümesinde gözaltı sınıfına ait örnek sayısının azalmış olmasıdır. Örgüt sınıfının örnek sayısı düzensiz veride bir miktar artmasına rağmen başarı oranı %96'ya düşmüştür. Çalışmak ve operasyon sınıflarında düzensiz veride sayı artmasına rağmen başarı oranında bir artış görülmemiştir. Türkiye sınıfında örnek sayısı artmasına rağmen başarı oranı %99'dan %98'e düşmüştür. Dengesiz veri kümesinde bazı sınıf sayılarının adeti artmasına rağmen başarı oranında bir artış görülmemiştir.

Sonuç olarak 5 sınıflı çalışmada düzensiz veriden elde edilen karmaşıklık matrisinin başarı oranı daha düşüktür.



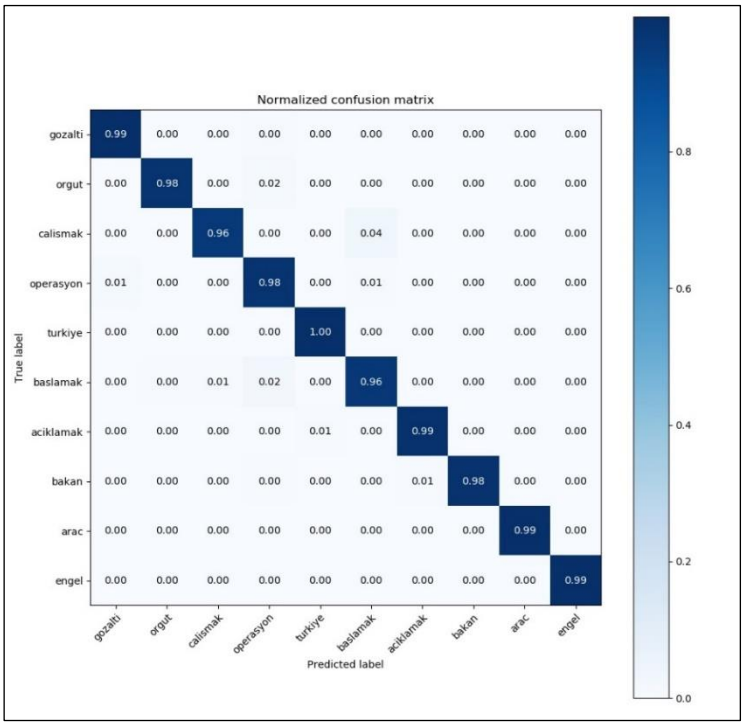
Şekil 4.1: P1 için karmaşıklık matrisi.



Şekil 4.2: P2 için karmaşıklık matrisi.

4.2. P3 ve P4 İçin Karmaşıklık Matrisi

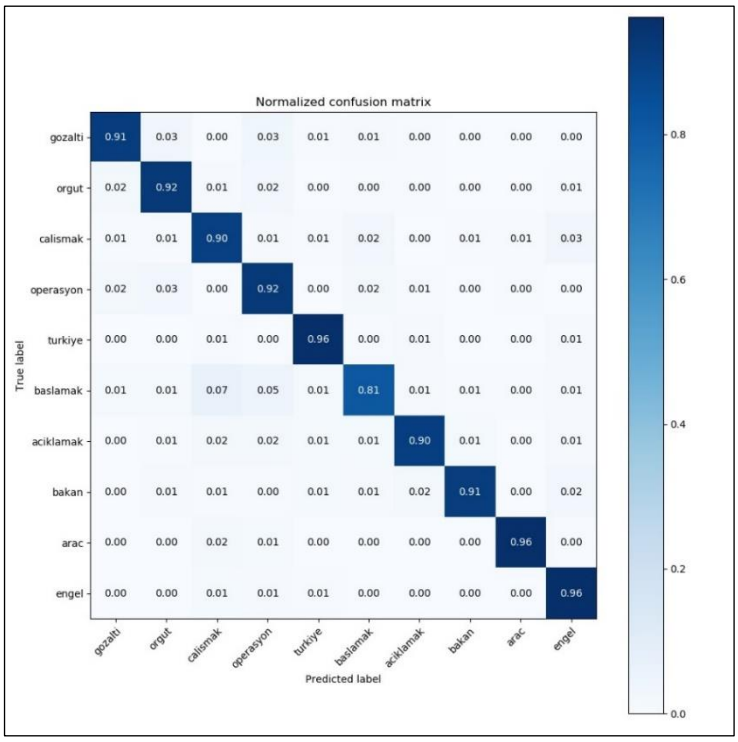
111 adet videonun içerisinde bulunan 10 farklı kelime sınıfından oluşan veri kümesinin %15'i test aşamasında kullanılmak üzere ayrılmıştır. Dengeli veri kümesi için 9.615 adet veri karmaşıklık matrisinde kullanılmıştır. Dengesiz veri kümesinin karmaşıklık matrisinde kullanılan veri adeti 14.442'dir. Her iki durum için de model bu veriyle daha önce hiç karşılaşmamıştır. İki durum için de hesaplanan karmaşıklık matrisinin normalize edilmiş hali aşağıda verilmiştir.



