RADYOLOJİ RAPORLARINDAN DOĞAL DİL İŞLEME TEKNİKLERİ KULLANILARAK VARLIK İSMİ ÇIKARIMI

YÜKSEK LİSANS TEZİ

Sedanur ORCİN

Danışman
Doç. Dr. Uçman ERGÜN

BİYOMEDİKAL MÜHENDİSLİĞİ ANABİLİM DALI

Haziran 2025

Bu tez çalışması, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından 2210-C Ulusal Lisansüstü Burs Programı ve 1002-A Hızlı Destek Programı kapsamında desteklenmiştir.

AFYON KOCATEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

RADYOLOJİ RAPORLARINDAN DOĞAL DİL İŞLEME TEKNİKLERİ KULLANILARAK VARLIK İSMİ ÇIKARIMI

Sedanur ORCİN

Danışman
Doç. Dr. Uçman ERGÜN

BİYOMEDİKAL MÜHENDİSLİĞİ ANABİLİM DALI

Haziran 2025

TEZ ONAY SAYFASI

Sedanur ORCİN tarafından hazırlanan "Radyoloji Raporlarından Doğal Dil İşleme Teknikleri Kullanılarak Varlık İsmi Çıkarımı" adlı tez çalışması lisansüstü eğitim ve öğretim yönetmeliğinin ilgili maddeleri uyarınca / / tarihinde aşağıdaki jüri tarafından **oy birliği** ile Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü **Biyomedikal Mühendisliği Anabilim Dalı'nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman **:** Doç. Dr. Uçman ERGÜN

Başkan **:** -
Afyon Kocatepe Üniversitesi, Mühendislik Fakültesi

Üye **:** -
Afyon Kocatepe Üniversitesi, Mühendislik Fakültesi

Üye **:** -
Afyon Kocatepe Üniversitesi, Mühendislik Fakültesi

Üye **:** -
Afyon Kocatepe Üniversitesi, Mühendislik Fakültesi

Üye **:** -
Afyon Kocatepe Üniversitesi, Mühendislik Fakültesi

Afyon Kocatepe Üniversitesi
Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
..... /..... /..... tarih ve
……. sayılı kararıyla onaylanmıştır.

…….
Prof. Dr. Bekir YALÇIN
Enstitü Müdürü

BİLİMSEL ETİK BİLDİRİM SAYFASI

Afyon Kocatepe Üniversitesi

Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında;

– Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,

– Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,

– Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,

– Atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,

– Kullanılan verilerde herhangi bir tahrifat yapmadığımı,

– Ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

13/05/2025

İmza

Sedanur ORCİN

ÖZET

Yüksek Lisans Tezi

RADYOLOJİ RAPORLARINDAN DOĞAL DİL İŞLEME TEKNİKLERİ KULLANILARAK VARLIK İSMİ ÇIKARIMI

Sedanur ORCİN
Afyon Kocatepe Üniversitesi
Fen Bilimleri Enstitüsü
Biyomedikal Mühendisliği Anabilim Dalı
Danışman: Doç. Dr. Uçman ERGÜN

Tıbbi metin madenciliği kapsamında, özellikle radyoloji raporlarında yer alan karmaşık yapılı ifadelerin anlamlandırılması büyük önem taşımaktadır. Bu raporlar, içerdiği çeşitli tıbbi terimler ve uzun bağlamsal ilişkiler nedeniyle hem otomatik analiz sistemleri hem de klinik karar süreçleri açısından zorluklar barındırmaktadır. Bu çalışmada, radyoloji raporlarından varlık isimlerini otomatik olarak çıkarabilen bir hibrit model geliştirilmiştir. Modelde BERT, BioBERT ve ClinicalBERT gibi önceden eğitilmiş dil modelleri BiLSTM ve CRF katmanlarıyla entegre edilmiş; hiper-parametre optimizasyonu ise GA ve PSO ile gerçekleştirilmiştir. Yöntem kapsamında, farklı model yapılandırmaları karşılaştırılmış ve en yüksek başarıyı sağlayan kombinasyon belirlenmiştir. Eğitim ve doğrulama süreçleri RadGraph veri kümesi üzerinden gerçekleştirilmiştir. MIMIC-CXR ve CheXpert test setlerinde yapılan değerlendirmelerde, GA ile optimize edilen ClinicalBERT+BiLSTM+CRF modeli sırasıyla %98,90 ve %97,92 F1 skorlarına ulaşmıştır. Elde edilen sonuçlar, önerilen yaklaşımın klinik metinlerde yüksek doğrulukla çalıştığını ve karar destek sistemlerine katkı sağlayabileceğini göstermektedir.

2025, ix + 107 sayfa

Anahtar Kelimeler: Doğal dil işleme (NLP), Radyoloji raporları, Varlık İsmi Çıkarımı (NER), Karar destek sistemleri

ABSTRACT

M.Sc. Thesis

NAMED ENTİTY RECOGNİTİON FROM RADİOLOGY REPORTS USİNG NATURAL LANGUAGE PROCESSİNG TECHNİQUES

Sedanur ORCİN

Afyon Kocatepe University

Graduate School of Natural and Applied Sciences

Department of Biomedical Engineering

Supervisor: Assoc. Prof. Dr. Uçman ERGÜN

Within the scope of medical text mining, it is of great importance to make sense of complex statements, especially in radiology reports. These reports pose challenges for both automated analysis systems and clinical decision processes due to the various medical terms and long contextual relationships they contain. In this study, a hybrid model that can automatically extract entity names from radiology reports is developed. In the model, pre-trained language models such as BERT, BioBERT and ClinicalBERT are integrated with BiLSTM and CRF layers, and hyperparameter optimization is performed with GA and PSO. Within the scope of the method, different model configurations were compared and the combination that provided the highest success was determined. The training and validation processes were performed on the RadGraph dataset. In the evaluations performed on MIMIC-CXR and CheXpert test sets, the GA-optimized ClinicalBERT+BiLSTM+CRF model achieved F1 scores of 98,90% and 97,92%, respectively. The results show that the proposed approach works with high accuracy on clinical texts and can contribute to decision support systems.

2025, ix + 107 pages

Keywords: Natural Language Processing (NLP), Radiology reports, Named Entity Recognition (NER), Decision support systems

TEŞEKKÜR

Bu tez çalışmasının hazırlanmasında öneri ve rehberliğiyle her aşamada yanımda olan, bilimsel yaklaşımı ve desteğiyle çalışmalarımın şekillenmesinde rol oynayan danışmanım Doç. Dr. Uçman ERGÜN'e teşekkürlerimi sunarım.

Gerek akademik danışmanlığı gerekse manevi desteğiyle her zaman yanımda olan, tüm sorularımı sabırla yanıtlayan ve moral verici katkılarıyla motivasyonumu artıran Araştırma Görevlisi Sezin BARIN'a özel olarak teşekkür ederim.

Tez sürecinde moral ve manevi desteğini esirgemeyen, her koşulda yanımda olan Mehmet ÇAKMAKTEPE'ye teşekkür ederim.

Bu çalışmanın gerçekleştirilmesine olanak sağlayan TÜBİTAK'a, 2210-C Yurt İçi Lisansüstü Burs Programı kapsamında sağladığı maddi destek nedeniyle teşekkür ederim.

Son olarak, en önemli teşekkürümü hayatım boyunca bana koşulsuz sevgi, destek ve cesaret veren aileme ediyorum. Her zaman yanımda olan annem Hatice ORCİN, babam Mustafa ORCİN, abim Mehmet Ali ORCİN ve kardeşim Sinem ORCİN'e sonsuz teşekkür eder, bu çalışmayı onların sevgisine ithaf ederim.

Sedanur ORCİN

Afyonkarahisar 2025

İÇİNDEKİLER DİZİNİ

Sayfa

SİMGELER ve KISALTMALAR DİZİNİ

Simgeler

h_t	Gizli durum
x_t	Her bir giriş
$^{y}_{t}$	Sıralı bir etiket dizisi
Q	Sorgu matrisi
K	Anahtar matrisi
d_k	Anahtar vektörünün boyutu
e	Euler sayısı
W_O	Başlıkların çıktısını birleştiren ağırlık matrisi
h	Başlık sayısı
b	Bias terimi
max (0, ·)	ReLU aktivasyon fonksiyonu
pos	Kelimenin dizideki pozisyonu
i	Pozisyonel vektörün boyutundaki indeksi
d_{model}	Gömme vektörünün boyutudur.
Sublayer(x)	Dikkat mekanizması veya FFN gibi alt katmanların çıktı
γ ve β	Öğrenilebilir parametreler

Kısaltmalar

ESK	Elektronik Sağlık Kayıtları
NLP	Doğal Dil İşleme
NER	Varlık İsmi Tanıma
RNN	Tekrarlayan Sinir Ağı
LSTM	Uzun Kısa Süreli Bellek
ANAT-DP	Kesin Anatomik Yapı
OBS-DP	Kesin Gözlem
OBS-DA	Kesin Olarak Yok Gözlem
OBS-U	Belirsiz Gözlem
BILOU	Başlangıç, İç, Son, Dış, Tekil
BILSTM	Çift Yönlü Uzun Kısa Süreli Bellek Ağı
FFN	İleri Beslemeli Ağ
ReLU	Doğrultulmuş Lineer Birim
MLM	Maskeleme ile Dil Modellemesi
NSP	Sonraki Cümle Tahmini
RMSProp	Karekök Ortalama ile Yayılım Optimizasyonu
SGD	Stokastik Gradyan Azaltımı
CART	Sınıflandırma ve Regresyon Ağacı
SVM	Destek Vektör Makineleri
APSO	Uyarlanabilir Parçacık Sürü Optimizasyonu
MIMIC-III	Yoğun Bakım için Medikal Bilgi Veritabanı

ŞEKİLLER DİZİNİ

Sayfa

ÇİZELGELER DİZİNİ

Sayfa

1. GİRİŞ

Sağlık hizmetlerinin dijitalleşmesiyle birlikte tıbbi bilgi yönetimi süreçleri, geleneksel yaklaşımların ötesine geçerek daha sistematik, erişilebilir ve analiz edilebilir yapılar hâline gelmiştir. Önceki dönemlerde hasta kayıtlarının basılı belgeler, el yazısı notlar veya manuel sistemler aracılığıyla tutulması, bilgiye erişim sürecinde gecikmelere, hata riskinin artmasına ve karar destek mekanizmalarının etkinliğinin sınırlı kalmasına neden olmuştur (Preethi vd. 2021). Ayrıca, standardizasyon eksikliği ve kurumlar arası veri paylaşımının sınırlı olması hem hasta bakımının sürekliliğini sağlamayı hem de bilimsel araştırmaların gerçekleştirilmesini zorlaştırmıştır (Parameshwari vd. 2022).

Elektronik sağlık kayıtları (ESK) ve dijital tıbbi veri tabanlarının yaygınlaşması, sağlık sektöründe büyük hacimli yapılandırılmamış verilerin ortaya çıkmasına neden olmuştur (Yoon vd. 2024). Radyoloji raporları, doktor notları ve epikriz özetleri gibi metin tabanlı klinik belgeler, önemli miktarda bilgi içermelerine rağmen, yapısal olmamaları nedeniyle doğrudan analiz edilememektedir (Meier-Diedrich vd. 2025). Bu durum, özellikle klinik karar destek sistemlerinin işlevselliği açısından otomatik bilgi çıkarımı süreçlerinin geliştirilmesini zorunlu kılmaktadır.

Doğal dil işleme (NLP) yöntemleri, yapılandırılmamış tıbbi metinlerden anlamlı bilgi elde edilmesine yönelik etkili araçlar sunmaktadır. Bu yöntemler arasında, tıbbi varlıkların (hastalık, ilaç, prosedür, anatomik yapı vb.) tanımlanmasını sağlayan Varlık İsim Tanıma (NER) uygulamaları, klinik metin madenciliği ve otomatik raporlama sistemlerinde temel bileşenler olarak değerlendirilmektedir (Hu vd. 2024, Wierzbicki vd. 2024). Tıbbi metinlerde yer alan ifadelerin dilsel çeşitliliği, terim yoğunluğu ve bağlamsal karmaşıklığı, NER görevlerini diğer doğal dil işleme alanlarına kıyasla daha güç hâle getirmektedir. Geleneksel kural tabanlı ve istatistiksel yaklaşımlar, sınırlı örüntü tanımları ve yoğun insan müdahalesi gerektirmesi nedeniyle özellikle büyük ve heterojen veri kümelerinde yetersiz kalmaktadır (Huang vd. 2024, Magnini vd. 2025). Bu doğrultuda, bağlamsal ilişkileri temsil etmekte yüksek başarı gösteren derin öğrenme modellerinin kullanımı giderek yaygınlaşmaktadır (Torfi vd. 2020, Hu vd. 2025). Sıralı veri işleme görevlerinde öne çıkan Tekrarlayan Sinir Ağları (RNN) ve Uzun-Kısa Süreli

Bellek (LSTM) yapıları, tıbbi metin madenciliği alanındaki pek çok erken dönem çalışmada başarıyla uygulanmıştır. Bununla birlikte, söz konusu modellerin uzun süreli bağıntıları öğrenmedeki sınırlılıkları, daha derin bağlamsal ilişkileri yakalayabilen alternatif mimarilerin geliştirilmesini gerekli kılmıştır (Kandadi vd. 2025). Bu gereksinim doğrultusunda, Vaswani ve arkadaşları tarafından geliştirilen Transformer mimarisi (Vaswani vd. 2017), NLP alanında paradigmik bir dönüşüm sağlamış ve bağlamsal temsillerin paralel biçimde öğrenilebilmesini mümkün kılmıştır.

Transformer mimarisinin temel yapıtaşı olan "Self-Attention" mekanizması, modelin tüm kelimeler arasındaki ilişkileri paralel olarak değerlendirebilmesini sağlar. Bu yapı sayesinde, bağlamdan bağımsız tekil kelime temsilleri yerine, bağlama duyarlı ve semantik açıdan zengin temsiller elde edilmektedir (Kumar ve Solanki 2023). Bu mekanizmaya dayalı olarak geliştirilen BERT, çift yönlü öğrenme yetisi sayesinde kelimeler arasındaki bağlamı daha derinlikli analiz edebilmekte ve doğal dil işleme görevlerinde yüksek performans göstermektedir (Chandra vd. 2024). Tıbbi metin işleme alanında BERT'in özelleştirilmiş sürümleri olan BioBERT ve ClinicalBERT modelleri, sırasıyla biyomedikal literatür ve klinik notlar üzerinde ön eğitim alarak, alan terminolojisine özgü bağlamsal temsilleri daha isabetli biçimde öğrenebilmektedir (Mcinnes vd. 2025, Vakili vd. 2024). Bu modellerin BiLSTM ve CRF katmanlarıyla entegrasyonu ise dizisel etiketleme görevlerinde etiket geçişlerinin daha tutarlı biçimde tahmin edilmesine olanak tanımakta; bu da NER performansında anlamlı iyileştirmeler sağlamaktadır (Alamro vd. 2024, Yang vd. 2023, Lu vd. 2021).

Derin öğrenme modellerinin etkinliğini yalnızca mimari yapı değil, aynı zamanda eğitim sürecinde kullanılan hiper-parametrelerin doğru biçimde seçilmesi de doğrudan etkilemektedir. Grid Search (Izgara tarama) ve Random Search (rastgele tarama) gibi geleneksel yöntemler, yüksek boyutlu arama uzaylarında hesaplama açısından verimsiz kalmakta ve genellikle küresel optimum yerine yerel minimumlarda sonlanmaktadır (Ettaik ve Habib 2021). Bu nedenle, Genetik Algoritma (GA) ve Parçacık Sürü Optimizasyonu (PSO) gibi meta-sezgisel yöntemler, parametre uzayını daha esnek ve etkili biçimde tarayarak modelin genel başarımını optimize etmede önemli avantajlar sunmaktadır (Kollapally ve Geller 2023, Koshiry vd. 2024). Literatürdeki mevcut

çalışmalar dil modelleme ve hiper-parametre optimizasyon süreçlerini genellikle birbirinden bağımsız ele almaktadır. Bu durum, uzun ve bağlamsal açıdan zengin tıbbi metinlerde NER başarımını sınırlamaktadır (Pakhale 2023). Aynı şekilde, çoğu çalışmada hiper-parametreler önceden belirlenmiş sabit değerlere dayalı olarak uygulanmakta ve bu durum modelin farklı veri kümeleri üzerindeki genelleme kapasitesini zayıflatmaktadır.

