Teknoloji Fakültesi

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

YAPAY SİNİR AĞLARI TEKNOLOJİLERİ KULLANILARAK MULTİMODÜLER VERİ SETİ ÜZERİNDE DUYGU ANALİZİ

BİTİRME PROJESİ

1. ARA RAPORU

Bilgisayar Mühendisliği Bölümü

DANIŞMAN

Doç. Dr. Ayşe Berna ALTINEL

İSTANBUL, 2025

İÇİNDEKİLER

ÖZET

YAPAY SİNİR AĞLARI TEKNOLOJİLERİ KULLANILARAK MULTİMODÜLER VERİ SETİ ÜZERİNDE DUYGU ANALİZİ

Bu tez çalışmasında, çok kişili diyaloglarda duygu analizi yapılmasının zorluğu ile ilgili problemlere çözüm olarak yapay sinir ağları teknolojilerine dayanan çözümler araştırılmıştır. MELD veri seti üzerinde multimodüler yapıya sahip farklı modellerin kullanılması ile duygu analizi konusunda en optimal modellerin tespit edilmesi için çalışmalar gerçekleştirilmiştir. Deneylerin yapıldığı bazı modeller textCNN, çift yönlü RNN yapısına sahip bcLSTM ve GRU tabanlı DialogueRNN modelleridir.

Mart, 2025

Sude Nur Tungaç

Rabia Şevval Aydın

ABSTRACT

EMOTION ANALYSIS ON MULTIMODAL DATASET USING ARTIFICIAL NEURAL NETWORKS TECHNOLOGIES

In this thesis project, solutions based on artificial neural network technologies were investigated to address the challanges in performing emotion analysis in multi-person dialogues. Studies were carried out to determine the most effective model for sentiment analysis by using various multimodular structure models on the MELD dataset. Some of the models used in the experiments include textCNN, bcLSTM with a bidirectional RNN structure and GRU-based DialogueRNN models.

March, 2025

Sude Nur Tungaç

Rabia Şevval Aydın

SEMBOLLER

σ **:** aktivasyon fonksiyonu

$\sum$ **:** toplam sembolü

KISALTMALAR

AVEC **:** audio/visual emotion challange

ERC **:** emotional recognition in conversations

GRU **:** gated recurrent unit

IEMOCAP: interactive emotional dyadic motion capture

MELD **:** multimodal emotion lines dataset

MOSEI **:** multimodal corpus of sentiment intensity

MOUD **:** multimodal opinion utterances dataset

ŞEKİL LİSTESİ

TABLO LİSTESİ

1. GİRİŞ

1.1. Problem Tanımı

Günümüzde duygu analizi sağlık alanında psikolojik rahatsızlıkların tespiti ve tedavisi, hukuk alanında suç analizi ve ifade analizi, pazarlama alanında kullanıcı tepkisi, kişi davranış tespiti gibi çeşitli amaçlar ile kullanılmaktadır. Geleneksel duygu analizi çalışmaları genellikle tek tip veri türüne odaklanmaktadır. Konuşmacılardan alınan ifadelerde metin verisi üzerine duygu analizi, mimikler kullanılarak görsel veri ile duygu analiz veya konuşma kayıtlarından ses verisi ile duygu analizi yapılması gibi. Bu tez çalışması duygu analizinin diyaloglar üzerine uygulanmasına odaklanacaktır.

Diyaloglar insan doğası gereği çoklu modüler yapıdadır. Konuşmacılar sözlü ifadelerine ek olarak mimikler, ses tonu, vücut hareketleri kullanarak duygularını karşı tarafa aktarırlar. Bu sebeple karşılıklı konuşmalarda duygu tespiti (ERC-Emotional Recognition in Conversations) çalışmalarında bu yapıya uygun veri seti ve yöntemlerinin kullanılması gerekmektedir [1].

Yıllar içerisinde ses, metin ve görsel veriler ile yapılan çalışmalar ile duygu analizi konusunda gelişmeler elde edilmesine rağmen diyaloglar üzerine olan çalışmalar yetersiz kalmaktadır. Bu durumun en büyük nedeni olan veri eksikliği sorununu çözmek için [1] tarafından diyaloglar üzerine çoklu modaliteye sahip geniş örnek sayısına sahip veri seti MELD (Multimodal EmotionLines Dataset) sunulmuştur. Proje kapsamında kullanılacak olan bu veri setinin kullanılmasında, modalitelerden verimli özellik çıkarılması, modalitelerin senkronize olarak kullanılması gibi sorunlara çözümler bulunması gerekmektedir.

1.2. Projenin Amacı ve Hedefleri

Projenin nihai hedefi multimodüler duygu analizi için en iyi performans gösteren yapay sinir ağı yöntemlerini bulmak ve birlikte çalışacakları kapsamlı bir sistem geliştirmektir. Bu amaç doğrultusunda öncelikle MELD veri setinin farklı modaliteleri incelenerek tek bir modalite üzerinden temel analizler gerçekleştirilecektir. Ardından diğer modaliteler

de eklenerek, çoklu modaliteler üzerine analiz gerçekleştirebilen bir yapı oluşturulacaktır. Proje boyunca her aşamada her bir modalite için çeşitli yapay sinir ağları modelleri test edilecek ve analiz değerlendirme raporu oluşturulacaktır. Bu rapor ile her bir modalite için optimal yöntemin bulunması ve farklı yöntemlerin birleştirilerek en iyi sonuç veren multimodal modelin geliştirilmesi hedeflenmektedir.

1.3. Proje Takvimi

Projenin geliştirilmesi esnasında ana görevlerin gerçekleştirilme sırası ve ne kadar sürecekleri ile ilgili hazırlanan proje takvimi aşağıdaki gibidir.

