Üniversite	**:**	**İstanbul Kültür Üniversitesi**
Enstitü	**:**	**Lisansüstü Eğitim Enstitüsü**
Anabilim Dalı	**:**	**Bilgisayar Mühendisliği**
Programı	**:**	**Bilgisayar Mühendisliği**
Tez Danışmanı	**:**	**Dr. Öğr. Üyesi Fatma Patlar AKBULUT**
Tez Türü ve Tarihi	**:**	**Yüksek Lisans – Mayıs 2023**

ÖZET

MODLAR ARASI TRANSFER ÖĞRENİMİ İLE SES SİNYALLERİNDEN DUYGU TANIMA

Fahreddin Raşit Kılıç

İnsanların konuşma sırasında ifade ettikleri duyguları anlamak, uygun şekilde tepki vermek için önemlidir. Ses sinyallerinden anlayabileceğimiz bilgileri maksimize edebilmek için ilgili ses ve görüntünün transfer öğrenimi yöntemi ile analiz edilmesi önemlidir. Duygu tanıma çalışmalarıyla alakalı olarak derin öğrenme ve yapay zekâ algoritmalarıyla araştırmalar hız kazanmıştır. Özellikle yapay zekâ ve robotik sistemlerde, doğal ve empatik bir insan-makine etkileşimi sağlamak için ses sinyallerinden duygu analizi esastır. Bu sistemler sayesinde kullanıcı deneyimini zenginleştirerek daha etkili ve tatmin edici hizmetler sunulabilmektedir.

Duygu analizi sağlık sektöründe de önemli bir rol oynamaktadır. Psikolojik hastalıkların teşhis ve takibinde, hastaların duygu durumlarını doğru bir şekilde tespit etmek, uygun tedavi ve müdahalelerin gerçekleştirilmesi için kritiktir. Eğitim sektöründe ise, öğrencilerin ve öğretmenlerin duygusal durumlarını anlamak, eğitim ve öğretim süreçlerini daha etkili hale getirmektedir. Reklam ve pazarlama alanında, tüketici duygularını analiz etmek, müşteri memnuniyetini ve marka sadakatini artırarak satışları ve karlılığı yükseltmektedir. Ayrıca, duygu analizi, oyun endüstrisinde daha gerçekçi ve etkileyici oyun deneyimleri sunmak için de kullanılmaktadır.

Bu tez çalışmasında, ses sinyallerinden ve ses sinyallerine ait ilgili görüntülerden transfer öğrenme yöntemi ile bu verilerin duygu durumlarını tespit etmeye yönelik gelişmiş sınıflandırma ve analiz yöntemlerini kullanılarak doğru duygu tahminlerinde bulunmayı hedeflenmektedir. Bu çalışmada veri seti nötr, sakin, mutlu, üzgün, kızgın, korkulu, tiksinme ve şaşırmış olmak üzere 8 farklı duygu durumu kullanılmıştır. Ses verilerini analiz edebilmek için MFCC ve Log Mel Filter Bank olmak üzere iki yöntem, Dense ve LSTM olmak üzere iki derin öğrenme tekniği kullanılmıştır. Video veri setini analiz edebilmek içinse CNN ağ modeli kullanılmıştır. Toplamda 11 farklı uygulama gerçekleştirilen bu uygulamada modellerin başarısı analiz edilmiş ve sonuç olarak görüntü verilerinden sınıflama gerçekleştiren modelden konuşma ses sinyalleri verilerinden sınıflama gerçekleştiren modele transfer öğrenmesi yöntemi ile bilgi aktarımı gerçekleştirilip %6,78'lik başarı artışı sağlanmıştır. Ayrıca MFCC yönteminin LMFB'a göre daha başarılı olduğu, şarkı ses türünün ise konuşma ses türüne göre daha yüksek doğrulukla etiketlendiği görülmüştür.

Anahtar Kelimeler : Ses sinyalleri, duygu analizi, derin öğrenme, transfer öğrenmesi, CNN, LSTM, MFCC, LMFB, Dense

University	**:**	**İstanbul Kültür University**
Institute	**:**	**Institute of Graduate Education**
Branch	**:**	**Computer Engineering**
Program	**:**	**Computer Engineering**
Thesis Advisor	**:**	**Dr. Fatma Patlar AKBULUT**
Thesis Type and Date	**:**	**Master Degree – May 2023**

ABSTRACT

EMOTION RECOGNITION FROM AUDIO SIGNALS WITH CROSS-MODAL TRANSFER LEARNING

Fahreddin Raşit Kılıç

Understanding the emotions people express during conversation is important to responding appropriately. In order to maximize the information we can understand from audio signals, it is important to analyze the relevant audio and video with the transfer learning method. Related to emotion recognition studies, research has gained momentum with deep learning and artificial intelligence algorithms. Especially in artificial intelligence and robotic systems, emotion analysis from sound signals is essential to providing a natural and empathetic human-machine interaction. Thanks to these systems, more effective and satisfying services can be provided by enriching the user experience.

Sentiment analysis also plays an important role in the healthcare industry. In the diagnosis and follow-up of psychological diseases, it is critical to accurately determine the mood of the patients and to carry out appropriate treatment and interventions. In the education sector, understanding the emotional states of students and teachers makes education and training processes more effective. In the field of advertising and marketing, analyzing consumer emotions increases sales and profitability by increasing customer satisfaction and brand loyalty. In addition, sentiment analysis is also used in the gaming industry to provide more realistic and immersive gaming experiences.

The objective of this thesis is to accurately predict emotional states based on audio signals and simultaneous facial images using a cross-domain transfer learning approach. Eight distinct states of emotion were used in this analysis: neutral, calm, happy, sad, angry, fearful, disgusted, and surprised. Traditional signal processing methods, such as MFCC and Log Mel Filter Bank, and deep learning techniques, such as Densely connected network, CNN, and LSTM, were used to analyze the audio data. The videos were evaluated by employing a CNN network model. In this application, which consisted of a total of 11 different applications, the success of the models was analyzed, and as a result, information was transferred from the model performing the classification from the image data to the model performing the classification from the speech audio signal data using the cross-modal transfer learning, resulting in a 6.78 percent improvement in success. In addition, it was discovered that the MFCC method was more effective than the LMFB, and that the song voice type was identified with greater precision than the speech voice type.

Keywords: Audio signals, sentiment analysis, deep learning, transfer learning, CNN, LSTM, MFCC, LMFB