Bu tez çalışmasının temel amacı, tıbbi metinlerden etkin ve doğru varlık çıkarımı yapabilen hibrit modeller geliştirmek ve bu modellerin hiper-parametrelerini meta-sezgisel algoritmalar aracılığıyla optimize ederek yüksek doğruluk, kararlılık ve genelleme yeteneği elde etmektir. Önerilen mimarilerde, bağlamsal temsili güçlü olan BERT, BioBERT ve ClinicalBERT modelleri ile sıralı etiketleme görevlerinde etkinliği kanıtlanmış BiLSTM ve CRF katmanları bir araya getirilmiştir. Bu bütünleşik yapı sayesinde hem semantik hem de sentaktik ilişkilerin daha etkili biçimde modellenmesi hedeflenmiştir. Modelin başarımını artırmak ve parametre ayarlarının rastgele belirlenmesini önlemek amacıyla GA ve PSO teknikleri kullanılarak hiper-parametre optimizasyonu gerçekleştirilmiştir. Bu yaklaşım, geleneksel yöntemlerin sınırlı kaldığı geniş parametre uzaylarının daha verimli şekilde taranmasına olanak tanımıştır.

Çalışma kapsamında, RadGraph veri seti (Jain vd. 2021), kullanılarak eğitim ve doğrulama süreçleri gerçekleştirilmiş, MIMIC-CXR ve CheXpert test setleri üzerinde performans değerlendirmeleri yapılmıştır. Elde edilen deneysel sonuçlar, GA ile optimize edilen ClinicalBERT+BiLSTM+CRF modelinin sırasıyla 0,9890 (MIMIC-CXR) ve 0,9792 (CheXpert) F1 skorları ile literatürdeki güncel yaklaşımlara kıyasla üstün performans sergilediğini ortaya koymaktadır. Bu çalışma, iki temel özgün katkı sunmaktadır: (1) Klinik bağlamda önceden eğitilmiş dil modellerinin sıralı etiketleme katmanları ile entegrasyonu sayesinde tıbbi varlık tanıma görevlerinde daha derin bağlamsal analiz yapılabilmesi, (2) hiper-parametre optimizasyonunun meta-sezgisel algoritmalar aracılığıyla gerçekleştirilmesi sayesinde modelin genel başarımının ve adaptasyon kabiliyetinin artırılması. Bu yönleriyle önerilen yöntem, hem tıbbi metin madenciliği alanında bilimsel literatüre katkı sunmakta hem de klinik karar destek sistemlerinin güvenilirliğini artırabilecek nitelikte bir çözüm önermektedir.

2. LİTERATÜR BİLGİLERİ

2.1 Klinik Metinler ve Elektronik Sağlık Kayıtları

Dijital sağlık teknolojilerindeki gelişmelerle birlikte, sağlık kurumlarında hasta verilerinin toplanması, saklanması ve yönetilmesi süreçleri büyük ölçüde dijital ortamlara taşınmıştır. ESK sistemleri, hasta bilgilerini sistematik şekilde kaydeden ve sağlık hizmetlerinin sürekliliğini sağlayan temel bilgi yönetim araçları haline gelmiştir (Meystre vd. 2008). ESK sistemleri; tanılar, reçeteler, laboratuvar sonuçları, vital bulgular ve hekim notları gibi çok çeşitli klinik verileri içermekte, bu yönüyle hem bireysel hasta takibi hem de araştırma faaliyetleri için önemli bir kaynak oluşturmaktadır. Şekil 2.1'de Radgraph (Jain vd. 2021) veri setinde yer alan İngilizce dilinde örnek bir radyoloji raporu sunulmuştur.

FINAL REPORT EXAMINATION:

CHEST (PORTABLE AP) INDICATION:

History: ___ F with at placement History: Upright AP view of the chest COMPARISON: None. Patient is currently listed as EU critical. FINDINGS: Endotracheal tube tip terminates approximately a 4,6 cm from the carina. Enteric tube tip terminates within the distal esophagus and should be advanced by at least 11 cm. Heart size is normal. The mediastinal and hilar contours are grossly unremarkable. Pulmonary vasculature is not engorged. Apart from minimal atelectasis in the lung bases, the lungs appear clear. No large pleural effusion or pneumothorax is present though the extreme right apex is slightly obscured by the patient 's chin projecting over this area. No displaced fractures are present. No acute osseous abnormality is seen. Gaseous distension of the stomach is noted. IMPRESSION: 1. Standard positioning of the endotracheal tube. 2. Enteric tube tip is sub optimally located within the distal esophagus and should be advanced by at least 11 cm. Gaseous distention of the stomach is also noted. 3. No focal consolidation.

Şekil 2.1 İngilizce dilinde radyoloji raporu örneği.

ESK içerisinde yer alan radyoloji raporları, klinik karar destek süreçlerinde kritik rol oynamaktadır. Bu raporlar, genellikle radyologlar tarafından görüntüleme sonuçlarına dayanılarak yazılan serbest metinlerden oluşur. Radyoloji raporları; anatomik bölgeler, radyolojik bulgular, patolojik değerlendirmeler ve önerilen ileri tetkikler gibi unsurları barındırır (Nobel vd. 2022).

Klinik metinlerin işlenmesinde karşılaşılan başlıca zorluklardan biri, terminolojik çeşitlilik ve dilsel düzensizliktir. Aynı klinik kavram, farklı uzmanlar tarafından farklı

terimlerle ifade edilebilmekte; kısaltmalar, yazım hataları, eksik sözdizimi ve bağlama özgü jargon kullanımı gibi faktörler metinlerin anlamlandırılmasını zorlaştırmaktadır (Landolsi vd. 2023). Bu nedenle, klinik serbest metinlerin makine tarafından işlenebilir hale getirilmesi yalnızca araştırma ve geliştirme faaliyetleri için değil, aynı zamanda klinik karar destek sistemlerinin güvenilirliğini ve etkinliğini artırmak açısından da stratejik bir gerekliliktir. Bu bağlamda, doğal dil işleme yöntemleri, özellikle NER ve ilişki çıkarımı gibi alt görevler aracılığıyla klinik metinlerden anlamlı, yapılandırılmış ve yeniden kullanılabilir bilgi üretimini mümkün kılmaktadır. Son yıllarda geliştirilen derin öğrenme tabanlı NLP yaklaşımları, bu süreci otomatikleştirmede önemli ilerlemeler sağlamış ve sağlık bilişimi alanında yeni bir paradigma oluşturmuştur (Hahn ve Oleynik 2020).

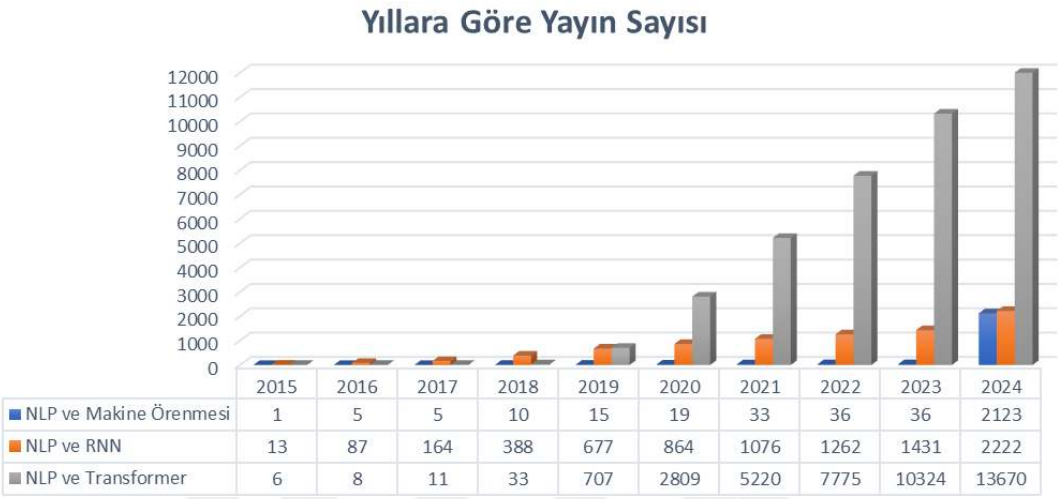
2.2 Doğal Dil İşleme (NLP)

NLP, insan dilinin bilgisayar sistemleri tarafından anlaşılması, analiz edilmesi ve üretilebilmesini amaçlayan yapay zekâ alt alanıdır. Temel hedefi, bilgisayarların doğal dili insanlar gibi anlamlandırarak çeşitli görevleri yerine getirebilmesidir. Bu bağlamda NLP; dil modelleme, metin sınıflandırma, bilgi çıkarımı, konuşma tanıma ve metin üretimi gibi çok sayıda görevi içermektedir (Dogra vd. 2022).

NLP'nin gelişimi üç temel evreden geçmiştir: Başlangıçta dilbilgisel kurallara dayanan yaklaşımlar, daha sonra teknikler istatistiksel modellerle desteklenmiş; son olarak derin öğrenme tabanlı mimariler (özellikle BERT, GPT gibi Transformer temelli modeller) ile insan düzeyine yakın başarılar elde edilmiştir (Shen ve Kejriwal 2023).

Bu evrimi destekleyen bibliyometrik bulgular da literatürdeki yayın eğilimlerinin zamanla nasıl değiştiğini gözler önüne sermektedir. Scopus veri tabanında 2015–2024 yılları arasında "NLP ve Makine Öğrenmesi", "NLP ve RNN" ve "NLP ve Transformer" anahtar kelimeleriyle yapılan yayın sayılarına ilişkin karşılaştırmalı bir analiz Şekil 2.2'de sunulmuştur. Bu arama Nisan 2025 tarihi itibariyle yapılmış olup, 2025 yılına ait makale sayıları net olmadığı için grafikte yer almamıştır**.** Yayın sayıları analiz edildiğinde özellikle Transformer tabanlı modellerin 2018 sonrası belirgin

biçimde öne çıktığı görülmektedir. Bu eğilim, derin öğrenme mimarilerinin yalnızca akademik alanda değil, aynı zamanda klinik uygulamalarda da hızla yaygınlaştığını göstermektedir (Rezaeenour vd. 2023).



	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
NLP ve Makine Örenmesi	1	5	5	10	15	19	33	36	36	2123
NLP ve RNN	13	87	164	388	677	864	1076	1262	1431	2222
NLP ve Transformer	6	8	11	33	707	2809	5220	7775	10324	13670

Şekil 2.2 2015-2024 yılları arası yayın sayıları.

Şekil 2.2 NLP'nin mimari gelişimini sayısal olarak ortaya koyarken, bu ilerleme uygulama çeşitliliğini de beraberinde getirmiştir. NLP'nin güncel kullanım alanları Şekil 2.3'de özetlenmektedir.

Görselde özetlendiği üzere, doğal dil işleme sistemlerinin temel görev alanları çok yönlü olup, dilin hem yapısal hem de anlamsal bileşenlerini kapsamaktadır. Bu görevler, farklı sektörlerdeki uygulamalara entegre edilerek insan-bilgisayar etkileşimini önemli ölçüde iyileştirmektedir.

- Adlandırılmış Varlık Tanıma: Metin içerisindeki kişi adları, yer isimleri, kuruluşlar, hastalıklar gibi özel anlam taşıyan varlıkların otomatik olarak tespit edilerek etiketlenmesini sağlar. Bu görev, bilgi çıkarımı ve bilgiye dayalı karar destek sistemleri için temel bir yapı taşını oluşturmaktadır.
- Duygu Analizi: Metinlerdeki duygusal tonun (örneğin pozitif, negatif veya nötr) belirlenmesine yönelik bir görevdir. Sosyal medya analizleri, müşteri geri bildirimlerinin değerlendirilmesi ve kamuoyu yoklamalarında yaygın biçimde kullanılmaktadır.

- Metin Sınıflandırma: Belirli bir amaca yönelik olarak metinlerin önceden tanımlı kategorilere ayrılması işlemidir. Örneğin, e-postaların spam/is olarak ayrılması ya da haber başlıklarının konuya göre sınıflandırılması bu kapsamda değerlendirilmektedir.
- Cümle Analizi: Cümlelerin dilbilgisel yapılarının ve anlamlarının analiz edilmesini içermektedir. Bu görev, dilin yapısal kurallarına uygun çözümler üretmekte ve karmaşık cümle yapılarının bilgisayarlar tarafından işlenmesini mümkün kılmaktadır.
- Cümle-Metin Tamamlama**:** Eksik cümle ya da metin parçalarının bağlama uygun biçimde otomatik olarak tamamlanmasını amaçlar. Dil modeli eğitimi ve otomatik yazım destek sistemlerinde sıklıkla kullanılmaktadır.
- Çeviri: Bir dildeki metnin, anlamını koruyacak biçimde başka bir dile otomatik olarak çevrilmesini sağlar. Küresel iletişim ve çok dilli sistemlerde temel NLP görevlerinden biridir.
- Metin Özetleme: Uzun metinlerin anlam bütünlüğünü bozmadan daha kısa ve öz biçimde yeniden yapılandırılmasını ifade eder. Bu görev, özellikle bilgiye hızlı erişim gereken alanlarda (ör. haber, tıbbi belgeler) önemli rol oynamaktadır.
- Konuşma Tanıma: Sesli girdilerin yazılı metne dönüştürülmesini sağlayan görevdir. Sesli asistanlar, hasta-hekim görüşme çözümleri ve çağrı merkezi sistemlerinde yaygın olarak kullanılmaktadır.
- Metin Üretimi: Yapılandırılmış verilerin (örneğin sayısal tablolar, grafik verileri) doğal dil cümleleri hâline dönüştürülmesini sağlar. Raporlama sistemleri ve otomatik içerik üretiminde sıklıkla tercih edilmektedir.

Bu görevlerin her biri, eğitim, sağlık, medya, hukuk ve e-ticaret başta olmak üzere pek çok sektörde yaygın biçimde kullanılmakta; veri analitiği, karar destek sistemleri ve insan-makine iletişiminin etkinliğini artırmaktadır.

Şekil 2.3 NLP sistemlerinin başlıca görev alanları.

2.3 Varlık İsmi Tanıma (NER)

NER, metin içerisinde yer alan özel anlam taşıyan birimlerin (örneğin kişi adları, lokasyonlar, organizasyonlar, tarih ve zaman ifadeleri gibi) otomatik olarak tespit edilmesi ve etiketlenmesini amaçlayan temel bir NLP alt görevdir. NER sistemleri, metin madenciliği, bilgi çıkarımı ve anlamsal anlama gibi birçok üst seviye NLP görevine temel teşkil etmektedir (Miwa ve Bansal 2016, Singh 2018).

Bu alan ilk olarak 1990'lı yıllarda Message Understanding Conference serileri kapsamında ön plana çıkmış ve başlangıçta kural tabanlı sistemlerle uygulanmıştır (Chinchor ve Marsh 1998). Ancak bu yaklaşımlar yüksek düzeyde manuel özelleştirme gerektirmesi ile dil ve alan bağımlılığı gibi önemli sınırlılıklar taşımaktaydı. Sonraki yıllarda ise CRF (Lafferty vd. 2001) ve Hidden Markov Modelleri (HMM) gibi istatistiksel yöntemler ön plana çıkarak bağlamsal yapıların daha esnek şekilde modellenmesini mümkün kılmıştır. HMM, bir gözlem dizisinin ardında yatan gizli durumları modelleyerek çalışır ve NER bağlamında bu gizli durumlar, kelimelerin ait olduğu varlık etiketlerini temsil eder. HMM'ler, kelime dizilerinde önceki duruma

dayalı olarak geçiş olasılıklarını ve gözlem olasılıklarını kullanarak etiketleme yapar (Rabiner 1989). Ancak HMM modelleri yalnızca geçmiş durumları dikkate aldığından bağlamsal bütünlüğü sınırlı düzeyde yansıtabilir. Bu sınırlamaları aşmak üzere geliştirilen CRF modeli, etiketler arasındaki geçiş ilişkilerini ve gözlemlenen özellikleri birlikte dikkate alarak daha başarılı sonuçlar sunmuştur. Sonraki dönemde ise derin öğrenme temelli çözümler, alanın dinamiklerini kökten değiştirerek, kelimeler arası bağlamsal ilişkileri daha etkili biçimde modelleyen mimarilerle birlikte daha yüksek doğruluk oranlarına ulaşılmasını mümkün kılmıştır (Yadav ve Bethard 2019).