Tablo 1. Proje takvim tablosu

Proje Aşaması	**Süre**
Literatür taraması	2 hafta
Veri setinin analizi	2 hafta
Modalitelerden özellik çıkarılması	3 hafta
Tekli modaliteye sahip modellerin oluşturulması	4 hafta
Multimodal model yapısının oluşturulması	5 hafta
Analiz performans raporunun oluşturulması	2 hafta
Sonuçların değerlendirilmesi	3 hafta
Rapor ve sunum hazırlığı	2 hafta

2. MULTİMODÜLER YAPIDAKİ DİYALOG VERİSİ ÜZERİNDE YAPAY SİNİR AĞLARI KULLANILARAK DUYGU ANALİZİ GERÇEKLEŞTİRME YÖNTEMLERİ

Multimodüler duygu analizinde farklı modalitelerden (ses, metin ve görüntü) gelen verilerin birleştirilmesiyle tekli modalite duygu analizinden elde edilen başarıyı arttırması hedeflenir. Ses verisinin tonlama, perde ve vurgu gibi akustik özellikler taşımasının yanı sıra metin verisi anlamsal bilgiler sunar. Görüntü verisi ise içeriğinde jest ve mimikler bulunur. Bu modalitelerin birbirleri ile entegre edilerek bir arada kullanılması ile duygu analizinde daha yüksek doğruluk elde edilir. Bu bağlamda, temel olarak kullanılan üç çıktı modeli bulunur. Bu çıktı modelleri şu şekildedir:

- Erken Birleşme (Early Fusion): Farklı modaliteler için ham verilerin işlenmeden birleştirilmesinin ardından modelin girdi katmanına verilmesi şeklinde gerçekleşir.
- Geç Birleşme (Late Fusion): Her modalitenin ayrı ayrı işlenerek özelliklerinin çıkarılmasının ardından birleştirilmesi durumudur.
- Orta Düzeyde Birleştirme (Hybrid Fusion): Modalitelerin belirli bir seviyeye kadar ayrı ayrı işlenmesinin ardından daha sonrasında bir arada işlenmesiyle gerçekleştirilen yöntemdir [2].

2.1. Literatür Taraması

Yapay sinir ağları ile çoklu modüler yapıda duygu analizi yapılması güncel ve önemli görülen bir konudur. Literatürde bu konuda yapılmış birçok araştırma ve yöntem bulunur. Duygu analizinde diyalog sürecini anlamak ve duygu akışını takip etmek için [2] tarafından yapılan çalışmada geliştirilmiş DialogueRNN modelinden bahsedilir. Geliştirilen DialogueRNN modeli ve araştırmada incelenen diğer modeller IEMOCAP ve AVEC veri setleri üzerinde test edildiğinde en başarılı performansı DialogeRNN varyantlarından biri olan BiDialogRNN+Attn varyantı gösterir.

İkili diyalogların yanında çok kişili diyaloglar üzerinde duygu analizi için [1] tarafından yapılan çalışmada MELD veri seti üzerinde test edilen üç ana model vardır. Bu modeller

text-CNN, bcLSTM ve DialogueRNN modelleridir. text-CNN konuşmanın bağlamını dikkate almaz, yani ifadelerin sırasını veya önceki ifadelerin durumlarını kullanmaz. İki yönlü RNN kullanan bcLSTM ise konuşmayı bir bütün olarak işler, konuşmacı değişikliğini dikkate almaz. DialougeRNN ise çok kişili diyaloglarda konuşmacı durumlarını takip ederek her bir konuşmacı için özel bağlam oluşturur.

[3] tarafından yapılan çalışmada bağlam duyarlılığı bağımlılıklarının yanı sıra konuşmacı duyarlılığı bağımlılıkları da dikkate alınır. Bu özelliklere odaklanan bir model olan ConGCN modeli geliştirilir. Modelde graf bazlı evrişimli sinir ağı kullanılır. Graf tabanlı modellemede her bir ifade ile her bir konuşmacı için birer düğüm bulunur. Bağlamsal bağımlılık için aynı konuşmadaki düğümlerin ilgili kenarları birbirine bağlanırken konuşmacı bağımlılığının sağlanmasında ifade düğümü ile konuşmacı düğümü arasında bir bağlantı oluşturulur. Oluşturulan GCN modelinin performansının MFN, BC-LSTM, CMN, ICON ve DialogueRNN modellerinin performansları ile karşılaştırıldığında en yüksek başarıyı verdiği görülür.

[4] tarafından yapılan çalışmada MELD veri setinde bulunan üç veri türü için de derin öğrenme modülleri yapılandırılarak ince ayar yapılır. Metin modalitesinin ön işlemesinde GPT, ses modalitesinin ön işlemesinde WaveRNN, görüntü modalitesinin ön işlemesinde ise FaceNet modalitesinden yararlanılır. Duygunun tahmini aşamasında çapraz modalite füzyon transformatörü ve füzyon için EmbraceNet mimarisi kullanılır.

[5] tarafından yapılan çalışmada ise farklı veri setleri üzerinde metin verisi ile yapılan çalışmalarda metnin bağlam durumu, içsel durum, dışsal durum, niyet durumu ve duygu durumu değişkenleri kullanılır. Ayrıca RoBERTa large modelinin ince ayar edilerek bağımsız özellik çıkarımında kullanılmasının başarıyı artırdığı görülür.