Şekil 4.3: P3 için karmaşıklık matrisi.

Her iki karmaşıklık matrisinde diyagonal elemanlara bakıldığında oranın çok yüksek olduğu görülmektedir. Buna rağmen aralarında bazı farklılıklar bulunmaktadır. Örneğin dengeli veri kümesinde gözaltı sınıfının başarı oranı %99'dur fakat dengesiz veride bu oran %91'e düşmüştür. Bunun sebebi düzensiz veri kümesinde gözaltı sınıfına ait örnek sayısının azalmış olmasıdır. Benzer şekilde örgüt sınıfının başarı oranı da %99'dan %91'e düşmüştür. Türkiye, çalışmak ve operasyon sınıflarının miktarı her iki veri kümesinde de değişmemesine rağmen başarı oranı dengesiz veri kümesinde düşmüştür. Dengesiz veri kümesinde başlamak, açıklamak ve bakan sınıflarında başarı oranı çok yüksek oranda düşmüştür. Bu veri kümesinde bu 3 sınıfın

frekans sayısında da ciddi bir azalma olmuştur. Başarı oranının düşmesi bununla ilişkilidir. Dengesiz veri kümesinde sadece engel sınıfının frekans sayısında artış olmuştur. Bu sınıfın başarı oranı da nispeten daha az azalmıştır.



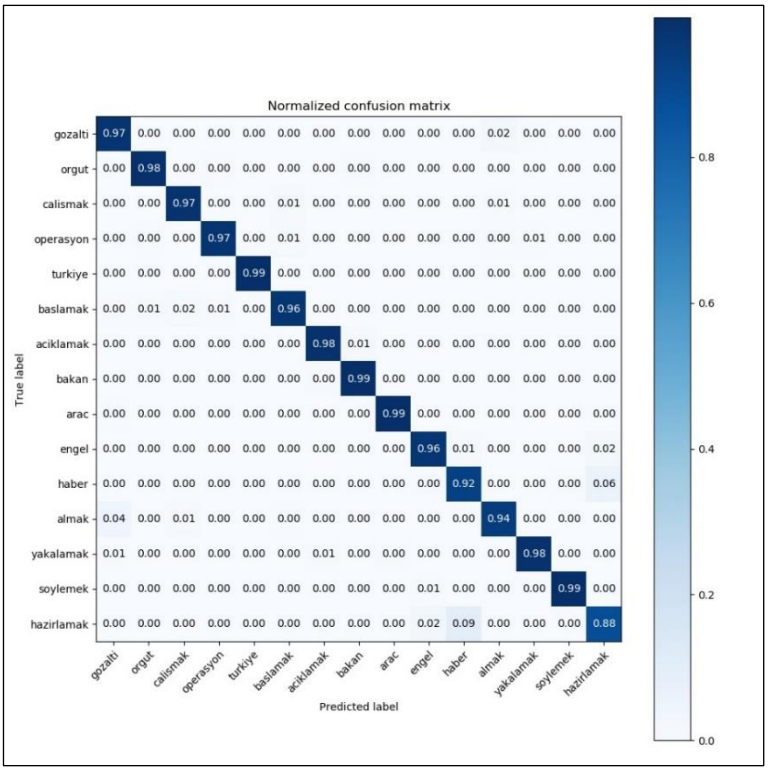
Şekil 4.4: P4 için karmaşıklık matrisi.

10 sınıf için oluşturulan dengesiz veri kümesinde bir sınıf hariç diğer sınıfların frekans sayıları ya azalmış ya da aynı kalmıştır. Bölüm 4.1 de bulunan 5 sınıflı veri kümesine oranla bu veri kümesinde başarı oranı daha da azalmıştır. Sonuç olarak 10 sınıflı çalışmada düzensiz veriden elde edilen karmaşıklık matrisinin başarı oranı daha düşüktür.