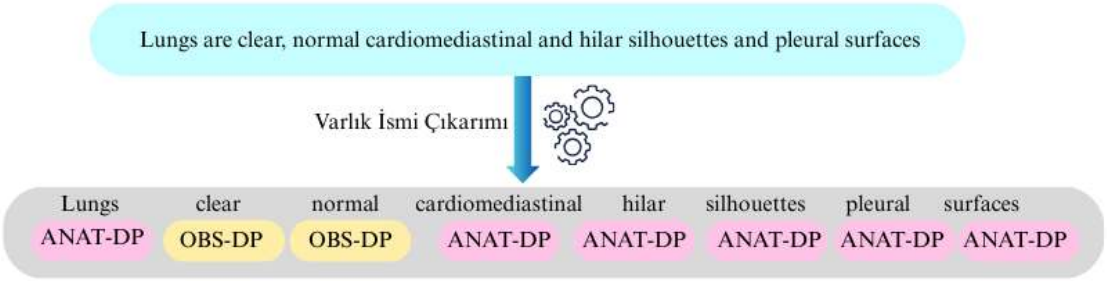
Bu gelişmelerin ardından, derin öğrenme tabanlı yöntemler, özellikle 2010'ların ortasından itibaren NER alanında yeni bir dönem başlatmıştır. RNN, LSTM ve çift yönlü versiyonları olan BiLSTM gibi mimariler, kelimeler arası bağlamsal ilişkileri çok daha etkili biçimde modelleme kapasitesine sahiptir (Lample vd. 2016). Bu süreçte en radikal kırılma ise 2017 yılında tanıtılan Transformer mimarisi ile yaşanmıştır (Vaswani vd. 2017). Özellikle ön eğitimli dil modelleri olan BERT (Devlin vd. 2019), BioBERT (Lee vd. 2020) ve ClinicalBERT (Alsentzer vd. 2019) gibi modeller, bağlamı çift yönlü dikkate almaları ve paralel hesaplama yetenekleri sayesinde, klinik NER görevlerinde son derece başarılı sonuçlar vermiştir (Turchin vd. 2023).

Bu teknolojik ilerlemeler, sağlık bilişimi alanında yapılandırılmamış metinlerden bilgi çıkarımı açısından büyük önem taşımaktadır. ESK, radyoloji raporları, hekim notları ve epikrizler gibi belgeler serbest biçimli ve yüksek hacimli olduğundan, geleneksel analiz yöntemleriyle verimli şekilde işlenmeleri oldukça güçtür (Meystre vd. 2008). Bu noktada NER sistemleri, tıbbi metinlerden anlamlı bilgi çıkarımı için kritik bir rol üstlenmektedir

Tıbbi doğal dil işleme uygulamaları içerisinde en yaygın ve etkili biçimde kullanılan görevlerden biri NER'dir. Bu görev; hastalık, semptom, ilaç, tanı testleri ve tedavi yöntemleri gibi kavramların elektronik sağlık kayıtları içerisinden otomatik olarak tanımlanmasını ve yapılandırılmasını sağlamaktadır. Sheikhalishahi ve arkadaşları, NER uygulamalarının özellikle kronik hastalıkların takibi, klinik karar destek sistemlerinin güçlendirilmesi ve tıbbi araştırmalara yönelik yapılandırılmış veri

üretiminde yüksek doğruluk sunduğunu vurgulamaktadır (Sheikhalishahi vd. 2019). Bu sistemler sayesinde hem manuel analiz yükü azalmış hem de hata oranları minimize edilmiştir. Ayrıca, geçmişe dönük hasta analizlerinden proaktif risk tahminine kadar geniş bir kullanım yelpazesi oluşmuştur.

NER sistemlerinin işleyişi, aşağıda sunulan örnekle daha net anlaşılabilir. Şekil 2.4'de, "Lungs are clear, normal cardiomediastinal and hilar silhouettes and pleural surfaces" ifadesi üzerinden yapılan etiketleme süreci gösterilmektedir. Bu ifadede; 'Lungs', 'cardiomediastinal', 'hilar', 'silhouettes', 'pleural' ve 'surfaces' gibi kelimeler ANAT-DP (Anatomik Yapılar), 'clear' ve 'normal' gibi ifadeler ise OBS-DP (Gözlemsel Bulgular) olarak etiketlenmiştir. Bu ayrım, metinlerdeki tıbbi bilgilerin yapısal hale getirilmesini kolaylaştırmakta ve semantik analizlerin doğruluğunu artırmaktadır (Hossain vd., 2023).



Şekil 2.4 Örnek bir klinik metinde NER süreci ve etiketleme örneği.

2.4 NER Görevinde Kullanılan Veri Setleri ve Etiketleme Şemaları

NER görevlerinde model başarımı, büyük ölçüde kullanılan veri setlerinin niteliğine ve kapsamına bağlıdır. Gerek eğitim gerekse değerlendirme aşamasında yüksek kaliteli, etiketlenmiş klinik veri kümeleri, modellerin gerçek dünya senaryolarında daha güvenilir ve genellenebilir sonuçlar üretmesini sağlamaktadır. Özellikle biyomedikal alanda, yapılandırılmamış serbest metinlerin etiketlenmesi oldukça zahmetli ve uzmanlık gerektiren bir süreçtir. Bu nedenle, açık erişimli ve farklı tıbbi alanları kapsayan etiketli veri setleri, NER araştırmaları için hayati öneme sahiptir (Lee vd. 2020).

Son yıllarda, İngilizce klinik metinlerde NER görevleri için yaygın biçimde kullanılan pek çok açık veri seti geliştirilmiştir. Çizelge 2.1'de literatürde yer alan bazı veri setleri sunulmuştur. MIMIC-CXR, CheXpert ve RadGraph gibi geniş kapsamlı radyoloji veri setleri, özellikle göğüs hastalıklarına odaklanırken; CHIFIR, invazif mantar enfeksiyonları ve CORAL, pankreas ve meme onkolojisi gibi daha özel klinik senaryolara özgülenmiştir. Bu veri setlerinin önemli bir avantajı, yalnızca görüntüleme raporlarını içermekle kalmayıp, aynı zamanda raporlar içerisindeki semantik varlıkları etiketlenmiş biçimde sunmalarıdır. Böylece, tıbbi anlam taşıyan ifadelerin otomatik çıkarımı ve sınıflandırılması mümkün hale gelmektedir. Ayrıca, bu veri kümeleri üzerinde eğitilen modeller, ileri düzey NER görevlerinde, özellikle de varlık tipine özgü ayırt edici örüntülerin öğrenilmesinde yüksek başarı oranları sunmaktadır. Bu sayede, geliştirilen sistemlerin hem tanı süreçlerini desteklemesi hem de araştırma ve raporlama süreçlerini otomatikleştirmesi mümkün hâle gelmektedir.

Çizelge 2.1 Literatürde yer alan veri setleri.

Veri Seti	Dil	Veri Sayısı	Bölge	URL
MIMIC-CXR Veri Seti (Johnson vd, 2019)	İngilizce	227.835	Göğüs	(İnt.Kyn.1)
CHEXPERT Veri Seti (Irvin vd., 2019)	İngilizce	65.240	Göğüs	(İnt.Kyn.2)
CHIFIR Veri Seti (Rozova vd., 2023)	İngilizce	283	İnvazif Mantar	(İnt.Kyn.3)
RadGraph Veri Seti (Jain vd., 2021)	İngilizce	547	Göğüs	(İnt.Kyn.4)
CORAL Veri Seti (Sushil vd., 2024)	İngilizce	40	Pankreas ve Meme	(İnt.Kyn.4)

Bu veri setlerinin her biri, kendi içinde belirli bir klinik bağlama (örneğin göğüs radyolojisi, onkoloji, nöroloji) odaklanmakta ve farklı varlık türlerinin çıkarımını mümkün kılacak şekilde özenle etiketlenmiştir. Bu etiketleme süreci, NER modellerinin eğitilmesi ve değerlendirilmesinde belirleyici bir rol oynamaktadır. Ancak bu sürecin başarılı olabilmesi için yalnızca nitelikli veri seti kullanımı yeterli değildir; verilerin nasıl etiketlendiği, yani kullanılan etiketleme şeması, model performansı üzerinde doğrudan etkili olan kritik bir bileşendir (Katona ve Farkas 2014).

Klinik NER sistemlerinde yaygın olarak kullanılan etiketleme şemaları arasında BIO (Beginning, Inside, Outside) ve BILOU (Beginning, Inside, Last, Outside, Unit) biçimleri öne çıkmaktadır. Bu şemalar, her kelimenin bulunduğu konuma ve ait olduğu varlığa göre sistematik bir şekilde etiketlenmesini sağlar (Marquez vd.2005).

- BIO şeması, her varlık için yalnızca üç etiket türü kullanır:
  - B- (Beginning): Varlık adının ilk kelimesi,
  - I- (Inside): Aynı varlığa ait devam eden kelimeler,
  - O- (Outside): Herhangi bir varlıkla ilişkili olmayan kelimeler.

Örneğin, "akciğer nodülü" ifadesinde "akciğer" kelimesi B-ANAT, "nodülü" ise I-ANAT etiketiyle işaretlenir. Bu yapı, sadeliği sayesinde özellikle büyük ölçekli veri setlerinde hızlı etiketleme ve modelleme imkânı sunar. Ancak, çok kelimeli varlıkların sınırlarını hassas biçimde belirlemede sınırlı kalabilir.

- BILOU şeması, bu sınırlamaları aşmak için geliştirilmiş daha ayrıntılı bir yaklaşımdır:
  - B- (Beginning), I- (Inside), L- (Last): Çok kelimeli bir varlığın başlangıç, iç ve son kelimelerini,
  - U- (Unit): Tek kelimelik varlıkları,
  - O- (Outside): Varlık dışı kelimeleri ifade eder.

Örneğin, "hemorajik inme" ifadesi B-DISEASE, L-DISEASE şeklinde etiketlenebilirken, "trombüs" gibi tek kelimelik bir varlık U-DISEASE etiketiyle temsil edilir. Bu ayrım, modellerin varlık uzunluklarına ve yapısal sınırlarına daha duyarlı hale gelmesini sağlar (Malik vd. 2016).

Etiketleme şeması seçimi, çoğu zaman veri setinin dil özelliklerine, varlık türlerinin ortalama uzunluğuna ve kullanılan model mimarisine göre belirlenir. Örneğin, transformer tabanlı modeller bağlamsal temsilleri daha iyi öğrendiği için, BILOU gibi ayrıntılı şemalarla birlikte daha yüksek doğruluk oranlarına ulaşabilmektedir (Eberts ve Ulges 2020). Öte yandan, BIO şeması hâlâ pek çok çalışmada tercih edilmekte; çünkü daha az parametre gerektirmesi, sade yapısı ve düşük hesaplama maliyeti ile küçük veri setleri veya daha basit görevler için yeterli performans sunabilmektedir (Ozyurt 2020).

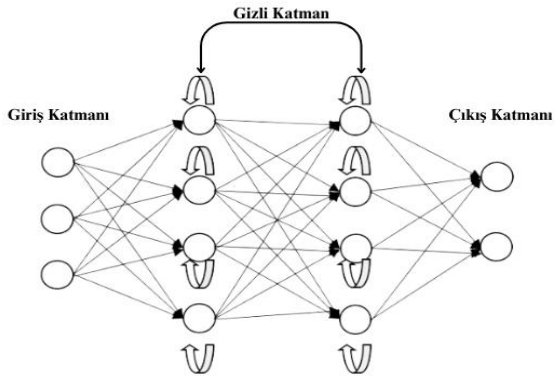
Ayrıca şema seçiminin etkisi sadece doğrulukla sınırlı olmayıp; yanlış negatif üretme eğilimi, varlık sınırlarını hatalı belirleme riski ve tahmin belirsizliği gibi durumlarda da belirleyici rol oynayabilir. Bu nedenle, klinik metinlerdeki karmaşık yapılar göz önünde bulundurulduğunda, şema seçimi hem istatiksel başarı hem de pratik kullanılabilirlik açısından dikkatle değerlendirilmelidir.

2.5 Derin Öğrenme ve Transformer Tabanlı Modeller

2.5.1 Tekrarlayan Sinir Ağları (RNN)

NLP alanında kelimeler arasındaki sıralı ilişkileri modellemek için geliştirilen ilk yapay sinir ağı mimarilerinden biri olan RNN, sıralı veri üzerinde öğrenme gerçekleştirebilme kapasitesiyle dikkat çekmiştir (Elman, 1990). Bu mimari, her bir zaman adımındaki girdiyi işlerken, önceki zaman adımındaki bağlamsal bilgiyi (gizil durumu) de hesaba katarak çalışır. Böylece, kelime dizileri gibi sıralı yapılarda bağlamsal anlamın öğrenilmesine olanak tanır.

RNN'lerin temel avantajı, her giriş verisinin yalnızca kendi değerine bağlı olarak değil, aynı zamanda geçmişteki verilere dayalı olarak yorumlanmasıdır. Bu sayede model, bir kelimenin anlamını belirlerken onu önceki kelimelerle birlikte değerlendirme yeteneği kazanır. Özellikle metin sınıflandırma, makine çevirisi ve NER gibi ardıl işlemler gerektiren görevlerde bu yapıdan sıklıkla yararlanılmaktadır. RNN mimarisi genel olarak üç temel katmandan oluşur: giriş katmanı, gizli katman ve çıkış katmanı. Bu katmanların mimari düzeni Şekil 2.5'te gösterilmektedir (Mienye vd. 2024).



Şekil 2.5 RNN mimarisinin genel katman yapısı.

Giriş katmanında model, her kelimeyi sayısal olarak temsil edebilmek için genellikle önceden eğitilmiş kelime gömme (embedding) vektörlerini kullanır. Örneğin, "akciğer grafisi temiz" gibi bir cümledeki her kelime, belirli bir vektörle gösterilerek dizisel giriş elde edilir. Bu giriş dizisi, sırasıyla gizli katmana iletilir.

Gizli katman, RNN'in esas öğrenme işleminin gerçekleştiği katmandır. Her nöron, yalnızca kendi girdisine değil, aynı zamanda bir önceki zaman adımındaki gizil duruma da bağlıdır. Bu geri döngüsel bağlantılar (recurrent connections), modelin geçmiş bağlam bilgisini bellekte tutmasını sağlar. Böylece ağ, giriş sırasındaki kelimelerin bağlamını dikkate alarak çıktı üretir.

Gizli durumların güncellenmesi süreci, aşağıdaki matematiksel denklem ile tanımlanır.

$$h_t = \tanh\left(W_{xh}x_t + W_{hh}h_{t-1} + b_n\right) \tag{2.1}$$

Burada;

h_t: Zaman adımı t'deki gizil durum,

x_t: Zaman adımı t'deki giriş vektörü,

h_{t-1}: Bir önceki zaman adımının gizil durumu,

W_{xh} = Girişten gizil duruma olan ağırlık matrisi,

W_{hh}= Önceki gizil durumdan geçiş ağırlığı,

b_h = Bias terimi,

Tanh = Aktivasyon fonksiyonu olarak kullanılır.

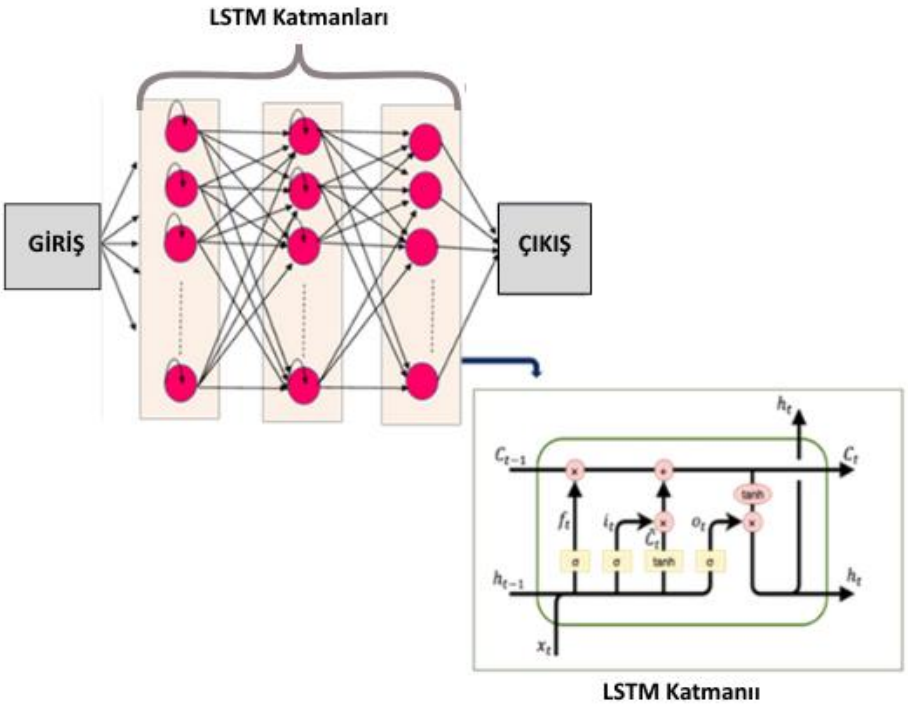
Bu yapı, her zaman adımında önceki bilginin aktarılmasını mümkün kılar; ancak bu aktarım zamanla zayıflayabilir.

Her gizli durum, görev bağlamına bağlı olarak çıkış katmanına aktarılır. Örneğin NER görevinde, her bir kelimenin hangi sınıfa ait olduğunu tahmin eden bir softmax katmanı kullanılır. Bu katman, etiket olarak örneğin B-DISEASE, I-ANAT, O gibi çıktıların üretilmesini sağlar.