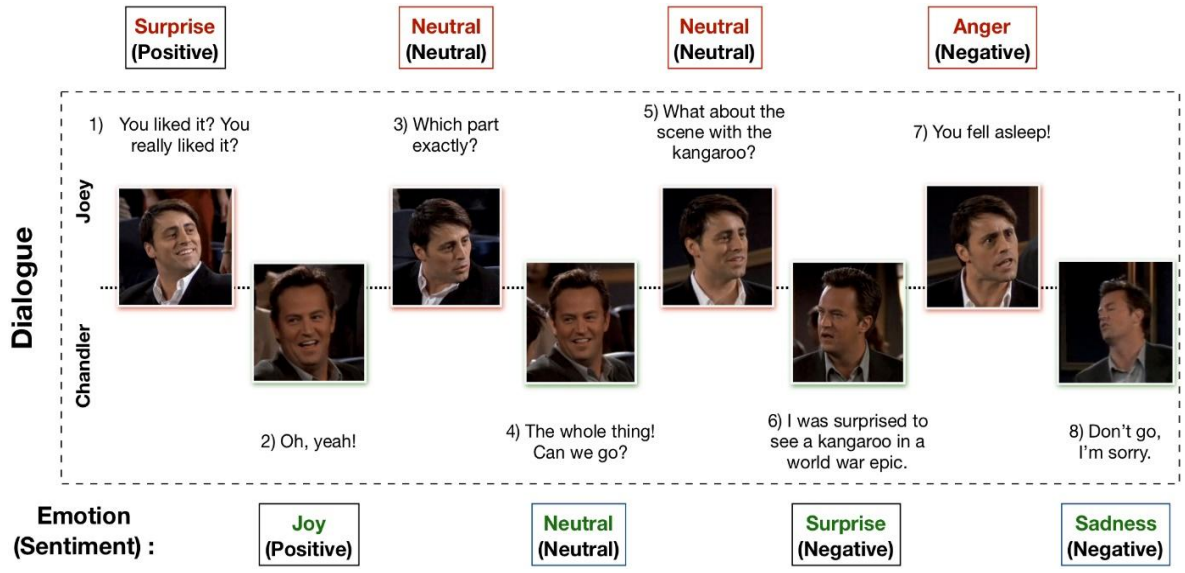
[6] tarafından yapılan çalışmada konuşmalardaki duygu tanıma problemi (ERC) ele alınmıştır ve bu probleme çözüm olarak konuşmayı hem konuşmacılar arasındaki bağımlılık hem de kendi kendine bağımlılık incelemesi ile duygusal olarak sınıflandıran DialogueGCN modeli tanıtılmıştır.

2.2. Veri Seti Analizi

2.2.1 MELD Veri Seti

Multimodüler yapıda duygu analizi için pek çok veri seti bulunmaktadır. Fakat bu veri setlerinin çoğunluğu tekli ifadeler içermesi sebebiyle sınırlı çalışma imkânı sunmaktadır. Örneğin CMU-MOSEI, CMU-MOSI ve CMU-MOUD veri setleri sadece tekli ifadelerden oluştuğundan dolayı diyalog üzerine analiz imkanı sağlayamamaktadır. IEMOCAP VE SEMAINE veri setleri ikili ifadeler içerdiklerinden diyaloglar üzerine çalışma imkanı sağlamaktadır ancak örnek sayısı bu çalışmada kullanılacak olan MELD veri setine göre çok daha azdır. Örnek sayısına ek olarak MELD veri seti ikiden fazla kişiden oluşan diyaloglara yer verdiği için daha kapsamlı bir çalışma fırsatı sunmaktadır.

MELD veri seti hazırlanırken diyaloglar video klipleri ile beraber değerlendirilmiştir. Veri seti Friends adlı televizyon serisine ait 1433 diyalogdan alınmış 13000 ifadeden oluşur. Her bir ifade için 7 duygu (sinir, tiksinti, üzüntü, mutluluk, nötr, şaşkınlık, korku) etiketine ek olarak duygular pozitif, negatif ve nötr sınıflarında gruplandırılmıştır. Sinir, tiksinti, üzüntü, korku duyguları negatif, mutluluk pozitif, nötr ise nötr sınıfı içerisinde gruplanmıştır. Şaşkınlık hem pozitif hem de negatif olarak ifade edilebilen bir duygu örneğidir. Şekil 1'de veri setine ait örnek bir diyalog yer almaktadır.



Şekil 1. MELD veri setinden diyalog örneği

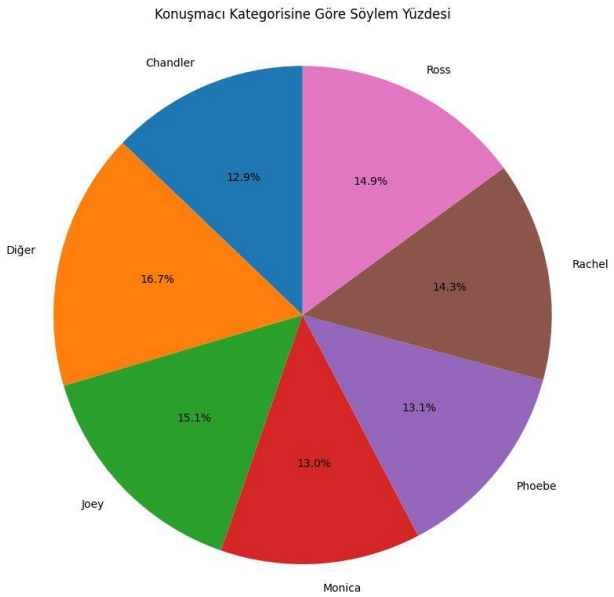
Tablo 2'de veri setinde her bir duygu sınıfı için kaç adet ifade olduğu gösterilmektedir.