4.3. P5 ve P6 için Karmaşıklık Matrisi

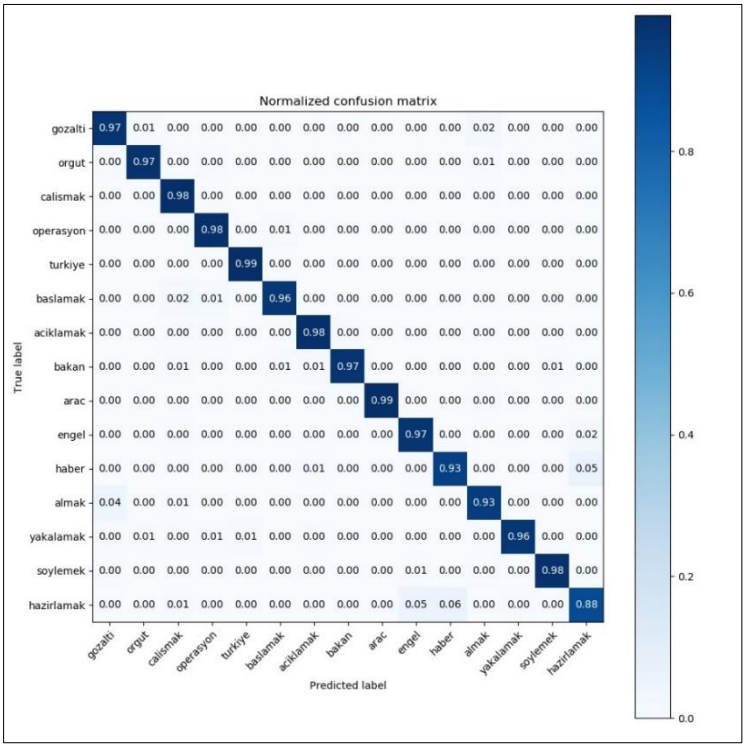
111 adet videonun içerisinde bulunan 15 farklı kelime sınıfından oluşan veri kümesinin %15'i test aşamasında kullanılmak üzere ayrılmıştır. Dengeli veri kümesi için 10.685 adet veri karmaşıklık matrisinde kullanılmıştır. Dengesiz veri kümesinin karmaşıklık matrisinde kullanılan veri adeti 17.330'dur. Her iki durum için de model

bu veriyle daha önce hiç karşılaşmamıştır. İki durum için de hesaplanan karmaşıklık matrisinin normalize edilmiş hali aşağıda verilmiştir.



Şekil 4.5: P5 için karmaşıklık matrisi.

Her iki karmaşıklık matrisi için diyagonal elemanlara bakıldığında oranın çok yüksek olduğu görülmektedir. 15 sınıf için oluşturulan ve dengeli ve dengesiz veri setlerinden elde edilen karmaşıklık matrislerinin başarı oranı 5 ve 10 sınıflı veri setlerinden elde edilen karmaşıklık matrislerinin başarı oranlarından daha yüksektir. Aralarındaki başarı oranları birbirine çok yakındır. Birçok sınıfta frekans sayısı değişmesine rağmen her iki matris içinde başarı oranı aynı kalmıştır. Gözaltı, Türkiye, başlamak, açıklamak araç, hazırlamak sınıflarının başarı yüzdelerinde herhangi bir değişim görülmemiştir. Buna rağmen bazı sınıflarda farklılıklar bulunmaktadır. Örneğin dengeli veri kümesinde örgüt sınıfının başarı oranı %98'dir fakat dengesiz veride bu oran %97'ye düşmüştür. Bunun sebebi düzensiz veri kümesinde gözaltı sınıfına ait örnek sayısının azalmış olmasıdır. Benzer şekilde bakan sınıfının başarı oranı da %99'dan %97'e düşmüştür. Fakat bu veri kümesinde dengeli veri kümesinden daha fazla öğrenen sınıflarda olmuştur. Bu sınıflar haber, engel, operasyon ve çalışmak sınıflarıdır.



Şekil 4.6: P6 için karmaşıklık matrisi.