2.5.2 Uzun Kısa Süreli Bellek (LSTM)

RNN mimarileri sıralı verilerde kısa vadeli bağımlılıkları başarıyla öğrenebilse de uzun dizilerde bilgi taşıma yetenekleri gradyan kaybolması (vanishing gradient) problemi nedeniyle ciddi şekilde sınırlıdır (Bengio vd. 1994). Bu sorunu aşmak amacıyla geliştirilen LSTM ağı, RNN mimarisinin bir uzantısı olup, uzun-bağlam bağımlılıklarını daha etkili biçimde modelleyebilme yeteneğine sahiptir (Hochreiter ve Schmidhuber, 1997).

LSTM, klasik RNN'den farklı olarak, her zaman adımında bilgiyi taşıyıp unutmaya karar verebilen özel bir hücre yapısına sahiptir. Bu yapı, ağın önemli bilgileri uzun süre boyunca belleğinde tutmasını ve önemsiz bilgileri unutmasını sağlar. Şekil 2.6'da LSTM mimarisinin genel yapısı ve hücresel işleyişi sunulmuştur.



Şekil 2.6 LSTM mimarisinin genel katman ve hücre yapısı.

LSTM ağı, giriş vektörlerini birden fazla LSTM katmanından geçirerek çıktı üretir. Her bir katmanda, belirli bir zaman adımındaki giriş (x_t) ve bir önceki zaman adımındaki gizli durum (h_{t-1}) birlikte işlenir. Klasik RNN hücresinden farklı olarak, LSTM

yapısında cell state (hücre durumu) adı verilen ek bir taşıyıcı bileşen bulunur ve bu kanal, bilgiyi zaman boyunca kayıpsız şekilde taşımak üzere tasarlanmıştır. Cell state yapısının güncellenmesi, hücre içindeki üç temel kapı mekanizması aracılığıyla yönetilir (Tarwani ve Edem 2017).

İlk adımda, unutma kapısı (forget gate) hücre durumunda hangi bilgilerin tutulup hangilerinin unutulacağına karar verir. Bu mekanizma sigmoid aktivasyon fonksiyonu ile tanımlanır ve matematiksel olarak şu şekilde ifade edilir (Lu vd. 2021):

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{2.2}$$

Burada:

f_t: Zaman adımı ttt'deki unutma kapısı çıktısıdır.

σ: Sigmoid aktivasyon fonksiyonudur.

W_f: Unutma kapısı ağırlık matrisidir.

h_{t-1}: Bir önceki zaman adımındaki gizli durumdur.

x_t: O anki giriş vektörüdür.

b_f: Unutma kapısı bias terimidir.

Ardından giriş kapısı devreye girerek, yeni bilgilerin hücre durumuna nasıl ekleneceğini kontrol eder. Bu süreç iki adımdan oluşur. Öncelikle, hangi bilginin ekleneceği sigmoid fonksiyonu ile belirlenir:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2.3}$$

Burada:

i_t: Giriş kapısının çıktısıdır.

W_i: Giriş kapısı ağırlık matrisidir.

b_i: Giriş kapısı bias terimidir.

Daha sonra, eklenecek aday hücre durumu *tanh* aktivasyonu kullanılarak üretilir:

$$\tilde{C}_t = tanh(W_C[h_{t-1}, x_t] + b_C) \tag{2.4}$$

Burada:

$\tilde{C}_t$: O adımda eklenmesi düşünülen aday hücre durumu vektörüdür.

W_C: Hücre durumu için ağırlık matrisidir.

b_C: Hücre durumu bias terimidir.

Bu adımlar sonucunda yeni hücre durumu *(C_t)* şu şekilde güncellenir:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t) \tag{2.5}$$

Burada:

C_t*:* Zaman adımı *t*'deki güncellenmiş hücre durumudur.

C_{t-1}*:* Bir önceki zaman adımındaki hücre durumudur.

$*$: Eleman bazında çarpma işlemini temsil eder.

Son adımda, çıkış kapısı hücrenin üreteceği çıktıyı belirler. Öncelikle çıkış kapısının değeri hesaplanır:

$$o_t = \sigma(W_o[h_{t-1}, x_t)] + b_o) \tag{2.6}$$

Burada:

O_t: Çıkış kapısının çıktısıdır.

W_o: Çıkış kapısı ağırlık matrisidir.

b_o*:* Çıkış kapısı bias terimidir.

Son olarak, hücrenin yeni gizli durumu *(h_t)* hesaplanır:

$$h_t = o_t * \tanh(C_t) \tag{2.7}$$

Burada:

h_t*:* Zaman adımı *t*'deki yeni gizli durumdur.

t*anh*: Hiperbolik tanjant aktivasyon fonksiyonudur.

Bu çıktı hem anlık tahminler için kullanılır hem de bir sonraki zaman adımı için gizli durum olarak devredilir. LSTM yapılarının bu kapı mekanizmaları, modelin hem kısa vadeli hem de uzun vadeli bağımlılıkları aynı anda öğrenebilmesine olanak

sağlamaktadır. Bu özellikler, LSTM'i doğal dil işleme alanında, metin sınıflandırma, makine çevirisi, duygu analizi ve NER gibi görevlerde yaygın ve etkili bir yöntem hâline getirmiştir (Rahman vd. 2021).

2.5.3 Çift Yönlü Uzun Kısa Süreli Bellek (BiLSTM)

LSTM mimarileri, zaman içerisindeki ardışık bağımlılıkları modellemede önemli ilerlemeler sağlamış olsa da yalnızca geçmiş bilgiye dayalı çalışmaları nedeniyle gelecekteki bağlam bilgisinden yararlanamamaktadır. Oysa doğal dil gibi sıralı veri yapılarında bir kelimenin anlamı hem öncesinde hem de sonrasında gelen kelimelerle birlikte şekillenmektedir. Bu nedenle, yalnızca ileri yönde bilgi işleyen standart LSTM yapıları, özellikle uzun ve karmaşık cümlelerde bağlamı tam olarak kavrayamamakta; önemli bilgileri kaçırabilmektedir (Kandadi ve Shankarlingam 2025). Bu eksikliği gidermek amacıyla geliştirilen BiLSTM yapıları, her zaman adımında girdiyi hem ileri hem de geri yönde işleyerek daha zengin bağlamsal temsiller üretmeyi amaçlamaktadır (Schuster ve Paliwal 1997).

BiLSTM yapısında, giriş vektörleri $(x_1, x_2, ..., x_n)$ bir ileri LSTM hücresinden ve bir geri LSTM hücresinden geçirilir. İleri yönlü LSTM, diziyi başlangıçtan sona doğru işlerken, geri yönlü LSTM diziyi sondan başa doğru işler. Her iki yönden elde edilen gizli durumlar daha sonra birleştirilerek tek bir çıktı vektörü oluşturulur (Xu vd. 2019). Bu işlem matematiksel olarak şu şekilde ifade edilmektedir:

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}] \tag{2.8}$$

Burada:

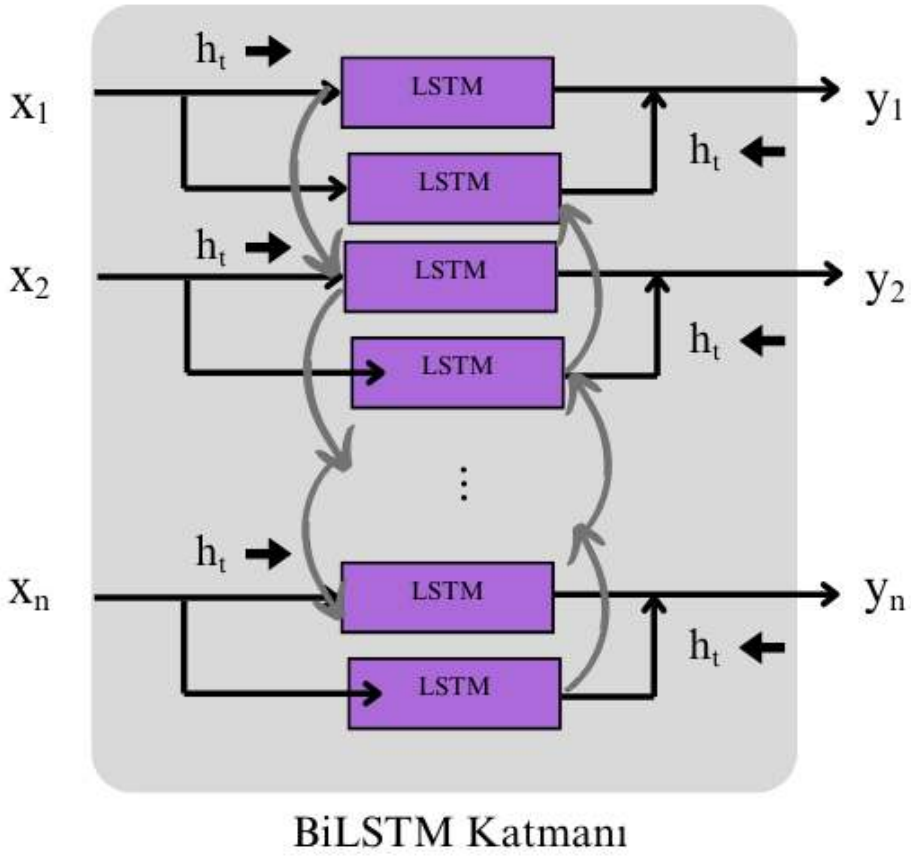
$\overrightarrow{h_t}$: Zaman adımı t 'de ileri yönlü LSTM tarafından üretilen gizli durum vektörüdür,

$\overleftarrow{h_t}$: Zaman adımı t 'de geri yönlü LSTM tarafından üretilen gizli durum vektörüdür,

$[\cdot;\cdot]$ ileri ve geri yönlü gizli durum vektörlerinin birleştirilmesini ifade etmektedir.

BiLSTM yapısının genel akışı ve işleyişi Şekil 2.7'de gösterilmiştir.

Şekilde de görüldüğü üzere, her bir giriş vektörü hem ileri hem de geri LSTM katmanlarında işlenmekte, sonuçta her zaman adımı için zenginleştirilmiş bir temsil *(h_t)* elde edilmektedir. Bu temsil, çıktı üretiminde *(y_t)* kullanılmaktadır.

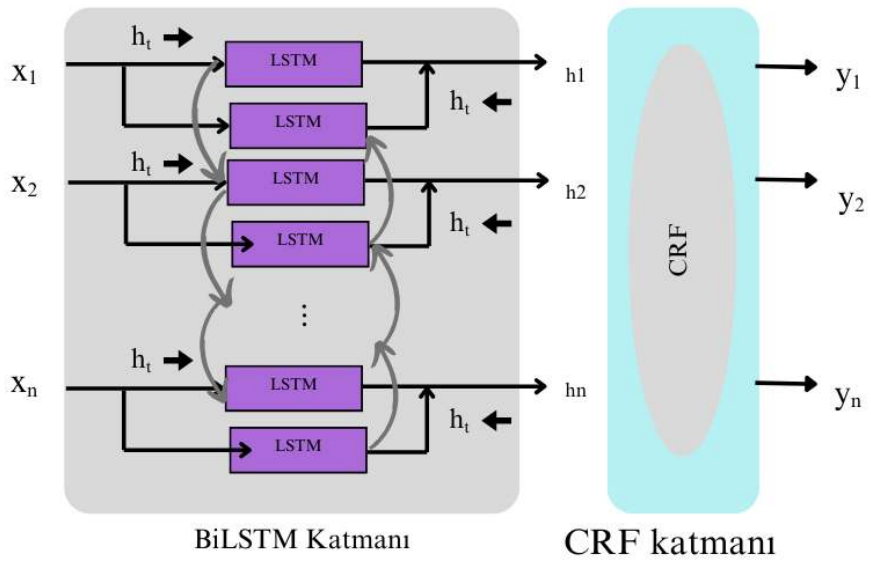


Şekil 2.7 BiLSTM mimarisinin genel yapısı.

BiLSTM yapıları, özellikle dil modelleme, duygu analizi, makine çevirisi ve NER gibi doğal dil işleme görevlerinde önemli avantajlar sağlamaktadır. Klinik metinler gibi uzun, karmaşık ve yoğun bilgi içeren verilerde, BiLSTM' in sunduğu çift yönlü bağlam modelleme yeteneği, varlık sınırlarının daha doğru belirlenmesi ve bağlamsal ilişkilerin daha etkili şekilde yakalanmasını mümkün kılmaktadır.

2.5.4 Koşullu Rastgele Alan (CRF)

NLP alanında, özellikle sıralı etiketleme problemlerinde, bireysel kelime sınıflandırmalarının ötesinde, etiketler arasındaki geçişlerin de doğru modellenmesi büyük önem taşımaktadır. Özellikle NER gibi görevlerde, kelimelerin anlamı sadece kendilerine değil, çevresindeki kelimelerle olan ilişkilerine ve bu kelimelere atanacak etiketlerin sıralı tutarlılığına da bağlıdır. Örneğin, bir varlık başlangıç etiketi (B-DISEASE) tespit edildiğinde, onu bir devam etiketi (I-DISEASE) izlemesi beklenir. Ancak klasik sınıflandırıcılar, kelime bağımsızında çalıştıkları için bu tür geçiş tutarlılıklarını doğal olarak modelleyemezler. Bu sorunu aşmak için BiLSTM gibi bağlamsal temsiller üreten modellerin çıkışlarına CRF katmanı entegre edilmektedir (Huang vd. 2015; Panchendrarajan ve Amaresan 2018,). Şekil 2.8'de BiLSTM-CRF yapısının genel mimarisi sunulmaktadır.



Şekil 2.8 BiLSTM-CRF Mimarisi.

CRF, gözlemlenen veri dizilerine karşılık gelen etiket dizilerini koşullu olasılık kullanarak modelleyen ayrıştırıcı (discriminative) bir modeldir (Lafferty vd. 2001). BiLSTM katmanı her bir zaman adımı için bir gizli durum *(h_t)* üretir; CRF katmanı ise bu gizli durumları kullanarak tüm sekans boyunca en olası etiket dizisini (y_1, y_2, ..., y_n) belirler. Böylece sadece bireysel kelime tahminlerine değil, aynı zamanda kelimeler arasındaki etiket geçişlerine de odaklanılır.

BiLSTM katmanında her giriş (x_t) ileri ve geri yönde işlenerek birleştirilmiş bir gizli durum (h_t) elde edilmekte; bu gizli durumlar daha sonra tek bir CRF katmanına aktarılmakta ve çıktı olarak sıralı bir etiket dizisi (y_t) üretilmektedir.

CRF katmanında, verilen bir gizli durum dizisi $(h= (h_1, h_2, ..., h_n))$ ve bir etiket dizisi $(y= (y_1, y_2, ..., y_n))$ için bir toplam skor fonksiyonu tanımlanır:

$$s(h, y) = \sum_{t=1}^{n} (A_{y_{t-1}, y_t} + P_{t, yt}) \tag{2.9}$$

Burada:

$A_{yt,\ yt}$: Önceki etiket y_{t-1}'den mevcut etiket y_t'ye geçişin skorudur.

$P_{t,yt}$: Zaman adımı t'deki gizli durumun, y_t etiketine karşı verdiği tahmin skorudur.

n: Dizideki zaman adımı sayısıdır.

Model, gözlemlenen veri dizisi için en yüksek toplam skoru üreten etiket dizisini bulmayı amaçlar:

$$\hat{y} = \underset{y}{argmax} s(h, y) \tag{2.10}$$

Burada;

$\hat{y}$: Modelin tahmin ettiği en olası etiket dizisidir.

$argmax$: Belirtilen fonksiyonun (burada $s(h,y)$), en yüksek değer aldığı y etiket dizisini bulma işlemidir.

y: Olası tüm etiket dizileri kümesini temsil eder.

s(h,y): Gizli durumlar *(h)* ve etiket dizisi *(y)* arasındaki toplam skor fonksiyonudur.

Bu yapı sayesinde, kelimeler arasındaki bağımlılıklar doğrudan model içine gömülerek, sekans genelinde tutarlı ve anlamlı etiketlemeler yapılabilmektedir. Klinik metinler gibi karmaşık ve uzun bağımlılık zincirleri içeren verilerde, BiLSTM-CRF mimarisi hem kelime düzeyindeki doğruluğu artırmakta hem de çıktı sekanslarının semantik uyumunu sağlamaktadır.

2.5.5 Transformer Mimarisi

NLP alanında sıralı veri modelleme amacıyla geliştirilen ilk sinir ağı mimarilerinden biri RNN olmuştur. Ancak RNN yapıları, uzun bağımlılık ilişkilerini öğrenmede karşılaştıkları gradyan sönmesi problemi nedeniyle sınırlı bir performans sergilemiştir (Elman 1990). Bu sorunu aşmak için geliştirilen LSTM mimarisi, daha uzun bağımlılıkları öğrenebilme kapasitesiyle dikkat çekmiş ve uzun yıllar boyunca sıralı modelleme görevlerinde standart yöntem olarak kullanılmıştır (Hochreiter ve Schmidhuber 1997).