Tablo 2. MELD veri setinin duygu sınıflarına ait örnek sayıları

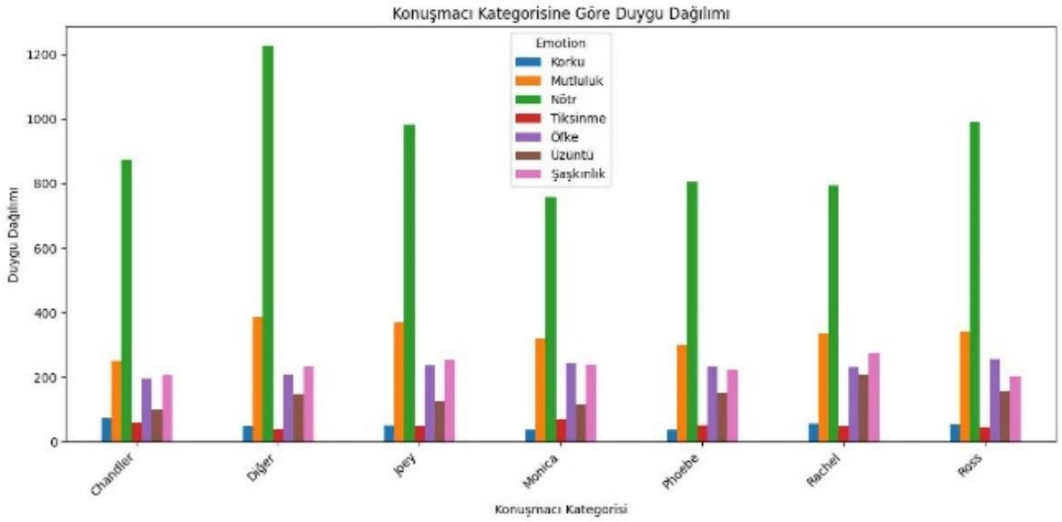
Duygular	Nötr	Şaşkınlık	Korku	Üzüntü	Sevinç	Tiksinti	sinir
MELD	6436	1636	358	1002	2308	361	1697

Konuşmacılara göre ifade sayısı Şekil 2'de yer almaktadır. Ana karakterlere ait olan konuşmalar haricindeki konuşmalar 'Diğer' başlığı altında toplanmıştır. Konuşmacılara göre duygu dağılımı ise Şekil 3'te görülmektedir. Konuşmacılara göre his dağılımı Şekil

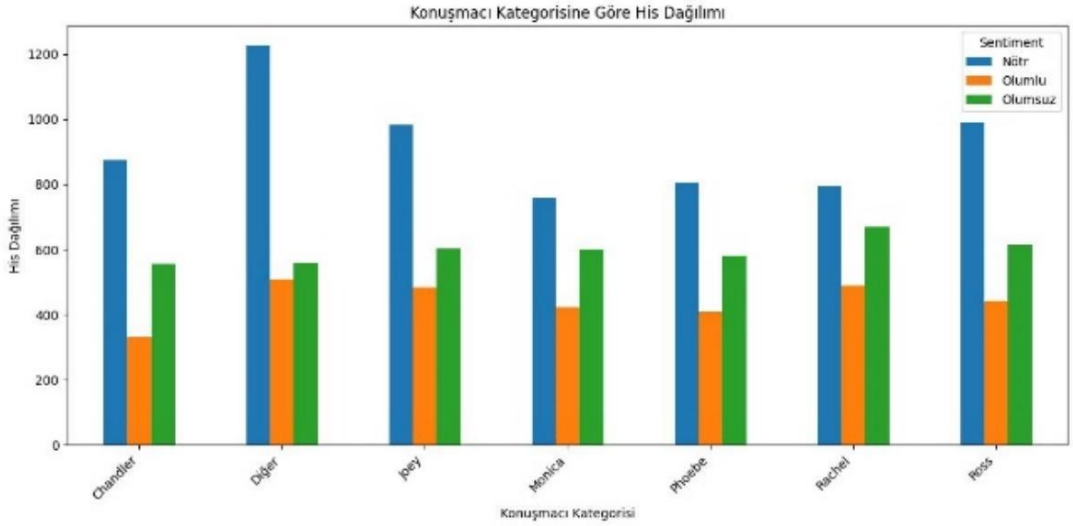
4'te bulunmaktadır. Bu görsellerin incelenmesi sonucunda veri serinde nötr sınıfına ait verilerin diğer sınıflara oranla daha fazla olduğu gözlemlenmektedir. Korku ve tiksinme sınıflarında ise bu oranın düşük olduğu görülmektedir.