15 sınıf için oluşturulan dengesiz veri kümesinde oluşturulan diğer veri setlerinden farklı olarak, 2 sınıf hariç diğer sınıfların frekans sayıları ya artmış ya da aynı kalmıştır. Bundan dolayı 5 ve 10 sınıflı veri setlerine oranla bu veri kümesinde başarı oranı daha da artmıştır. Sonuç olarak 15 sınıflı çalışmada düzensiz veriden elde edilen karmaşıklık matrisinin başarı oranı diğer veri setlerine oranla daha yüksektir.

5.SONUÇ ve GELECEK ÇALIŞMALAR

Bu tez kapsamında, TİD için zayıf güdümlü yöntem kullanarak veri kümesi oluşturma ve oluşturulan veri kümesiyle TİD'in farklı sınıf sayıları için sınıflandırması ele alınmıştır. Böylelikle problemin çözümü için gerekli olan veri kümesi oluşturma kısmı algoritmik olarak toplanıp modellemeye hazır hale getirilmiştir. Bu yöntemi, çalışmanın maliyetinin azalmasına büyük katkı sağlamıştır. Veri kümesi oluşturulduktan sonra YSA modeli öne sürülmüştür.

Bu bağlamda öncelikle veri kümesinin oluşturulabilmesi için çevrim içi platformlardan videolar bulunmuştur. Kullanılan video adedi 111'dir. Bu videolar çeşitli görüntü işleme kütüphaneleriyle düzenlenmiştir. Videoların içerisinde toplam 2 adet işaretçi bulunmaktadır. Düzenleme sonucunda videoların içerisinde TİD'in geçtiği kısımlar görüntü olarak kullanılmak üzere alınmıştır. Daha sonra videolarda geçen her bir kelime için alt yazı dosyaları elde edilmeye çalışılmıştır. Bunun için farklı uygulamalar denenmiştir. İlk denenen uygulama PyTranscriber [28] uygulamasıdır. Bu uygulama sonucunda alt yazı dosyaları elde edilebilse de her bir kelime için ulaşılan zaman bilgisinin doğruluk oranı yüksek değildir. Bundan dolayı Google'ın Speech to Text API'yı [30] denenmiştir. Bunun sonucunda video içerisindeki her bir kelime için, başlangıç ve bitiş zaman bilgisi elde edilmiştir. Böylelikle her bir kelime geçtiği videodan alınarak düzenli bir şekilde klasörlere koyulmuş bu şekilde veri kümesi oluşturulmuştur.

Veri kümesi 111 videodan elde edilen 510 adet farklı kelimeden oluşturulmuştur. Bu kelimeler TİD'de mevcut olan kelimelerdir. Konuyla ilgili yapılan önceki çalışmalarda oluşturulan veri kümelerinde çoğunlukla bir kişilere, farklı arka planlara ve bunları kaydedebilmek için cihaza ihtiyaç duyulmaktadır. Bu çalışmada ise oluşturulan veri kümesi çevrim içi platformlardan algoritmik olarak elde edildiği için böyle bir ihtiyaç yoktur. Bu duruma ek olarak ihtiyaç duyulduğu takdirde oluşturulan veri kümesi yine çevrim içi platformlar kullanılarak güncellenebilir. Veri kümesini oluştururken bu platformların kullanılması çalışmanın gerçek hayata uygulanabilirliği açısından özgünlük içermektedir. Ayrıca veri kümesi önceden belirlenmiş sabit bir sınıf sayısı için oluşturulmadığından, istenildiği takdirde sınıf sayısı değiştirilerek farklı problemler için eğitim gerçekleştirilebilir.

Çalışmanın ikinci bölümünde oluşturulan veri kümesiyle farklı problemlerin çözümü ele alınmıştır. Toplam 6 adet problem bulunmaktadır. Bu problemler 3 farklı sınıf sayısında dengeli ve dengesiz veri kümesi için TİD'in tanınması problemidir. Her bir problem için öncelikle ESA kullanılarak öznitelik çıkarımı sonrasında elde edilen yeni veri kümesiyle UKSB yöntemi kullanılarak sınıflandırma yapılmıştır. Tüm problemler için aynı modeller kullanılmıştır. ESA'da kullanılan model DenseNet [31] mimarisine benzer bir mimaridir. UKSB de ise 3 katmanlı basit bir mimari kullanılmıştır.