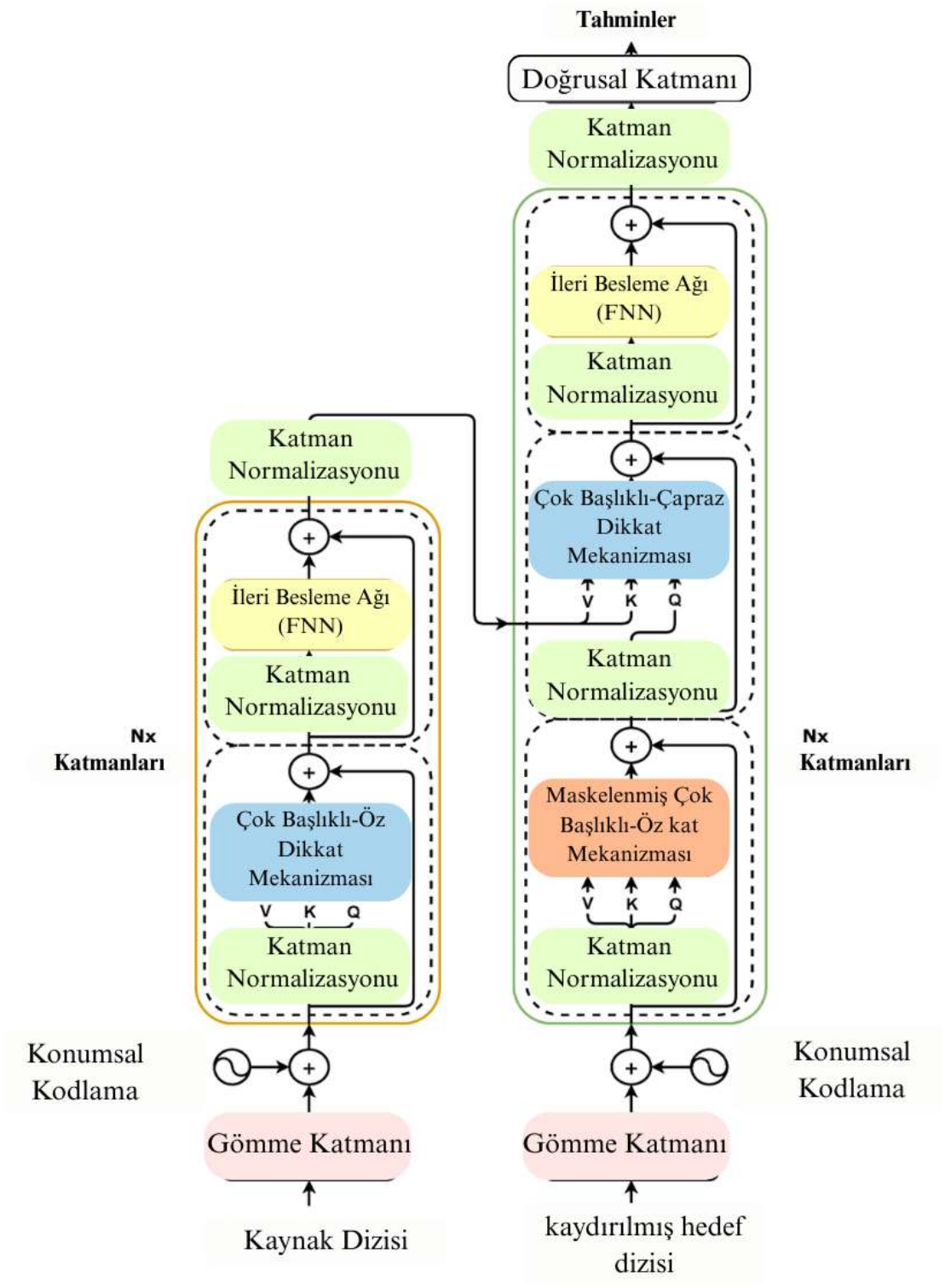
Buna rağmen, RNN ve LSTM gibi ardışık yapılar verilerin zaman adımlarına göre sıralı işlenmesi zorunluluğu taşıdığı için paralel veri işelmeye uygun değildir. Bu durum, özellikle büyük veri kümeleri üzerinde eğitim süresinin uzamasına ve uzun sekanslarda bilgi kaybı yaşanmasına neden olmuştur (Zhao vd. 2020).

Bu sınırlamaları aşmak amacıyla 2017 yılında Vaswani ve arkadaşları tarafından önerilen Transformer mimarisi, doğal dil işleme sistemlerinde yeni bir paradigma başlatmıştır (Vaswani vd. 2017). "Attention Is All You Need" başlıklı çalışmalarında, sıralı işlem gereksinimini tamamen ortadan kaldırarak sadece Attention mekanizması üzerine kurulu bir yapı tasarlamışlardır. Encoder-Decoder yapısı üzerinden giriş ve çıkış dizileri arasındaki ilişkileri etkili biçimde modellemekte ve tamamen paralel işlem yeteneği sayesinde büyük ölçekli veri kümeleri üzerinde verimli bir eğitim süreci sunmaktadır. Transformer mimari yapısı Şekil 2.9'da sunulmuştur.

Transformer mimarisi, doğal dil işleme ve sıralı veri modelleme alanlarında devrim niteliğinde değişiklikler yaratmıştır. Bu mimari, girdiler üzerindeki bağımlılıkları tamamen paralel şekilde modelleyebilme yeteneğine sahiptir. Böylece, önceki ardışık işleme zorunluluğu ortadan kaldırılarak uzun mesafe bağımlılıkları daha etkili biçimde öğrenilebilmekte ve eğitim süresi önemli ölçüde kısaltılmaktadır (Gillioz vd. 2020).

Transformer mimarisi "Çoklu Başlık Dikkat Mekanizması" (Multi-Head Attention) mekanizması kullanılarak farklı ilişki türlerinin paralel biçimde öğrenilmesini

sağlanmaktadır. Bunun yanı sıra, sıraya bağlılık bilgisinin modele dışsal olarak aktarılabilmesi amacıyla "Pozisyonel Kodlama" (Positional Encoding) tekniği geliştirilmiştir (Antipova ve Horban 2024).



Şekil 2.9 Transformer mimarisi: encoder ve decoder yapısı.

Şekil 2.9'da gösterildiği üzere, Transformer mimarisi iki temel bileşenden oluşmaktadır: Encoder ve Decoder. Giriş dizisi, kelime gömme (embedding) ve konumsal kodlama adımlarının ardından Encoder katmanlarında işlenerek bağlamsal temsillere dönüştürülür. Decoder katmanları ise bu temsilleri kullanarak çıktı dizilerini üretir. Encoder içinde çok başlıklı öz-dikkat mekanizmaları ve ileri beslemeli ağlar yer alırken; Decoder yapısında ek olarak maskelenmiş öz-dikkat mekanizması ve Encoder-Decoder arasında çapraz dikkat mekanizmaları bulunmaktadır (Ghojogh ve Ghodsi 2020).

2.5.5.1 Transformer Encoder-Decoder Yapısı

Transformer mimarisi, diziden diziye (sequence-to-sequence) öğrenme görevleri için geliştirilen bir encoder-decoder yapısına sahiptir. Bu yapıda, giriş dizisi encoder katmanları tarafından işlenerek bağlamsal temsillere dönüştürülmekte; ardından bu temsiller, Decoder katmanları aracılığıyla hedef çıktılar (örneğin başka bir metin dizisi veya sınıflandırma etiketleri) üretilmektedir (Moritz vd. 2020).

Encoder, birbiri ardına tekrarlanan *Nx* adet Encoder katmanından oluşmaktadır. Her Encoder katmanı üç ana bileşeni içerir (Nieminen 2023):

- Çok Başlıklı Öz-Dikkat Mekanizması (Multi-Headed Self-Attention): Giriş dizisindeki her token'ın, diğer tüm token'larla olan ilişkilerini paralel olarak modellemesini sağlar. Bu mekanizma, bağlamsal bilgi alışverişi sayesinde kelime temsillerinin zenginleşmesine olanak tanır.
- Katman Normalizasyonu (Layer Normalization): Öz-dikkat ve ileri beslemeli ağ bloklarının ardından uygulanır. Modelin öğrenmesini stabilize eder ve daha hızlı yakınsamayı sağlar.
- İleri Beslemeli Ağ (Feed-Forward Network-FFN): Her bir token'ın temsil vektörüne bağımsız olarak uygulanan iki katmanlı doğrusal dönüşüm ve aktivasyon fonksiyonundan oluşur.

Encoder girişinde, her kelime bir gömme vektörüne dönüştürülür ve pozisyonel kodlama eklenerek dizideki sıralı bilgiler modele kazandırılır. Encoder katmanlarından geçirilen temsiller, sonrasında Decoder'a aktarılır.

Decoder, encoder yapısına benzer şekilde *Nx* adet decoder katmanından oluşur, ancak birkaç önemli ek özellikle zenginleştirilmiştir. Her decoder katmanı üç temel dikkat mekanizması içerir (Zhang vd. 2021):

- Maskelenmiş Çok Başlıklı Öz-Dikkat Mekanizması (Masked Multi-Headed Self-Attention): Bir token'ın yalnızca kendisinden önce gelen token'lara dikkat edebilmesini sağlar. Bu mekanizma, sıralı çıktı üretimi için gereklidir ve gelecekteki bilgilerin yanlışlıkla kullanılmasını engeller.
- Çok Başlıklı Çapraz Dikkat Mekanizması (Multi-Headed Cross-Attention): Decoder, Encoder'dan gelen bağlamsal temsillere dikkat ederek giriş bilgisiyle çıktı üretimi arasındaki ilişkiyi kurar. Query vektörleri Decoder'dan, Key ve Value vektörleri Encoder çıktılarından gelir.
- Katman Normalizasyonu ve İleri Beslemeli Ağ (FFN): Her dikkat mekanizmasının ardından katman normalizasyonu yapılır ve ileri beslemeli ağ uygulanır.

Decoder sonunda, çıktı vektörleri bir doğrusal katmandan geçirilir ve softmax fonksiyonu uygulanarak her pozisyon için bir olasılık dağılımı üretilir. Bu dağılım sayesinde tahmini kelimeler veya etiketler seçilir.

2.5.5.2 Dikkat Mekanizmaları

Transformer mimarisinin temel taşı, Dikkat Mekanizması (Attention Mechanism) kavramıdır. Bu mekanizma, her bir giriş öğesinin dizideki diğer öğelere olan bağlamsal ilişkisini dinamik olarak modellemesine olanak tanır (Ghojogh ve Ghodsi 2020).

Özellikle Transformer'da kullanılan dikkat mekanizması iki ana bileşene ayrılır:

- Ölçeklenmiş nokta çarpımı dikkat mekanizması (scaled dot-product attention)
- Çok başlıklı dikkat mekanizması (Multi-head attention)

Ölçeklenmiş nokta çarpımı dikkat mekanizması (scaled dot-product attention) her bir sorgu (query) vektörünün, tüm anahtar (key) vektörleri ile karşılaştırılarak ilgili değer

(value) vektörlerinden bilgi çekmesine olanak tanır (Sohn vd. 2024). İşlem adımları şu şekildedir:

1. Sorgu (Q), Anahtar (K) ve Değer (V) Matrislerinin Oluşturulması: Giriş vektörleri, üç farklı ağırlık matrisi aracılığıyla Query, Key ve Value matrislerine dönüştürülür.
2. Uyum Skoru Hesaplama (Similarity Score): Her Query vektörü ile tüm Key vektörleri arasında nokta çarpımı alınır. Bu işlem, her öğenin diğer öğelerle olan benzerliğini ölçer.
3. Ölçekleme: Dikkat skorlarının büyüklüğünü stabilize etmek için her skor, Key vektör boyutunun karekköküne bölünür.
4. Softmax Uygulama: Skorlar softmax fonksiyonu ile normalize edilerek olasılık dağılımına dönüştürülür.
5. Ağırlıklı Toplama: Elde edilen ağırlıklar Value vektörleriyle çarpılarak çıktı vektörü hesaplanır.

Bu işlemler aşağıdaki denklemle özetlenebilir:

$$Attention\,(Q,K,V) = softmax\,(\frac{QK^T}{\sqrt{d_k}})V \tag{2.11}$$

Burada:

Q: Sorgu matrisi

K: Anahtar matrisi

V: Değer matrisi

d_k*:* Anahtar vektörünün boyutu

Ölçeklenmiş nokta çarpımı dikkat mekanizmasında, sorgu (Query) ve anahtar (Key) vektörleri arasındaki benzerlik skorları, bir olasılık dağılımına dönüştürülmelidir. Bu amaçla softmax fonksiyonu kullanılmaktadır (Gao vd. 2021).

Softmax fonksiyonu, giriş değerlerini normalize ederek toplamları 1 olan bir olasılık dağılımı üretir ve böylece modelin hangi öğelere daha fazla dikkat vereceğini belirlemesini sağlar. Softmax fonksiyonunun matematiksel ifadesi şu şekildedir:

$$softmax(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \tag{2.12}$$

Burada:

z_i: Giriş skorlarından biridir,

e: Euler sayısını (yaklaşık 2.718) temsil eder.

Bu süreç, yüksek skorların daha baskın hâle gelmesini sağlarken, düşük skorların etkisini azaltır. Böylece her bir Query vektörünün, ilgili Value vektörlerinden dikkatli ve kontrollü şekilde bilgi çekmesi mümkün olur

Transformer modelleri tek bir dikkat mekanizması yerine, birden fazla bağımsız dikkat başlığı (attention head) kullanır. Bu yapı, modelin farklı alt uzaylarda paralel olarak farklı ilişki türlerini öğrenmesini sağlar (Mahdavi 2023). İşlem adımları şu şekildedir:

1. Çoklu Başlıklar için Ayrı Projeksiyonlar: Giriş vektörleri farklı ağırlık matrisleri ile birçok farklı Q_i, K_i, V_i alt uzayına projekte edilir.
2. Paralel Scaled Dot-Product Attention Uygulaması: Her başlık kendi scaled attention skorlarını hesaplar.
3. Başlıkların Birleştirilmesi: Tüm başlıklardan elde edilen çıktı vektörleri birleştirilir.
4. Son Doğrusal Dönüşüm: Birleştirilen vektörler bir çıkış ağırlık matrisiyle (W^O) doğrusal dönüşüme tabi tutulur.

Bu işlemler aşağıdaki genel formülle ifade edilir:

$$Multihead\ (Q, K, V) = Concat(head_1, head_2, \ldots . . head_n)W^O \tag{2.13}$$

$$where\ head_i = Attention\ (QW_i^Q, KW_i^K, VW_i^V) \tag{2.14}$$

Burada:

W_i^Q, W_i^K, W_i^V : Her başlık için farklı ağırlık matrisleri,

W^O: Başlıkların çıktısını birleştiren ağırlık matrisi,

h: Başlık (head) sayısıdır.

2.5.5.3 İleri Beslemeli Ağlar (FFN)

Transformer mimarisinde yer alan İleri Beslemeli Ağ (Feed-Forward Network, FFN) bileşeni, her encoder ve decoder katmanında dikkat mekanizmasının ardından gelen, pozisyona bağımlı olmayan doğrusal dönüşümlerden oluşan temel bir modüldür. Bu yapı, her bir giriş vektörünü sırasıyla iki doğrusal katmandan geçirerek hem temsil gücünü artırmakta hem de doğrusal olmayanlık ekleyerek daha karmaşık özelliklerin öğrenilebilmesini sağlamaktadır (Sonkar ve Baraniuk 2023).

Transformer'daki FFN modülü, tüm giriş dizisi boyunca her pozisyona ayrı ve bağımsız olarak uygulanır. Bu, modelin paralel işlem kapasitesini koruyarak aynı anda birden fazla konumda işlem yapılmasına olanak tanır. FFN mimarisi, tipik olarak şu şekilde formüle edilir (Alammar 2021):

$$FFN(x) = \max(O, xW_1 + b_1)\, W_2 + b_2 \tag{2.15}$$

Burada:

x: Giriş vektörüdür.
W_1, W_2: Ağırlık matrisleridir.
b_1, b_2: Bias terimleridir.
max (0, ·): ReLU (Rectified Linear Unit) aktivasyon fonksiyonudur.