Şekil 2. MELD veri setindeki konuşmacıların ifade dağılımı yüzdeleı



Şekil 3. MELD veri setindeki konuşmacıların duygu dağılımları

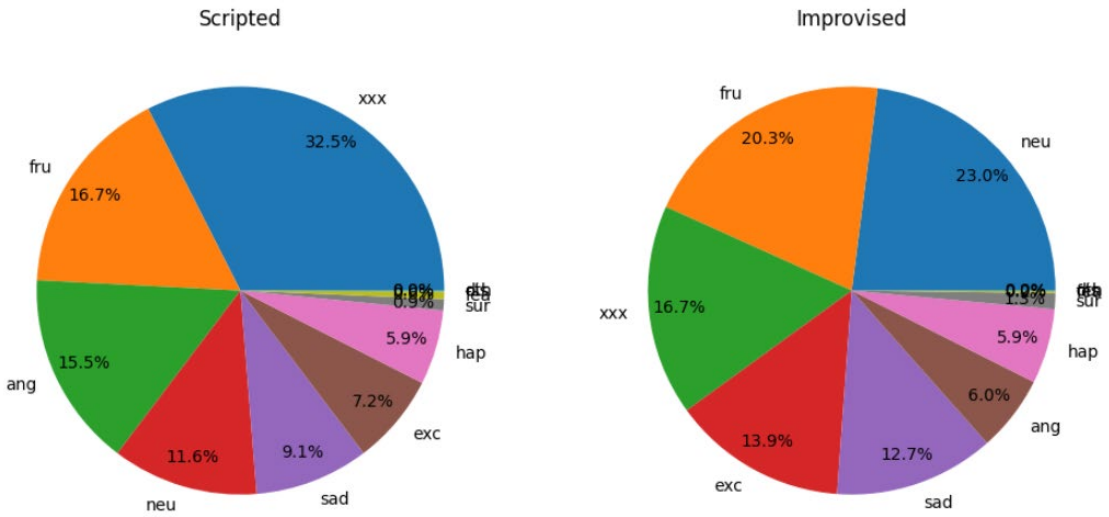


Şekil 4. MELD veri setindeki konuşmacıların his dağılımları

2.2.2 IEMOCAP Veri Seti

Interactive Emotional Dyadic Motion Capture (IEMOCAP) veri seti 5 erkek, 5 kadın olmak üzere 10 kişilik bir grup tarafından gerçekleştirilen ikili diyalog çiftlerinden oluşur. 151 diyaloğun 2 konuşmacı için de video görüntülerini alarak 302 video kaydı sunar. Her bir diyalog ifadesi sinir, heyecan, korku, üzüntü, şaşkınlık, mutluluk, memnuniyetsizlik, hayal kırıklığı ve nötr olmak üzere 9 duygu sınıfı ile etiketlenmiştir. Her bir diyalog için ifadelerin yazılı verisi, konuşmacıların ses kayıtları, video kayıtları olmak üzere üç modalite için de veri bulunur.

Toplamda 10039 tane veri örneği bulunur. Bu veri örnekleri iki farklı metot altında gruplanır: bir senaryoya bağlı gerçekleşen diyaloglar ve doğaçlama olarak gerçekleştirilen diyaloglar. İfadelerin 4784 tanesi doğaçlama, 5255 tanesi senaryoya bağlıdır. Şekil 5'te her bir grup için duygu etiketlerinin dağılım yüzdeleri gösterilmiştir.**[7]**



Şekil 5. IEMOCAP veri setindeki duyguların metot gruplarına göre dağılımları

2.3. Önerilen Yöntem ve Modeller

DialogueRNN / RoBERTa+DialogueRNN: DialogueRNN modeli çok modlu duygu sınıflandırmasında diyaloğun bağmanını anlamak ve duygu akışını takip etmekte verimli bir yöntem olarak karşımıza çıkar. Bu modelin üç modülü bulunmaktadır.

- Küresel durum (Global GRU), önceki ifadelerin ve konuşmacının durumunun dikkate alınması ile genel bağlamın temsil edilmesidir. g_t , küresel bağlamdaki zaman adımının t durumunu temsil eder. x_t, güncel girdinin vektörüdür. g_{t-1}, önceki zaman adımındaki küresel bağlam vektörünü ifade eder.

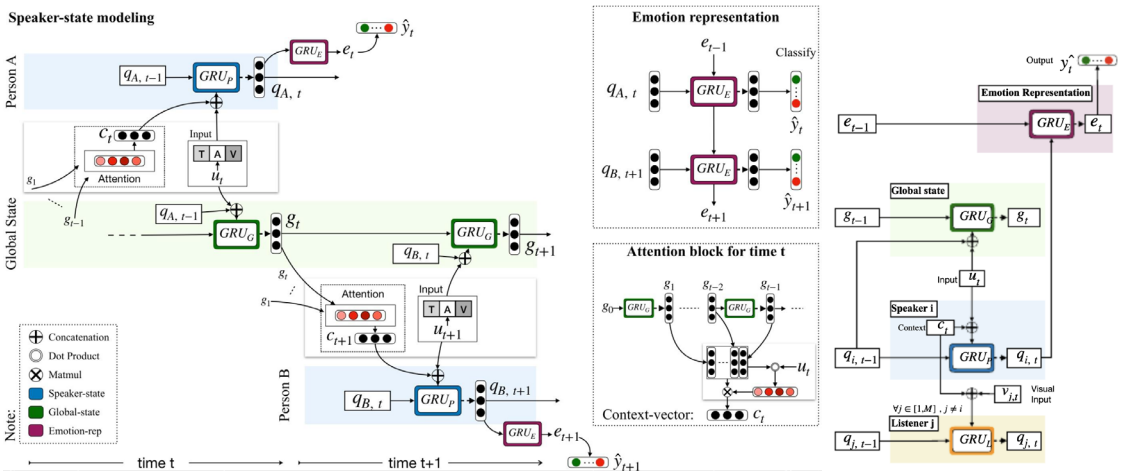
$$g_t = GRU_g(x_t, g_{t-1}) \tag{1}$$

- Konuşmacı durumu (Speaker GRU), konuşmacının önceki durumunun ve konuşmasının bağlamının ifadesidir. $s^i{}_t$, belirli bir konuşmacının belirli bir andaki durumudur. Eğer bir konuşmacı i, t zamanında konuşuyor ise durumu aşağıdaki gibi güncellenir.

$$s_t{}^i = GRU_s(x_t, s^i{}_{t-1}) \tag{2}$$

- Duygu durumu (Emotion GRU) kısmında ise konuşmacının durumunun ve önceki ifadelerin duygusal bağlamının birleştirilmesinin sonucunda duygusal temsilin oluşturulmasıdır. e_t, t zamanındaki duygusal durumu temsil eder.