Öncelikle 5 sınıf için veri kümesi için 2 adet veri kümesi oluşturulmuştur. Oluşturulan ilk veri kümesi dengeli ikincisi dengesiz bir veri kümesidir. Her iki veri kümesi de aynı modeller tarafından eğitilmiştir. Sınıflandırma işlemi sonunda her iki veri kümesi de başarılı sonuçlar vermiş olmasına rağmen dengeli veri kümesinin karmaşıklık matrisinin sonuçları dengesize göre daha başarılıdır. Sonrasında veri kümesi 5 sınıf daha arttırılarak 10 sınıflı 2 adet veri kümesi oluşturulmuş ve aynı eğitim sürecine tabii tutulmuştur. Sınıflandırma sonucunda dengeli veri kümesinin başarı oranı daha yüksektir. Ardından 15 sınıf için oluşturulan iki farklı veri kümesi için eğitim gerçekleştirilmiş ve benzer sonuçlar elde edilmiştir.

Sınıf sayıları arasındaki başarı oranı mukayese edildiğinde sınıf sayısı arttıkça başarı oranında artış görülmüştür. Buna rağmen video akışında ardışık olan kelimelerin sınıflandırmasında çeşitli karışıklıklar gözlemlenmiştir.

Bu tez kapsamında görüldüğü üzere derin öğrenme yöntemlerinde kullanılmak üzere yeni bir veri kümesi oluşturmak mümkündür. Gelecek çalışmalarda TİD için oluşturulan bu veri kümesi benzer yöntem kullanılarak arttırılabilir. Veri kümesi arttırılırken daha çok işaretçi ve farklı arka planlar kullanılabilir. Ardışık olduğu için birbiriyle karışan kelimelerden daha çok veri toplanarak bu karmaşıklık giderilebilir. Mevcut çalışmada DenseNet [31] mimarisine benzer bir mimari kullanılmıştır. Ek bir çalışmayla farklı mimari için veri kümeleri eğitilerek modellerin başarısı karşılaştırılabilir.

Bu çalışma kapsamında zayıf güdümlü bir yöntem kullanarak oluşturulan veri kümesinin tamamı modelleme aşamasında kullanılmamıştır. Buna rağmen başarı oranı oldukça yüksektir. Gelecek çalışmalarda eğitim verisi genişletilerek modelleme yapılabilir. Bu yaklaşım TİD içerisinde bulunan bütün ifade ve kelimeler için uygulanarak TİD'in gelişmesine katkı sağlayabilir.

KAYNAKLAR

[1] Oral A. Z., (2016), "Türk işaret dili çevirisi", Siyasal Kitabevi.

[2] Oktekin B., Nadire Ç., (2019)"İşitme ve Konuşma Engelli Bireyler için İşaret Tanıma Sistemi Geliştirme." Folklor/Edebiyat 25.97: 575-590.

[3] Web 1, (2022) https://www.aile.gov.tr/media/105389/eyhgm_istatistik_bulteni_mart_2022.pdf (Erişim Tarihi: 24/05/2022)

[4] Erkuş M., (2020), "Türk işaret dilinde kelime tabanlı derin öğrenme uygulaması", Yüksek Lisans Tezi, Başkent Üniversitesi, Fen Bilimleri Enstitüsü.

[5] Zhu X, Andrew B. G., (2009), "Introduction to semi-supervised learning." Synthesis lectures on artificial intelligence and machine learning 3.1: 1-130.

[6] Umbaugh S. E., (2010), "Digital image processing and analysis: human and computer vision applications with CVIPtools". CRC press,

[7] Krizhevsky A., Ilya S., Geoffrey E. H., (2012), "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25.

[8] Kyle J. G., Kyle J., Woll B., Pullen G., Maddix F., (1988), "Sign language: The study of deaf people and their language". Cambridge University Press.

[9] Web 2, (2022), https://orgm.meb.gov.tr/alt_sayfalar/duyurular/1.pdf , (Erişim Tarihi: 24/05/2022)

[10] Web 3, (2022), https://isaretdilisozlugu.com/anasayfa , (Erişim Tarihi: 24/05/2022)

[11] Web 4, (2022), https://isaretce.com/, (Erişim Tarihi: 24/05/2022)

[12] Web 5, (2022), https://www.kaggle.com/datasets/grassknoted/asl-alphabet, (Erişim Tarihi: 24/05/2022)

[13] Yalçınkaya Ö., Atvar A., Duygulu P. (2016). "Turkish sign language recognition application using Motion History Image". 24th Signal Processing and Communication Application Conference (SIU) (pp. 801-804). IEEE.