ReLU aktivasyon fonksiyonu, her bir bileşeni için negatif değerleri sıfıra eşitleyip pozitif değerleri olduğu gibi geçirerek, modele doğrusal olmayanlık kazandırır. ReLU fonksiyonu aşağıdaki gibi tanımlanır (Hinz vd. 2019):

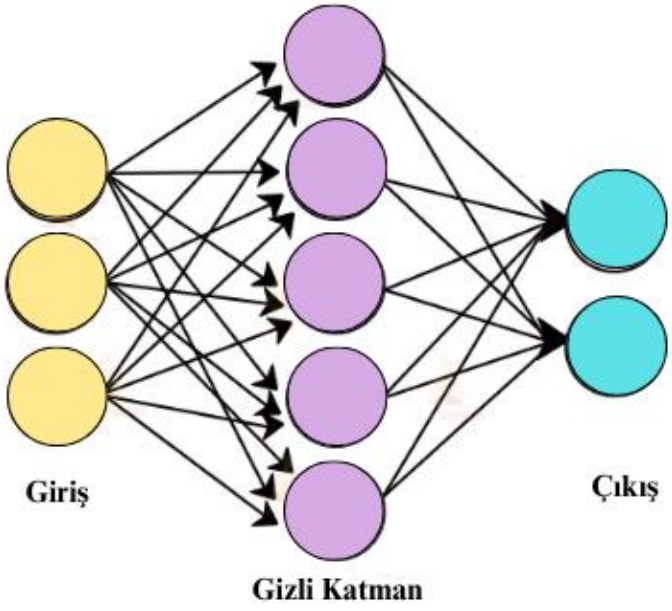
$$ReLU(z) = \max(O, z) \tag{2.16}$$

Bu doğrusal olmayanlık, ağın çok katmanlı yapısında parçalı lineer dönüşümler uygulayarak daha karmaşık karar yüzeyleri oluşturmasını sağlar. ReLU'nun sade yapısı, eğitim sırasında gradyan geçişinin korunmasına, dolayısıyla daha hızlı ve kararlı öğrenmeye katkı sağlar.

Transformer’da kullanılan FFN katmanı genellikle giriş boyutunun dört katı kadar genişletilen bir gizli katman içerir. Yani örneğin giriş boyutu 512 ise, gizli katman 2048 nörondan oluşur ve ardından tekrar 512 boyutuna projeksiyon yapılır. Bu yapı, hem parametre sayısını yönetilebilir seviyede tutmakta hem de temsillerin kapasitesini artırmaktadır.

Şekil 2.10’da gösterildiği üzere, FFN yapısı bir giriş katmanı (sarı), genişletilmiş tek bir gizli katman (mor) ve çıkış katmanından (mavi) oluşmaktadır. Bu yapı, Transformer mimarisindeki FFN modülünün mantıksal temsili olarak kullanılabilir.

FFN'nin her bir pozisyona bağımsız olarak uygulanıyor olması, Transformer’ın paralel bilgi işleme yeteneğinin korunmasını sağlar. Bu yapı, özellikle uzun diziler üzerinde hesaplamaların GPU/TPU gibi paralel işlemcilerde verimli yürütülmesine olanak tanıdığı için, mimarinin yüksek performanslı hale gelmesinde kritik rol oynamaktadır.



Şekil 2.10 Transformer mimarisinde kullanılan tipik bir FFN bileşeninin yapısı.

2.5.5.4 Gömme Katmanı ve Pozisyonel Kodlama

NLP uygulamalarında kullanılan derin öğrenme modelleri, metin girdilerini sayısal temsillere dönüştürmeden işleyemez. Bu dönüşüm, Gömme Katmanı (Embedding

Layer) aracılığıyla gerçekleştirilir. Gömme katmanı, her kelimeyi sabit boyutlu yoğun (dense) vektörlere çevirerek, kelimeler arasındaki anlamsal ilişkilerin sayısal olarak temsil edilmesini sağlar. Bu temsiller, önceden eğitilmiş gömme matrisleri (Word2Vec, GloVe) ya da eğitim sırasında öğrenilen parametreler ile elde edilebilir (Patil vd. 2023).

Ancak Transformer mimarisi, sıralı yapıları (RNN, LSTM vb.) kullanmadığı için, yalnızca kelimenin içeriği değil, sıralamadaki konumu da açıkça modele sunulmalıdır. Çünkü bir cümlede kelimelerin konumu, bağlamı anlamlandırmada kritik öneme sahiptir. Bu gereksinimden ötürü, Transformer mimarisinde her kelimeye ait gömme vektörüne, onun pozisyon bilgisini içeren bir Pozisyonel Kodlama Vektörü eklenir (Chen vd. 2021). Her bir kelime, önce Gömme Modeli tarafından vektörleştirilir. Aynı zamanda her kelimenin konumsal bilgisi, Pozisyonel Kodlama Modeli aracılığıyla hesaplanır. Elde edilen bu iki vektör kelimenin anlam vektörü ve konum vektörü element bazlı (bileşen bazlı) toplama işlemiyle birleştirilerek, modele beslenir (Kaliyar 2020).

Transformer mimarisinde önerilen pozisyonel kodlama yöntemi, öğrenilebilir parametreler içermeyen, sabit ve matematiksel olarak tanımlı bir yapıdır. Sinüs ve kosinüs fonksiyonlarına dayalı bu yapının formülleri aşağıda verilmiştir (Huangliang vd. 2023):

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right) \tag{2.17}$$

$$PE_{(pos,2i+1)} = cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \tag{2.18}$$

Burada:

pos: Kelimenin dizideki pozisyonudur.

i: Pozisyonel vektörün boyutundaki indeksidir.

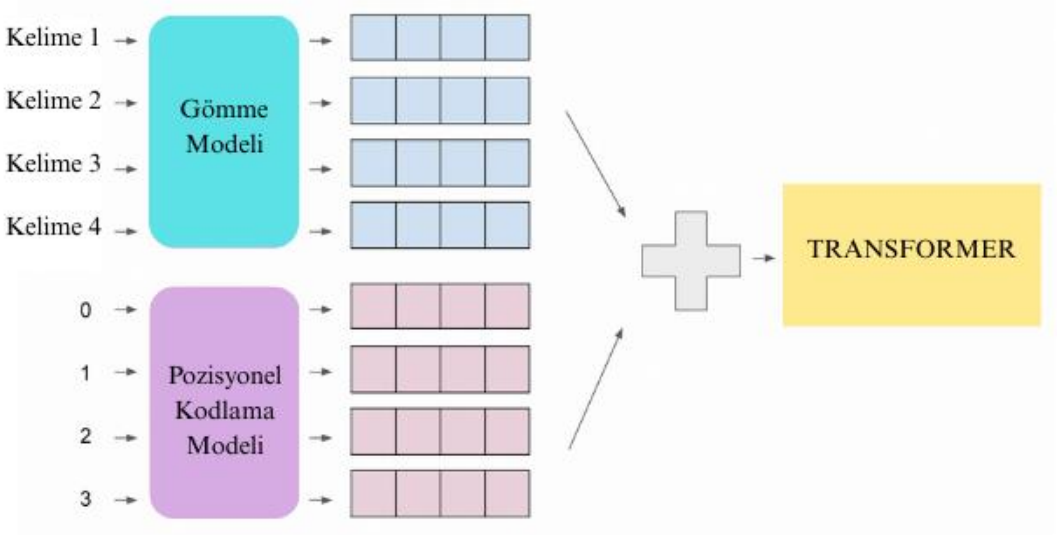
d_{model}: Gömme vektörünün boyutudur.

Sinüs ve *kosinüs* fonksiyonları ise farklı frekanslarda konumsal ayrım sağlamaktadır.

Bu yöntem sayesinde model, yalnızca mutlak konumları değil, aynı zamanda kelimeler arası göreli mesafeleri de öğrenebilir hale gelir. Böylece, dikkat mekanizması her kelimeye hem semantik içerik hem de konumsal bağlam temelinde işlem uygulayabilir.

Şekil 2.11'de gösterildiği üzere, her kelime dizisi ilk olarak Gömme Katmanı aracılığıyla sabit boyutta vektör temsiline dönüştürülür. Bu katman, kelimelerin anlamsal özelliklerini içeren yoğun vektör gösterimlerini üretir. Böylece örneğin "Kelime 1" ve "Kelime 2" gibi farklı girdiler, kendi anlam uzaylarında benzersiz konumlara yerleştirilmiş olur.

Ancak bu gömme temsilleri, kelimenin cümledeki sırasını içermediğinden, modelin sıralama bilgisine doğrudan erişimi yoktur. Bu eksikliği gidermek üzere, her kelimenin cümle içindeki konumuna ilişkin bir Pozisyonel Kodlama Vektörü üretilir. Şemada "Pozisyonel Kodlama Modeli" ile ifade edilen bu yapı, her pozisyon için sabit matematiksel fonksiyonlar (sinüs ve kosinüs) kullanarak benzersiz kodlamalar oluşturur.



Şekil 2.11 Gömme ve Pozisyonel Kodlama Temsili.

Gömme vektörleri ve pozisyonel kodlama vektörleri aynı boyutta olacak şekilde toplanır. Bu işlem sonucunda elde edilen bileşik vektör hem kelimenin anlamını hem de pozisyonunu temsil eder. Böylelikle Transformer mimarisi, giriş dizisindeki kelimelerin hem bağlamsal içeriğini hem de sıralama bilgisini birlikte işler.

Bu entegrasyon, modelin uzun mesafeli bağımlılıkları tanımasına olanak tanır ve dikkat mekanizmalarıyla birlikte, kelime konumlarının da dikkate alındığı güçlü bağlam

temsilleri oluşturur. Sonuç olarak, pozisyonel kodlama Transformer'ın bağlamsal duyarlılığını artıran temel bileşenlerden biri hâline gelir.

2.5.5.5 Katman Normalizasyonu ve Artık Bağlantılar

Transformer mimarisinde, her encoder ve decoder katmanı, dikkat ve FFN bileşenlerinden oluşmakla birlikte bu bileşenlerin ardından iki önemli yapı daha yer alır: Artık Bağlantılar (Residual Connections) ve Katman Normalizasyonu (Layer Normalization). Bu iki mekanizma, modelin derinliği arttıkça karşılaşılan optimizasyon sorunlarını hafifletmek ve eğitimi kararlı hâle getirmek amacıyla uygulanır (Vaswani vd. 2017).

Artık bağlantı, giriş ile çıkış arasına doğrudan bir geçiş yolu eklenmesini ifade eder. Bu yaklaşım, derin sinir ağlarında karşılaşılan gradyan sönmesi problemini azaltmakta ve öğrenmenin daha hızlı ve istikrarlı gerçekleşmesini sağlamaktadır (Kobayashi vd. 2021). Matematiksel olarak şu şekilde ifade edilir:

$$y = Sublayer(x) + x \tag{2.19}$$

Burada:

x: Giriş vektörüdür.
Sublayer(x): Dikkat mekanizması veya FFN gibi alt katmanların çıktısıdır.
y: Artık bağlantıdan sonra elde edilen ara çıktıdır.

Bu yapı sayesinde model, öğrenilmesi zor olan dönüşümleri sıfırdan başlatmak yerine, mevcut bilgiyi olduğu gibi aktarabilir ve yalnızca gerekli dönüşümleri öğrenmeye odaklanabilir.

Artık bağlantıyla elde edilen y vektörü, daha sonra Katman Normalizasyonu işlemine tabi tutulur. Bu teknik, her eğitim örneği için özellik bazlı olarak ortalama ve standart sapma ile normalize edilen bir dönüşümdür. Bu işlem, modelin daha hızlı öğrenmesini sağlar ve içsel kovaryans kaymasını azaltır (Ba vd. 2016). İşlem şu şekilde formüle edilir:

$$LayerNorm(y) = \gamma . \frac{y - \mu}{\sigma} + \beta \quad (2.20)$$

Burada:

μ: y'nin ortalamasıdır.

σ: y'nin standart sapmasıdır.

γ *ve* β: Öğrenilebilir parametrelerdir.

Katman normalizasyonu, her pozisyondaki vektöre bağımsız olarak uygulanır ve modelin farklı örnekler üzerinde daha istikrarlı şekilde davranmasını sağlar. Transformer mimarisinde bu işlem hem encoder hem de decoder bileşenlerinde, dikkat ve FFN modüllerinden sonra uygulanmaktadır (Xiong vd. 2020).

2.5.6 Transformer Mimarisi Tabanlı Modeller

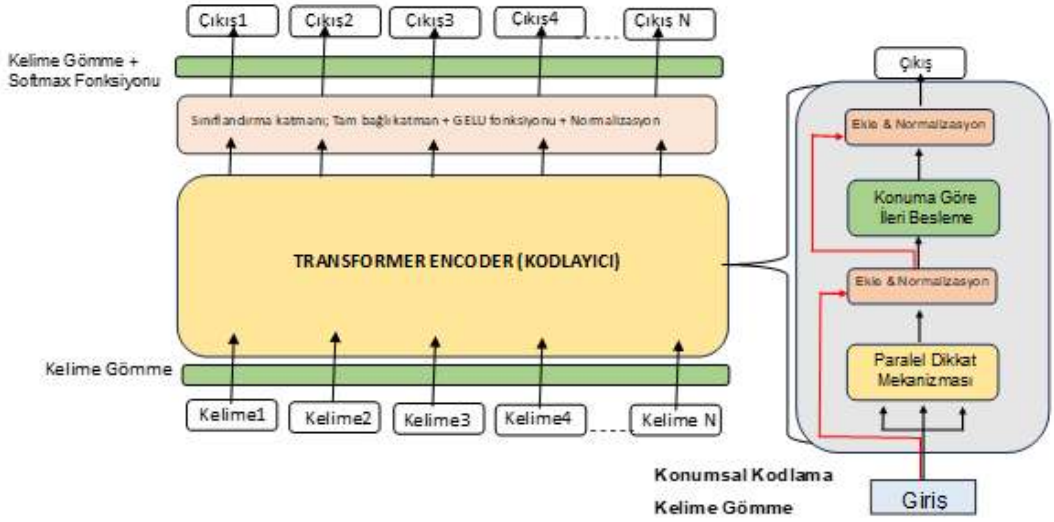
Transformer tabanlı modeller, doğal dil işleme alanında özellikle NER görevlerinde bağlamsal öğrenme yetenekleri sayesinde büyük ilerlemeler sağlamıştır. Bu başarı, klinik metin madenciliği gibi karmaşık semantik analiz gerektiren alanlarda da etkili bir şekilde kullanımlarını yaygınlaştırmıştır. Bu bölümde, Transformer mimarisi temelli klinik NER modelleri detaylı biçimde ele alınacaktır.

2.5.6.1 BERT

BERT (Bidirectional Encoder Representations from Transformers), Devlin ve arkadaşları tarafından geliştirilen ve yalnızca encoder katmanlarını temel alan bir Transformer modelidir. Bu mimari, klasik tek yönlü dil modellerinin aksine, bağlamı çift yönlü olarak dikkate alarak her bir kelimenin temsilini hem solundaki hem de sağındaki kelimelere dayanarak öğrenmektedir. BERT'in bu iki yönlü yapısı, özellikle bağlama duyarlı doğal dil işleme görevlerinde dikkate değer bir performans artışı sağlamıştır (Devlin vd. 2019).

Modelin girişine, önceden belirlenmiş sabit uzunlukta bir kelime dizisi verilir. Her bir kelime, kelime gömme, segment kodlaması ve pozisyonel kodlama bileşenlerinin

toplamı ile vektörleştirilerek modele aktarılır. Bu temsiller, Şekil 2.12'de gösterildiği üzere, ilk olarak gömme katmanından geçirilmekte, ardından pozisyonel kodlama ile zenginleştirilerek Transformer encoder katmanlarına iletilmektedir. Encoder katmanları içerisinde çok başlıklı öz-dikkat mekanizması yer almakta ve bu yapı, paralel dikkat başlıkları aracılığıyla her kelimenin bağlamına uygun temsiller üretmektedir. Dikkat bloklarının ardından ise FFN uygulanmakta, bu sayede modelin doğrusal olmayan ilişkileri öğrenebilme kapasitesi artırılmaktadır. Tüm bu işlemler sırasında katman normalizasyonu ve rezidüel bağlantılarla modelin eğitimi kararlı hale getirilir (Kenton vd. 2019).



Şekil 2.12 BERT mimarisi.

Modelin çıktısı, özel olarak eklenen [CLS] token'ına karşılık gelen temsil vektörüdür. Bu çıktı, üst katmanda yer alan sınıflandırma bloğuna aktarılır; burada tam bağlantı (fully connected) katmanı, GELU aktivasyon fonksiyonu ve normalizasyon uygulanarak nihai tahmin gerçekleştirilir. Son aşamada, softmax fonksiyonu aracılığıyla olasılık dağılımı elde edilir. Eğitim sürecinde ise model iki görevle önceden eğitilmiştir: rastgele maskelenmiş kelimeleri tahmin etmeyi amaçlayan Masked Language Modeling (MLM) ve ardışık iki cümle arasındaki ilişkiyi öğrenmeyi hedefleyen Next Sentence Prediction (NSP) görevleri (Devlin vd., 2019).