$$e_t = GRU_e(g_t, s_t, e_{t-1}) \tag{3}$$



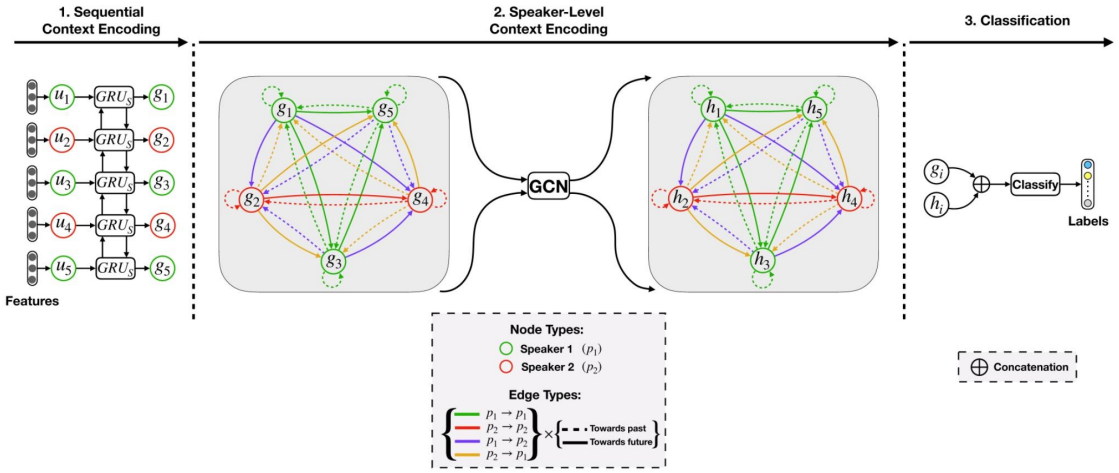
Şekil 6. DialogueRNN mimarisi ve t zamanı ifade etmek üzere bir diyalogdaki küresel, konuşmacı, dinleyici ve duygu durumlarının güncellenmesi şeması

Şekil 2.5'te M katılımcısı olan bir konuşmada kişi i konuşmacıdır ve kişiler $j \in [1,M]$ ve $j \neq i$ dinleyicidir. Mevcut ifade bu ifadelerin bir fonksiyon ile modellenmesinden elde

edilir [2]

DialogueRNN varyantları test edildiğinde en yüksek performansı BiDialogueRNN+Att varyansı elde eder. RoBERTa-large modelinin ince ayar edilmesi ile özellik çıkarılmasının ardından DialogueRNN kullanımında ise performansın artması söz konusudur [5].

DialogueGCN: Konuşmalardaki duygu tespitinin yapılması için geliştirilmiş olan bu model üç temel bileşenden oluşmaktadır.



Şekil 7. DialogueGCN mimarisine genel bakış

- Sıralı bağlam kodlayıcı (Sequential Context Encoder), çift yönlü GRU kullanarak konuşma verilerini çift yönlü işler. Bu aşamada konuşmanın sırası önemlidir ancak konuşmacı bilgisi göz önüne alınmaz. u_i, girdi olarak kullanılan konuşma cümlesini ifade eder. g_i, katmanın çıktısı olan sıralı bağlamsal özellik vektörüdür. $GRU^{\leftrightarrow}$, ise çift yönlü GRU hücresini sembolize etmektedir.

$$g_i = GRU^{\leftrightarrow}(g_{i-1}, u_i) \tag{4}$$

- Konuşmacı Seviyesinde Bağlam Kodlayıcı (Speaker-Level Context Encoder), konuşmadaki konuşmacılar arası bağlantıların ve kendi kendine bağımlılığın modellenmesinde graf yapısı kullanır. Konuşma cümlelerinin her biri birer düğüm olarak oluşturulur. Cümlenin konuşmacı ve zamansal bağımlılığına göre düğümün ilişkileri (kenarları) belirlenir. Kenar ağırlıkları benzerlik tabanlı dikkat

mekanizması kullanılarak ayarlanır. V, cümleleri ifade eden düğüm kümesidir. E, bağımlılıkları ifade eden kenar kümesidir. R, ilişki türleridir ve W, ağırlıkları temsil etmektedir.

$$G = (V, E, R, W) \tag{5}$$

Kenar ağırlıkları, benzerlik tabanlı dikkat mekanizmasının kullanılması ile belirlenir. a_{ij}, i ve j düğümleri arasındaki ağırlığı ifade etmektedir. W_e, öğrenilebilir parametre matrisidir. p ve f geçmiş ve gelecek pencere boyutlarını temsil etmektedir.

$$a_{ij} = softmax({g_i}^T W_e \ [g_{i-p}, \ldots, g_{i+f}]) \tag{6}$$

Graf evrişim işlemlerinde ilk adımda yerel komşuluk bilgisinin dönüştürülmesi şu şekilde gerçekleştirilir:

$${h_i}^{(1)} = \sigma(\sum_{r \epsilon R} \sum_{j \epsilon {N_i}^r} a_{ij} . {W_r}^{(1)} g_j + a_{ii} . {W_0}^{(1)} g_i) \tag{7}$$

Elde edilen değerler ikinci adımda bir sonraki aşamaya taşınır.

$${h_i}^{(2)} = \sigma(\sum_{j \epsilon {N_i}^r} {W_r}^{(2)} {h_j}^{(1)} + {W_0}^{(2)} {h_i}^{(1)}) \tag{8}$$

- Duygu Sınıflandırıcı (Emotion Classifier), konuşmanın hem sıralı hem de konuşmacı düzeyindeki bağlamsal özelliklerinin birleştirilmesi ile gerçekleştirilir. Tam bağlantılı katmanlar ve softmax fonksiyonu kullanılarak duygu sınıflandırması yapılır. Özellik birleştirme ve son duygu sınıflandırma işlemleri sırası ile aşağıdaki gibidir:

$$h_i = [g_i, {h_i}^{(2)}] \tag{9}$$

$$P_i = softmax(W_{smax}.RELU(W_l.h_i + b_l) + b_{smax}) \quad (10)$$

Bu model ile uzun süreli bağımlılığa sahip diyalogların bağlam modelleme sorunları çözülür. Konuşmacı düzeyindeki bağlamsal bilgiler DialogueRNN ve birçok modele göre daha iyi yakalanır. Çok katılımcısı olan konuşmalarda başarı performansı yüksektir. Sahip olduğu özellikler sayesinde duygu tanıma konusunda başarılıdır [6].

Proje kapsamında, bahsedilen modellerin veri seti üzerinde performansları test edilecektir.