[14] Makaroğlu B., Dikyuva H. (2017). Güncel Türk İşaret Dili Sözlüğü. Ankara: Aile ve Sosyal Politikalar Bakanlığı.

[15] Kın Z. B., (2019), “Türk işaret dili alfabesinin derin öğrenme yöntemi ile sınıflandırılması”, Yüksek Lisans Tezi, Başkent Üniversitesi Fen Bilimleri Enstitüsü.

[16] Beşer F., Kizrak M. A., Bolat B., Yildirim T., (2018), “Recognition of sign language using capsule networks”. 26th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

[17] Haberdar H., Albayrak S., (2005), “Real time isolated turkish sign language recognition from video using hidden markov models with global features”. In International Symposium on Computer and Information Sciences (pp. 677-687). Springer, Berlin, Heidelberg.

[18] Demircioğlu B., Bülbül G., Köse H., (2016), “Turkish sign language recognition with leap motion”. 24th Signal Processing and Communication Application Conference (SIU) (pp. 589-592). IEEE.

[19] Memiş A., Albayrak S., (2013), “Turkish Sign Language recognition using spatio-temporal features on Kinect RGB video sequences and depth maps”. 21st Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

[20] Camgöz N. C., Kındıroğlu A. A., Karabüklü S., Kelepir M., Özsoy A. S., Akarun L., (2016), “Bosphorussign: A turkish sign language recognition corpus in health and finance domains”. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (pp. 1383-1388).

[21] Kındıroğlu A. A., Oğulcan Ö., Akarun L., (2019), "Temporal accumulative features for sign language recognition." 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). IEEE.

[22] Özdemir O, Kındıroğlu A. A., Camgöz N. C., Akarun L., (2020), "BosphorusSign22k Sign Language Recognition Dataset." 9th Workshop on the Representation & Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives, Language Resources and Evaluation Conference 2020.

[23] Gökçe Ç., Özdemir O., Kındıroğlu A. A., Akarun L., (2020), “Score-level Multi Cue Fusion for Sign Language Recognition”. arXiv preprint arXiv:2009.14139.

[24] Sincan O. M., Keles H. Y., (2020), “AUTSL: A Large Scale Multi-Modal Turkish Sign Language Dataset and Baseline Methods”. IEEE Access, 8, 181340-181355.

[25] Türkmen B., Genç Y., (2020), "Handwriting Recognition via Visual Observation of Pen Movements," 5th International Conference on Computer Science and Engineering (UBMK), 2020, pp. 58-61, doi: 10.1109/UBMK50275.2020.9219381.

[26] Güney S., and Erkuş M., (2022), "A real-time approach to recognition of Turkish sign language by using convolutional neural networks." Neural Computing and Applications 34.5 (2022): 4069-4079.

[27] Web 6, (2022), https://opencv.org/ (Erişim Tarihi: 24/05/2022)

[28] Web 7, (2022), https://github.com/raryelcostasouza/pyTranscriber/, (Erişim Tarihi: 24/05/2022)

[29] Web 8, (2022), https://github.com/ahmetaa/zemberek-nlp, (Erişim Tarihi: 24/05/2022)

[30] Web 9, (2022), https://cloud.google.com/speech-to-text, (Erişim Tarihi: 24/05/2022)

[31] Huang G., Liu Z., Maaten L., Weinberger K. Q., (2017), "Densely Connected Convolutional Networks", Proceedings of the IEEE conference on computer vision and pattern recognition, 4700-4708.

ÖZGEÇMİŞ

Banuçiçek Kandemir Kılınççeker 2007 yılında başladığı Fırat Üniversitesi Fen Fakültesi Matematik bölümünü 2011 yılında başarı ile tamamladı. 2011-2012 yılları arasında Fırat Üniversitesi Eğitim Bilimleri Enstitüsü'nde tezsiz yüksek lisans yaparak pedagojik formasyonunu tamamladı. 2013-2017 yılları arasında Derince Belediyesi'nde Matematikçi olarak çeşitli birimlerde çalıştı. 2017-2021 yılları arasında Kocaeli Büyükşehir Belediyesi, Veri Analizi ve İstatistik Biriminde Matematikçi ve Veri Analisti olarak görev yaptı. 2018 yılında yüksek lisans eğitimine Gebze Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Matematik Anabilim Dalında başladı. 2021 yılından beri yüksek lisans aşamasında yoğun bir şekilde çalıştığı makine öğrenmesi ve derin öğrenme alanları ile ilgili çalışmalarına devam etmektedir.