Görsel olarak sunulan mimaride, veri ön işleme adımından itibaren kelimelerin gömme temsilleri oluşturulmakta, ardından konumsal bilgi eklenerek Transformer encoder katmanına iletilmektedir. Burada uygulanan çok başlıklı dikkat ve ileri besleme mekanizmaları, son aşamada sınıflandırma katmanıyla birleştirilerek modelin belirli NLP görevlerini yüksek doğrulukla yerine getirmesi sağlanmaktadır. Bu özellikleri sayesinde BERT, klinik metin madenciliği de dâhil olmak üzere birçok doğal dil işleme görevinde temel yapı taşı hâline gelmiştir. BERT modelinin temel yapı parametreleri Çizelge 2.2'te özetlenmiştir (Kaliyar 2020).

Çizelge 2.2'te sunulan yapısal veriler, BERT modelinin hem derinlik hem de parametre yoğunluğu açısından oldukça güçlü bir mimariye sahip olduğunu göstermektedir. Özellikle 12 katmandan ve her katmanda 12 dikkat başlığından oluşan çok başlıklı dikkat mekanizması, modelin giriş metinlerindeki bağlamsal ilişkileri zengin bir şekilde öğrenebilmesine olanak sağlamaktadır. 768 boyutlu gizli temsiller, kelimelerin hem anlamsal hem de bağlamsal özelliklerini yüksek doğrulukla yakalayacak şekilde optimize edilmiştir.

Çizelge 2.2 Bert modeli temel özellikleri.

Özellik	**Değer**
Encoder Katman Sayısı	12
Başlık (Head) Sayısı	12
Gizli Katman Boyutu	768
Toplam Parametre Sayısı	Yaklaşık 110 milyon
Maksimum Girdi Uzunluğu	512 token

Yaklaşık 110 milyon parametre içeren bu yapı, hem eğitim verilerinden kapsamlı genellemeler yapabilmekte hem de transfer öğrenme senaryolarında geniş bir görev yelpazesinde etkili bir başlangıç noktası sunmaktadır. Maksimum 512 token'lık giriş sınırı ise, uzun cümle ve paragrafları işleyebilme kapasitesini artırarak, modelin farklı doğal dil işleme görevlerinde esnek kullanımını mümkün kılmaktadır (Mohammed vd. 2021).

2.5.6.2 BioBERT

BERT mimarisi, bağlamsal dil modeli olarak doğal dil işleme alanında devrim niteliğinde bir gelişme sunmuş olsa da tıbbi terminolojiye özgü karmaşık yapıları yeterince temsil edememektedir. Bu nedenle, klinik ve biyomedikal metinlerde daha yüksek doğruluk elde etmek amacıyla BioBERT (Biomedical Bert) modeli geliştirilmiştir. BioBERT, temel BERT mimarisine dayalı olup, biyomedikal alana özgü büyük veri kümeleri üzerinde yeniden ön eğitimden geçirilmiş bir dil modelidir (Lee vd. 2020).

BioBERT'in geliştirilmesindeki temel amaç, PubMed ve PMC gibi biyomedikal metin kaynaklarından elde edilen literatürle modeli daha derin ve bağlama duyarlı hale getirmektir. Bu doğrultuda, BioBERT modeli; önce BERT-base mimarisi kullanılarak genel dil modellemesi ile önceden eğitilmiş, ardından PubMed ve PMC makaleleri üzerinde alan-adaptasyonu gerçekleştirilmiştir. Bu sayede model, genel dilden farklı olarak, semantik açıdan daha karmaşık ve bağlamsal farklılıklar içeren tıbbi kavramları daha iyi öğrenebilmektedir (Khadhraoui vd. 2022).

Yapısal olarak BioBERT, BERT-base ile aynı mimariyi korumaktadır: 12 adet encoder katmanı, her encoder'da 12 başlıktan oluşan çoklu dikkat mekanizması ve 768 boyutlu gizli katmanlardan meydana gelmektedir. Modelin toplam parametre sayısı yaklaşık 110 milyon civarındadır. Eğitimi, MLM ve NSP görevleriyle gerçekleştirilmiş; dolayısıyla temel öğrenme görevlerinde herhangi bir değişiklik yapılmamıştır (Srivastava vd. 2021).

BioBERT, özellikle NER, ilişki çıkarımı ve biyomedikal soru-cevap sistemleri gibi klinik doğal dil işleme görevlerinde literatürde dikkate değer başarılara imza atmıştır (Ruas vd. 2020, Yu vd. 2019). Bu yönüyle BioBERT, tıbbi metinlerdeki varlıkların doğru şekilde tanımlanması ve bağlamsal olarak işlenmesi açısından öncü modellerden biri hâline gelmiştir. Ayrıca, farklı klinik uygulamalara entegre edilerek karar destek sistemlerinin etkinliğini artırma potansiyeline sahiptir. BioBert temel özellikleri Çizelge 2.3'te sunulmuştur.

Çizelge 2.3 BioBERT modelinin temel özellikleri.

Özellik	Değer
Encoder Katman Sayısı	12
Başlık (Head) Sayısı	12
Gizli Katman Boyutu	768
Toplam Parametre Sayısı	Yaklaşık 110 milyon
Ön Eğitim Verisi	PubMed, PMC
Maksimum Girdi Uzunluğu	512 token

2.5.6.3 ClinicalBERT

ClinicalBERT, genel amaçlı BERT mimarisinin sağlık alanına özgü klinik metinlerle yeniden eğitilmiş özel bir versiyonudur. Modelin geliştirilmesindeki temel motivasyon, özellikle ESK gibi serbest biçimli klinik metinlerde karşılaşılan özel terimlerin, kısaltmaların ve stilistik farklılıkların genel dil modelleri tarafından yeterince temsil edilememesidir (Huang vd. 2019). ClinicalBERT modelinin temel özellikleri Çizelge 2.4'te verilmiştir.

Çizelge 2.4 ClinicalBERT modelinin temel özellikleri.

Özellik	Değer
Encoder Katman Sayısı	12
Başlık (Head) Sayısı	12
Gizli Katman Boyutu	768
Toplam Parametre Sayısı	Yaklaşık 110 milyon
Ön Eğitim Verisi	MIMIC-III Klinik Notları
Maksimum Girdi Uzunluğu	512

ClinicalBERT modeli, BioBERT'ten farklı olarak yalnızca bilimsel yayınlara değil, doğrudan MIMIC-III veri kümesi (Johnson vd. 2016) gibi klinik ortamlarda üretilmiş hasta notlarına dayalı olarak ön eğitimden geçirilmiştir. Bu özellik, modeli doğrudan klinik bağlamda üretilen metinlerin işlenmesi açısından daha yetkin hâle getirmektedir.

Mimari olarak ClinicalBERT, BERT-base yapısını temel almaktadır. 12 encoder katmanı, her biri 12 çoklu başlıktan oluşan dikkat mekanizması ve 768 boyutlu gizli katman boyutuna sahiptir. Toplamda yaklaşık 110 milyon parametre içermektedir. Modelin eğitimi sırasında MLM ve NSP görevleri kullanılmış, böylece hem kelime düzeyinde hem de cümleler arası bağlam modelleme mümkün kılınmıştır (Franz vd. 2020).

2.6 Derin Öğrenme Tabanlı Modellerin Eğitim Parametreleri

Derin öğrenme modellerinin başarımı, büyük ölçüde doğru şekilde belirlenmiş eğitim parametrelerine bağlıdır (Shrestha ve Mahmood 2019). Bu parametreler, modelin öğrenme sürecinin temel yapı taşlarını oluşturarak, hem eğitim verisinden genelleme yeteneği kazanmasını hem de aşırı öğrenmeden kaçınmasını sağlar. Bu bölümde, çalışmada kullanılan hibrit modellerin eğitiminde dikkate alınan temel hiper-parametreler detaylı biçimde tanıtılacaktır.

Ele alınan başlıca parametreler arasında öğrenme oranı (learning rate), maksimum giriş uzunluğu (max input length), yığın boyutu (batch size) ve epok sayısı (epoch) yer almaktadır. Her bir parametrenin modelin eğitim sürecine olan etkisi, seçim gerekçeleri ve parametrik optimizasyon süreçleri kapsamlı olarak ele alınacaktır.

Hiper-parametrelerin doğruluk oranları yanında eğitim süresi, genelleme kapasitesi ve etiketleme tutarlılığını da doğrudan etkilediği dikkate alınarak değerlendirmeler yapılmıştır. Bu bölümde sunulan bilgiler, ilerleyen bölümlerde yer alan optimizasyon stratejilerinin ve başarı ölçümlerinin temelini oluşturması bakımından önem arz etmektedir.

2.6.1 Öğrenme Oranı

Derin öğrenme tabanlı doğal dil işleme modellerinde öğrenme oranı, modelin ağırlıklarını her yinelemede ne kadar değiştireceğini belirleyen temel bir hiper-parametredir ve optimizasyon sürecinin başarısını doğrudan etkiler. Bu parametre, gradyan inişi tabanlı optimizasyon algoritmalarının her adımda hedef fonksiyonun

eğimine göre ağırlıkları ne kadar güncelleyeceğini tanımlar. Çok düşük öğrenme oranları modelin yavaş yakınsamasına ve yerel minimumlara takılmasına neden olabilir. Aşırı yüksek öğrenme oranları ise ağırlıklarda kararsızlığa yol açarak öğrenmeyi engelleyebilir (Jacobs 1998, Behara vd. 2006).

Özellikle önceden eğitilmiş Transformer modelleri (örneğin BERT, BioBERT, ClinicalBERT) ile yapılan transfer öğrenme uygulamalarında, öğrenme oranının daha dikkatli seçilmesi gerekmektedir. Bu modeller genellikle büyük miktarda genel veya alan özgü verilerle eğitildiğinden, düşük bir öğrenme oranı kullanılarak modelin önceki bilgilerinin bozulmadan yeni göreve uyum sağlaması hedeflenir. Bu nedenle literatürde $1x10^{-5}$ ila $5x10^{-5}$ aralığında sabit veya yavaş azalan öğrenme oranları önerilmektedir (Devlin vd. 2019, Lee vd. 2020). Ayrıca, öğrenme oranı zamanla azalan programlar (learning rate schedules) da modelin eğitim süreci boyunca daha kararlı bir şekilde yakınsamasına katkı sağlamaktadır (Smith 2017).

2.6.2 Epok

Epok, derin öğrenme modellerinin eğitiminde kullanılan temel bir kavram olup, modelin eğitim veri setindeki tüm örnekleri bir kez görmesini ifade eder. Başka bir deyişle, modelin tüm eğitim örnekleri üzerinde bir kez ileri (forward) ve geri (backward) geçiş yapmasıyla tamamlanan her döngü bir epok olarak tanımlanır. Ancak modelin veriyi yalnızca bir epok boyunca görmesi çoğu zaman yeterli olmadığından, eğitim süreci genellikle çok sayıda epok'tan oluşur (Alzubaidi vd. 2021).

Derin öğrenme modellerinde epok sayısının seçimi, modelin genel başarımı, yakınsama süresi, aşırı öğrenme (overfitting) ve az öğrenme (underfitting) gibi durumlar üzerinde doğrudan etkili bir hiper-parametredir. Yetersiz epok sayısı, modelin eğitim verilerinden yeterli düzeyde öğrenememesine yol açarken; aşırı epok sayısı, modelin eğitim verilerine fazla uyum sağlamasıyla test verilerinde genelleme yeteneğinin azalmasına neden olabilir (Komatsuzaki 2019).
Doğal dil işleme görevlerinde, özellikle tıbbi metin madenciliği gibi yüksek bağlamsal karmaşıklığa sahip alanlarda, optimal epok sayısı genellikle model mimarisi, veri

setinin büyüklüğü, etiket dengesi ve kullanılan optimızasyon algortimasına bağlı olarak değişkenlik göstermektedir. Literatürde BERT tabanlı modeller için 3–5 epok başlangıç önerisi sunulmakla birlikte, BiLSTM-CRF gibi daha derin yapılarla entegre edilen modellerde 10'un üzerindeki epok sayıları daha kararlı sonuçlar vermektedir (Devlin vd. 2019, Wang vd. 2020).

2.6.3 Yığın Boyutu

Yığın boyutu, derin öğrenme modellerinin eğitim sürecinde aynı anda işlenen örnek sayısını ifade eden temel bir hiper-parametredir. Eğitim sırasında model, tüm veriyi aynı anda değil, küçük parçalar (mini-batch) hâlinde işler. Bu parçalar, her bir ileri besleme ve geri yayılım adımında modelin ağırlıklarını güncellemek için kullanılır. Bu nedenle yığın boyutu, eğitim sürecinin hem hesaplama verimliliğini hem de modelin öğrenme dinamiklerini doğrudan etkiler (Smith vd. 2017).

Yığın boyutunun seçimi, öğrenme sürecinde optimizasyon dinamikleri, bellek kullanımı, genelleme performansı ve yakınsama hızı gibi kritik faktörlerde belirleyici rol oynar. Küçük yığın boyutları (8, 16, 32), modelin ağırlıklarını daha sık güncellemesine olanak tanır. Bu, daha fazla gürültü içeren gradyan tahminleriyle öğrenmenin daha rastlantısal ilerlemesine neden olabilir. Bu durum genelleme açısından faydalı olabilir çünkü modelin yerel minimumlara sıkışması engellenebilir (Keskar vd. 2017). Ancak bu gürültü, öğrenme sürecinde kararsızlığa ve daha uzun eğitim süresine yol açabilir. Büyük yığın boyutları (64, 128 veya üzeri), her adımda daha doğru ve kararlı gradyan tahminleri sağlar. Bu durum modelin daha hızlı ve istikrarlı bir şekilde yakınsamasına katkıda bulunur. Ancak büyük yığınlar yüksek GPU belleği gerektirir ve bazen modelin aşırı ezberleme yapmasına veya genelleme kapasitesinin azalmasına yol açabilir (Masters ve Luschi 2018).

Literatürde yapılan çalışmalar, küçük yığın boyutlarının özellikle dengesiz veri kümelerinde ve karmaşık metinlerde daha iyi genelleme sağladığını ortaya koymuştur (Smith vd. 2017, Shwartz vd. 2023). Ayrıca, öğrenme oranı ile yığın boyutu arasında önemli bir etkileşim vardır. Genellikle büyük yığın boyutları daha yüksek öğrenme

oranlarıyla eşleştirilir.

2.6.4 Maksimum Giriş Uzunluğu

Maksimum giriş uzunluğu, bir doğal dil işleme modeline aynı anda sunulabilecek en uzun token dizisini (yani kelime veya alt kelime birimlerini) belirleyen sınırlayıcı bir hiper-parametredir. Transformer tabanlı dil modelleri sabit uzunluklu girişleri işler. Bu nedenle, bir metin parçası modelin kabul ettiği maksimum uzunluğu aştığında, bu metin ya kesilir ya da özel işlem teknikleriyle bölünerek modele sunulur. Maksimum giriş uzunluğu parametresi modelin şu yönlerini doğrudan etkiler:

- Bağlamsal Kapsam (Context Window): Uzunluk arttıkça model, cümleler arası ilişkileri ve uzun mesafeli bağımlılıkları daha iyi yakalayabilir. Özellikle klinik metinler gibi karmaşık yapılar içeren metinlerde, uzunluk sınırının yüksek tutulması, bağlamsal bütünlüğün korunmasına katkı sağlar (Devlin vd. 2019).
- Hesaplama Maliyeti: Transformer modelleri, giriş uzunluğunun karesiyle orantılı hesaplama maliyetine sahiptir. Bu nedenle uzunluk arttıkça bellek tüketimi ve işlem süresi de artar. Örneğin, giriş uzunluğunu 128'den 256'ya çıkarmak, bellek kullanımını yaklaşık dört kat artırabilir (Vaswani vd. 2017).
- Veri Kırpma veya Dolgulama (Truncation & Padding): Maksimum uzunluk değeri, kısa metinlerin padding ile uzatılmasına; uzun metinlerin ise truncation ile kesilmesine neden olur. Dolayısıyla bu değer çok küçükse önemli bağlamsal bilgiler kaybedilebilir; çok büyükse ise gereksiz boşluklar modele ek yük bindirir. Bu durum özellikle düşük veri hacmine sahip veya dengesiz etiket dağılımı olan tıbbi metinlerde hassasiyet gerektirir (Lee vd. 2020).

2.7 Optimizasyon Algoritmaları

Makine öğrenmesi ve derin öğrenme modellerinde başarı, büyük ölçüde modelin parametrelerinin etkin bir şekilde optimize edilmesine bağlıdır. Bu süreçte kullanılan optimizasyon algoritmaları, modelin öğrenme yeteneğini, genelleme kapasitesini ve eğitim süresini doğrudan etkiler. Klinik NER gibi yüksek doğruluk gerektiren

görevlerde hem geleneksel hem de meta-sezgisel (metaheuristic) optimizasyon teknikleri önemli avantajlar sunmaktadır.