2.4. Model Eğitimi ve Değerlendirme Teknikleri

Model eğitimi süreci çoklu modüler yapılar ve yapay sinir ağları kullanılarak çok kişili konuşmalarda duygu analizi yapılmasında önemli bir aşamadır. Verilerin ön işlenmesi, model mimarisinin belirlenmesi, verilerin bu modellere uygun boyuta getirilmesi, modellerin parametrelerinin uygun şekilde ayarlanması ve optimizasyon işlemlerinin yapılması, model performansının değerlendirilmesi adımlarını içerir. Bu bölümde modelin eğitimi ve değerlendirilmesi kısımları ile ilgili detaylı bilgiler verilecektir.

2.4.1. Model Eğitim Süreci

MELD ve IEMOCAP veri seti üzerinde uygulanan deneylerde çeşitli yöntemler uygulanmıştır. Verilerden özellik elde etme aşamasında yazılı veriler için önceden eğitilmiş GloVe vektörleri ve 1D-CNN kullanılarak metin özellikleri elde edilmiştir. Ses verileri için openSMILE aracı kullanılmıştır. Deneylerde görsel modülü video bazlı konuşmacı tespiti probleminden dolayı kullanılmamıştır.

Model eğitimi için yapay sinir ağları tabanlı modeller ve çoklu modüler yapıya sahip modeller tercih edilmiştir. Model mimarisinde metinsel ve ses verilerini birlikte ve ayrı ayrı işleyebilen yapılar tercih edilmiştir. Bundan dolayı CNN tabanlı bir text-CNN modeli, çift yönlü RNN tabanlı bir bcLSTM modeli ve 3 adet GRU yapısı kullanan DialogueRNN modeli eğitilmiştir. Aynı zamanda verilerden öznitelik çıkarımı

aşamasında RoBERTa modeli kullanılarak elde edilen öznitelikler DialogueRNN modeline verilmiştir. Benzer bir şekilde RoBERTa ile işlenen veriler COSMIC modeli kullanılarak duygu sınıflandırması yapılmasında da kullanılmıştı.r

Eğitim sürecinde modellerin aşırı öğrenmesinin önüne geçilmesi adına erken durdurma (early stopping) ve dropout teknikleri kullanılmıştır. text-CNN ve bcLSTM modellerinde optimizasyon algoritması için 'adam' seçilmiştir. Kayıp fonksiyonu olarak ise 'kategorisel çapraz entropi kaybı' (categorical crossentropy) tercih edilmiştir. DialogueRNN ve COSMIC modellerinde ise bu modellerden farklı olarak kayıp fonksiyonunda Masked NLL Loss tercih edilmiştir.

2.4.2. Model Değerlendirme Süreci

Eğitimin tamamlanmasının ardından modelin performansının doğru bir şekilde değerlendirilebilmesi için modelin başarısını ve genelleme yeteneğini ölçerken bazı temel metriklerden yararlanılmıştır. Bu metrikler doğruluk (accuracy), hassasiyet (precision), duyarlılık (recall), f1-skoru (f1-score) ve karmaşıklık matrisi (confusion matrix) yöntemleridir. Değerlendirme sürecinde, eğitim, validasyon ve test veri setleri kullanmıştır, bu yöntem ile modelin genelleme yeteneği ölçülmüştür.

3. BULGULAR VE TARTIŞMA

3.1. Yapılan Çalışmalar

Çok konuşmacılı diyaloglarda yapay sinir ağları ve çoklu modüler yapıların kullanılması ile konuşmanın bağlam takibinin yapılabilmesi önem taşımaktadır. Ayrıca farklı veri türlerinden yararlanılarak duygu sınıflandırmasında ulaşılan başarının tekli modaliteler ile yapılan duygu sınıflandırmasına göre daha yüksek başarı oranına sahip olması beklenmektedir. Çalışmada DialogueRNN modelinin MELD veri setindeki metin, ses ile hem metin hem ses verisi üzerinde çalıştırılması, COSMIC modelinin MELD ve IEMOCAP veri setinde yer alan metin verileri için çalıştırılması, text-CNN ve bcLSTM modellerinin ise MELD veri seti için hem metin hem ses verileri üzerinde çalışmasının ardından multimodal çalıştırılması sağlanmıştır. Base modeller ile MELD veri kümesi

üzerinde elde edilen sonuçlar Tablo 3'te karşılaştırılmıştır. DialogueRNN modelinin, RoBERTa+DialogueRNN ve RoBERTa+COSMIC kombinasyonlarının MELD ve IEMOCAP veri seti üzerinde elde ettiği sonuçlar ise Tablo 4'te gösterilmiştir.

Tablo 3. MELD veri setinde base modellerin ve DialogueRNN modelinin duygu sınıflandırmasına ait doğruluk ve f1-skorları

Modeller / Duygular	nötr	şaşkınlık	korku	üzüntü	sevinç	tiksinti	sinir	doğruluk
Base model text	0.6498	0.00	0.00	0.00	0.00	0.00	0.00	0.4812
Base model audio	0.4578	0.00	0.00	0.00	0.00	0.00	0.00	0.3390
Bimodel base text+audio	0.6518	0.0137	0.00	0.00	0.1240	0.00	0.272	0.4816

Tablo 4. MELD ve IEMOCAP veri setinde DialogueRNN ve COSMIC modellerinin duygu sınıflandırmasına ait doğruluk değerleri

Modeller / Veri Setleri	MELD			IEMOCAP
	text	audio	multimodal	
DialogueRNN	59.46	47.47	59.46	61.55
RoBERTa + DialogueRNN	53.64	47.47	55.52	-
RoBERTa + COSMIC	66.25	-	-	65.08

Çalışmada base model olarak bcLSTM modeli kullanılmıştır. Base modellerin her biri için epoch sayısı 100, batch sayısı 50 olacak şekilde eğitim gerçekleştirilmiştir. DialogueRNN modeli için ise epoch sayısı 100 iken batch sayısı 30 olacak şekilde eğitim gerçekleştirilmiştir. COSMIC modelinin eğitiminde epoch sayısı 60 olarak belirlenmiştir. Batch size ise 32 olarak kullanılmıştır.

3.2. Sonuçların Değerlendirilmesi

Sonuçlar incelendiğinde duygu sınıfları arasında nötr etiketine sahip verilerin genel olarak sınıflandırmada diğer sınıflara göre baskın olduğu görülmektedir. Şaşkınlık, korku, üzüntü, tiksinti gibi daha karmaşık duygulara ait verilerin sınıflandırılabilmesi için modellerin tek başına yetersiz kaldığı, öznitelik çıkarımı yöntemleri kullanılarak bu verilerden elde edilecek özellikler ile sınıflandırmanın yapılması gerektiği görülmektedir. Modalitelerde tekli kullanım ile ikili kullanım karşılaştırıldığında, tek modalite kullanımında sevinç ve sinir duygularında performans göstermediği, ikili modalite kullanımında ise bu durumun iyileştiği gözlemlenir. Bu duygu analizinde farklı modalitelerin beraber kullanılarak sınıflandırma sürecinin desteklenmesi gerektiği görüşünü desteklemektedir. Aynı zamanda metin verisinin kullanımıyla yapılan sınıflandırılmalarda COSMIC modelinin performansının diğer modellerin performansları ile karşılaştırıldığında çok daha iyi olduğu görüşmektedir.

3.3. Yapılacak Çalışmalar

Bu çalışmada kullanılan bcLSTM tabanlı base modellerin özellikle karmaşık duyguları sınıflandırmak konusunda yetersiz kaldığı görülmüştür. Özellikle şaşkınlık, korku, üzüntü ve tiksinti gibi duyguların sınıflandırılmasında modelin performansının düşük olması daha ileri çalışmalara ihtiyaç duyulduğunu gösterir.

- Farklı Model Yapıları ile Denemeler: DialogueGCN gibi grafik tabanlı modellerin duygu analizinde kullanılması için deneme yapılacaktır.

Bu çalışmaların gerçekleştirilmesi ile modelin genel performansının gelişmesi ve karmaşık duyguların sınıflandırılmasında başarı oranının artırılması amaçlanmaktadır. Geliştirilmiş modellerin sonuçları mevcut bulgularla karşılaştırılarak model seçiminde en iyi performans gösteren yapının belirlenmesi hedeflenmektedir.

4. SONUÇLAR

Tez çalışması kapsamında şu ana kadar yapılan çalışmalarda sonuçlar incelendiğinde

duygu sınıfları arasında nötr etiketine sahip verilerin genel olarak sınıflandırmada diğer sınıflara göre daha yüksek başarıya sahip olduğu görülmektedir. Şaşkınlık, korku, üzüntü, tiksinti gibi daha karmaşık duygulara ait verilerin sınıflandırılabilmesi için bcLSTM, textCNN ve DialogueRNN modellerinin kapsamlı bir duygu analizinde tek başlarına yetersiz kaldığı görülmüş, COSMIC ile yapılan metin temelli duygu sınıflandırmasının diğer modellere göre çok daha başarılı olduğu görüşmüştür. Bu sebeple DialogueGCN ve ConGCN modellerinin test edilmesi ile ilerlenmesi gerektiğine karar verilmiştir.

5. KAYNAKLAR

[1] S. Poria, D. Hazarika, N. Majumder, G. Naik, R. Mihalcea, and E. Cambria, "MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversation," 2018. doi: 10.48550/arXiv.1810.02508.

[2] N. Majumder, S. Poria, D. Hazarika, R. Mihalcea, A. Gelbukh, and E. Cambria, "DialogueRNN: An Attentive RNN for Emotion Detection in Conversations," *arXiv preprint*, Nov. 2018. doi: 10.48550/arXiv.1811.00405

[3] D. Zhang, L. Wu, C. Sun, S. Li, Q. Zhu, and G. Zhou, "Modeling both context- and speaker-sensitive dependence for emotion detection in multi-speaker conversations," in Proc. 28th Int. Joint Conf. Artif. Intell. (IJCAI-19), 2019, pp. 5415–5421. [Online]. Available: https://www.ijcai.org/proceedings/2019/0752.pdf.

[4] C. Bai, S. Kumar, J. Leskovec, M. Metzger, J. F. Nunamaker, and V. S. Subrahmanian, "Predicting the visual focus of attention in multi-person discussion videos," in *Proc. 28th Int. Joint Conf. Artif. Intell. (IJCAI-19)*, 2019. doi: 10.24963/ijcai.2019/626.

[5] D. Ghosal, N. Majumder, A. Gelbukh, R. Mihalcea, and S. Poria, "COmmonSense knowledge for eMotion Identification in Conversations," in *Findings of the Association for Computational Linguistics: EMNLP*, 2020, pp. 2470–2481. doi: 10.48550/arXiv.2010.02795.

[6] A. Gelbukh, N. Majumder, S. Poria, N. Chhaya, and D. Ghosal, "DialogueGCN:

A Graph Convolutional Neural Network for Emotion Recognition in Conversation," *arXiv preprint*, Aug. 2019. doi: 10.48550/arXiv.1908.11540.

[7] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee ve S. S. Narayanan, "IEMOCAP: Interactive emotional dyadic motion capture database," Language Resources and Evaluation, cilt. 42, ss. 335–359, Kasım 2008, doi: 10.1007/s10579-008-9076-6.

[8] D. Zhang, L. Wu, C. Sun, S. Li, Q. Zhu, and G. Zhou, "Modeling both context- and speaker-sensitive dependence for emotion detection in multi-speaker conversations," in Proc. 28th Int. Joint Conf. Artif. Intell. (IJCAI-19), 2019, pp. 5415–5421. [Online]. Available: https://www.ijcai.org/proceedings/2019/0752.pdf.