2.7.1 Geleneksel Optimizasyon Algoritmaları

Makine öğrenmesi ve derin öğrenme modellerinin eğitiminde, model parametrelerinin kayıp fonksiyonunu minimize edecek şekilde güncellenmesi kritik bir adımdır. Kullanılan geleneksel optimizasyon algoritmaları, özellikle gradyan tabanlı yöntemlerle çalışmakta ve büyük veri kümeleri ile yüksek boyutlu parametre uzaylarında etkili performans göstermektedir.

2.7.1.1 Stokastik Gradyan İnişi (SGD)

Stokastik Gradyan İnişi (SGD), makine öğrenimi modellerinin eğitiminde en temel yaklaşımlardan biridir. Geleneksel gradyan iniş yöntemlerinde, tüm veri kümesi kullanılarak kayıp fonksiyonunun gradyanı hesaplanırken, SGD yalnızca tek bir örnek veya küçük bir mini-batch üzerinden hesaplama yapar. Bu yöntem, büyük veri kümeleriyle çalışırken eğitim sürecini hızlandırmakta ve bellek verimliliği sağlamaktadır (Amari 1993).

Matematiksel olarak, SGD güncellemesi şu şekilde ifade edilir:

$$\theta_{t+1} = \theta_t - \eta . \nabla_\theta J(\theta; x^{(i)}, y^{(i)}) \tag{2.20}$$

Burada:

θ: Model parametrelerini temsil eder.
η: Öğrenme oranını temsil eder.
$\nabla_\theta J$: Parametrelere göre kayıp fonksiyonunun gradyanıdır.
$x^{(i)}$, $y^{(i)}$: Rastgele seçilmiş örnekleri temsil eder.

SGD, özellikle büyük ölçekli veri setleri üzerinde işlem yaparken önemli ölçüde zaman ve bellek avantajı sağlar. Ancak bu avantajlara rağmen, SGD'nin temel dezavantajı

gradyan yönünün veri örneklerine göre yüksek dalgalanma göstermesidir. Bu durum, öğrenme sürecinde kararsızlıklara ve hedef fonksiyonun minimumuna ulaşmada zorluklara yol açabilir. Ek olarak, sabit bir öğrenme oranı ile çalıştığı için, yakınsama sırasında dikkatli parametre ayarı yapılmadığında global minimuma ulaşmak zorlaşabilir.

2.7.1.2 Momentum

Momentum algoritması, SGD'nin aşırı salınım yapmasını azaltmak ve daha stabil bir yakınsama sağlamak amacıyla geliştirilmiştir. Bu yaklaşım, önceki adımlardan gelen gradyanları "hız vektörü" (momentum) olarak adlandırılan bir yapı içerisinde biriktirir ve bu vektörle ağırlık güncellemelerini yönlendirir. Böylece gradyanlar sadece anlık değil, tarihsel eğilimlere göre de değerlendirilir (Qian 1999).

$$\upsilon_t = \gamma_{t-1} + \theta_t - \eta . \nabla_\theta J(\theta), \qquad \theta_{t+1} = \theta_t - \nu_t \tag{2.21}$$

Burada;

γ: Momentum katsayısıdır (tipik olarak 0.9–0.99).
θ: Model parametrelerini temsil eder.
η: Öğrenme oranını temsil eder.
$\nabla_\theta J$: Parametrelere göre kayıp fonksiyonunun gradyanıdır.
v_t: Hız vektörüdür (birikmiş gradyan bilgisi).

Momentum, özellikle çok boyutlu parametre uzaylarında yerel minimumlara sıkışma olasılığını azaltır ve dar vadilerdeki salınımları sönümleyerek daha kararlı bir yakınsama eğrisi elde edilmesine olanak tanır (Nakerst vd. 2020).

2.7.1.3 RMSProp (Kök Ortalama Kare Yayılımı)

RMSProp algoritması, gradyan karelerinin üssel hareketli ortalamasını alarak her bir parametreye özgü öğrenme oranları üretir. Bu sayede model, her parametrenin güncelleme adımını o parametreye özgü olarak optimize edebilir (Tieleman ve Hinton 2012). Özellikle değişken ölçekli gradyanlara sahip ağ yapılarında etkilidir.

$$E[g^2]_t = pE[g^2]_{t-1} + (1-p)g^2{}_t \; ; \; \theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t \qquad (2.22)$$

Burada;

θ_t : *t.* zaman adımındaki model parametresidir.

g_t : *t*. adımda hesaplanan gradyandır *(∇J(θ)).*

η: Öğrenme oranını temsil eder.

$E[g^2]_t$: Gradyanların ikinci momentinin (karelerinin) üssel hareketli ortalamasıdır.

ρ*:* Çürüme katsayısıdır (decay rate) ve genellikle 0.9 civarındadır.

ϵ: Sayısal kararlılığı sağlamak amacıyla kullanılan çok küçük bir pozitif sabittir.

İlk denklemde yer alan $E[g^2]_t$ geçmiş adımlardaki gradyan karelerinin üssel ortalamasını ifade eder. Bu hesaplama, daha önceki gradyan değerlerine azalan ağırlık vererek daha yakın zamanlı gradyanların etkisini artırır. Böylece, parametrelerin güncellenmesinde kullanılan ölçek faktörü her parametreye özel hâle gelir.

İkinci denklem, bu adaptif ölçeklendirme sonrası yapılan parametre güncellemesini göstermektedir. Buradaki öğrenme oranı, karekök altında normalize edilen ikinci moment ile bölünür; bu da büyük gradyanlara sahip parametrelerde küçük güncellemeler yapılmasını, küçük gradyanlara sahip parametrelerde ise daha büyük adımlar atılmasını sağlar.

Bu yönüyle RMSProp, özellikle dengesiz gradyan yapısına sahip problemlerde, online öğrenme senaryolarında, NLP ve zaman serisi analizi gibi sıradüzenli verilerle çalışırken etkinliği kanıtlanmış bir yöntemdir (Tieleman ve Hinton 2012).

2.7.1.4 Adam (Uyarlanabilir Moment Tahmini)

Adam algoritması hem momentum hem de RMSProp yöntemlerinin avantajlarını birleştiren modern bir optimizasyon tekniğidir. Gradyanların ilk momentini (ortalama) ve ikinci momentini (kare ortalama) izleyerek, parametre güncellemelerini daha kararlı hâle getirir (Kingma ve Ba, 2014). Matematiksel olarak aşağdaki gibi ifade edilmektedir (Eşitlik 2.23).

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t, \qquad v_t = \beta_2 v_{t-1} + (1-\beta_2) g^2{}_t$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1}, \quad \hat{v}_t = \frac{v_t}{1-\beta_2^t} \;;\; \theta_{t+1} = \theta_t - \frac{\eta \hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{2.23}$$

Burada:

g_t: *t.* Zaman adımında hesaplanan gradyandıri

m_t*:* Gradyanların üssel ortalamasıdır (birinci moment, momentum benzeri).

v_t*:* Gradyan karelerinin üssel ortalamasıdır (ikinci moment, varyans tahmini).

β_1*:* Birinci moment için çürüme katsayısıdır (genelde 0.9).

β_2*:* İkinci moment için çürüme katsayısıdır (genelde 0.999).

$\hat{m}_t, \hat{v}_t$: Bias düzeltmeli birinci ve ikinci moment değerleridir.

ϵ: Sayısal kararlılık için küçük sabittir.

η: Öğrenme oranıdır.

Bu yapı, dengesiz gradyan dağılımlarının olduğu durumlarda dahi öğrenme sürecinin istikrarlı ve verimli ilerlemesine katkıda bulunur (Kingma ve Ba, 2014).

2.7.2 Meta-sezgisel Optimizasyon Algoritmaları

Meta-sezgisel algoritmalar, geleneksel türev temelli optimizasyon tekniklerinin yetersiz kaldığı, çok boyutlu, karmaşık ve genellikle türev bilgisi bulunmayan problem alanlarında etkili çözümler sunabilen sezgisel yöntemlerdir. Bu algoritmalar, doğa, biyoloji, evrim ve fizik gibi çeşitli doğal süreçlerden esinlenerek geliştirilmiş ve çözüm uzayında küresel optimuma yakın sonuçlar elde edebilmek amacıyla tasarlanmıştır (Yang 2010, Talbi 2009).

Geleneksel yöntemler çoğu zaman lokal minimumlara takılmakta, parametre hassasiyeti nedeniyle stabil çalışamamakta ya da karmaşık hedef fonksiyonlar üzerinde başarılı sonuçlar verememektedir. Meta-sezgisel yöntemler ise, çok sayıda aday çözümü paralel olarak değerlendirme kapasitesine sahip olduklarından, daha geniş bir arama uzayını kapsayarak çok modlu problemlerde etkinlik göstermektedir (Fister vd. 2013).

Bu algoritmaların ortak özellikleri şu şekilde özetlenebilir:

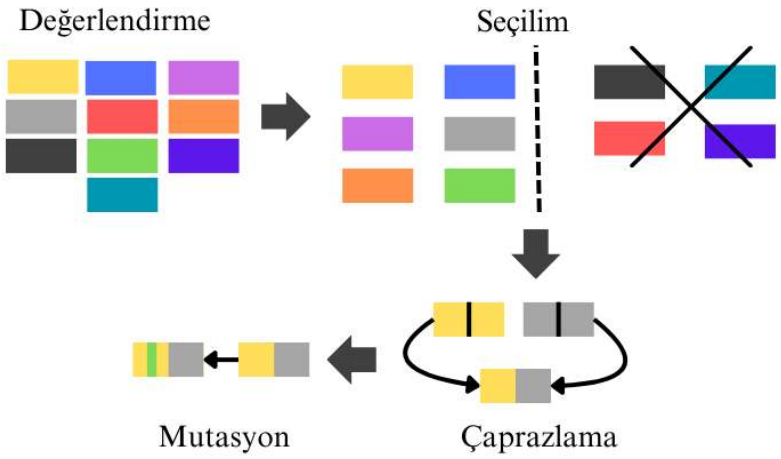
- Küresel arama yeteneğine sahip olmaları,
- Probleme özel bilgi gerektirmemeleri (türev, sürekli olma gibi),
- Stokastik yapıları sayesinde çeşitlilik sağlayarak lokal minimumlara saplanmayı azaltmaları,
- Paralel işlemeye uygunlukları nedeniyle yüksek boyutlu problemlerde avantaj sağlamaları,
- Çoklu çözüm üretme potansiyeliyle hiper-parametre optimizasyonu, model mimarisi seçimi gibi görevlerde esnek kullanım alanı sunmaları.

Meta-sezgisel algoritmalar; yapay sinir ağlarının ağırlıklarının öğrenilmesi, hiper-parametrelerin ayarlanması ve özellikle klinik NLP görevlerinde karmaşık model konfigürasyonlarının optimize edilmesi gibi birçok alanda yaygın biçimde kullanılmaktadır (Kaveh ve Mesgari 2023, Xue vd. 2021).

2.7.2.1 Genetik Algoritma (GA)

GA, doğal evrim sürecinden esinlenen ve John Holland tarafından 1975 yılında geliştirilen bir arama ve optimizasyon yöntemidir (Holland 1975). Evrimsel hesaplama ailesinin bir üyesi olan GA, problemler üzerinde çözüm üretmek için biyolojik evrimin temel ilkeleri olan seçilim, çaprazlama ve mutasyon mekanizmalarını kullanır. Özellikle çözüm uzayı çok büyük veya geleneksel türev temelli yöntemlerle optimize edilmesi güç olan problemlerde GA oldukça etkili bir alternatif olarak öne çıkmaktadır (Shapiro 1999). Genetik algoritmanın temel işleyişini gösteren adımlar Şekil 2.13'te ayrıntılı olarak gösterilmiştir.

GA, çözüm uzayında olası çözümleri temsil eden bireylerden (kromozomlardan) oluşan bir popülasyonla başlar. Bu bireyler, belirli bir uygunluk fonksiyonuna göre değerlendirilerek yeni nesiller üretmek amacıyla evrimsel işlemlere tabi tutulur. Süreç, belirli bir durdurma kriteri sağlanana kadar nesiller boyunca devam eder ve bu süreçte popülasyonun genel uygunluğu artar (Rooker 1991).



Şekil 2.13 GA temel işleyişini gösteren adımlar.

Genetik algoritmanın temel işleyişi, dört ana aşama etrafında şekillenmektedir. İlk olarak, değerlendirme sürecinde, popülasyonda yer alan her birey bir uygunluk (fitness) fonksiyonu aracılığıyla değerlendirilir. Bu uygunluk değeri, bireyin çözüme olan yakınlığını ifade eder ve sonraki adımlarda bireyin seçilme olasılığını doğrudan etkiler. İkinci aşama olan seçim sürecinde, daha yüksek uygunluk değerine sahip bireyler tercih edilir. Bu sayede, daha kaliteli genetik bilgiye sahip çözümler sonraki nesillere aktarılır ve evrimsel süreç yönlendirilmiş bir şekilde ilerletilir. Seçim işlemi genellikle rulet tekerleği, turnuva seçimi gibi yöntemlerle gerçekleştirilir. Ardından gelen çaprazlama aşamasında, seçilen bireyler rastgele çiftler hâlinde eşleştirilir. Bu bireyler genetik materyallerini değiş tokuş ederek yeni bireyler (çocuklar) oluştururlar. Çaprazlama işlemi, çözüm uzayında yeni ve potansiyel olarak daha iyi bireylerin oluşturulmasını sağlar. Son aşama olan mutasyon işleminde ise, çocuk bireylerin genetik yapısında belirli bir olasılıkla rastlantısal değişiklikler yapılır. Bu işlem, popülasyonun çeşitliliğini korumak, yerel optimumlara sıkışmayı engellemek ve çözüm uzayında keşfi sürdürülebilir kılmak açısından kritik bir rol oynamaktadır. Bu tez çalışmasında GA tercih edilmesinin temel nedeni, çok sayıda hiperparametre içeren karmaşık model yapılarında etkili ve esnek bir optimizasyon imkânı sunmasıdır.

2.7.2.2 Parçacık Sürü Optimizasyonu (PSO)

PSO, Kennedy ve Eberhart tarafından doğal yaşamdan ilhamla geliştirilen bir

popülasyon tabanlı optimizasyon algoritmasıdır (Kennedy ve Eberhart 1995). Kuşların veya balık sürülerinin toplu hâlde yiyecek ararken izledikleri sosyal davranış kalıplarını taklit ederek, çözüm uzayında etkili bir arama süreci gerçekleştirmeyi amaçlar. PSO, her bir olası çözümü bir "parçacık" olarak modelleyen ve bu parçacıkların pozisyonlarını uygunluk fonksiyonuna göre güncelleyerek küresel en iyi çözüme ulaşmayı hedefleyen bir yapıya sahiptir.

Algoritmanın temelinde, her bir parçacık; kendi en iyi konumu (personal best – pBest) ve tüm sürü içerisindeki en iyi konum (global best – gBest) bilgilerini dikkate alarak hareket eder. Her parçacık, bir hız vektörü ile temsil edilen yönelim bilgisine sahiptir ve bu vektör, parçacığın bir sonraki konumunu belirler (Kameyama 2009). Hız ve konum güncellemeleri aşağıdaki denklemler aracılığıyla gerçekleştirilir:

$$v_i(t+1) = w \,.\, v_i(t) + \, c_1 \,.\, r_1 \,.\, \big(pBest_i - x_i(t)\big) + c_2 \,.\, r_2 \,.\, \big(gBest - x_i(t)\big) \quad (2.24)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (2.25)$$

Burada:

$v_i\,(t)$: Parçacık i'nin t anındaki hızıdır.

$x_i\,(t)$: Parçacık i'nin t anındaki konumudur.

w: Atalet katsayısı (inertia weight), parçacığın önceki hareketini sürdürme eğilimini kontrol eder.

c_1 ve c_2: Bilişsel ve sosyal katsayılar (learning factors), genellikle 2 olarak seçilir.

r_1, r_2: [0-1] aralığında rastgele sayılardır.

$pBest_i$: Parçacığın şimdiye kadarki en iyi konumudur.

$gBest$: Sürünün genelindeki en iyi konumdur.

İlk terim olan $w.\ v_i(t)$, parçacığın önceki hızını temsil ederek hareketin sürekliliğini sağlar. İkinci terim, parçacığın kendi deneyiminden (bilişsel bileşen) öğrenmesini ifade ederken; üçüncü terim, sürüdeki diğer parçacıkların deneyimlerinden (sosyal bileşen) öğrenmeyi simgeler. Bu yapı sayesinde, PSO algoritması hem bireysel keşif hem de kolektif sömürü süreçlerini dengeli biçimde yürütür.

Parçacıkların bu şekilde yönlendirilmesi, çözüm uzayında çoklu bölgelerin etkin şekilde taranmasına ve küresel optimuma ulaşma olasılığının artmasına katkı sağlar. PSO, özellikle çok boyutlu, sürekli ve ayrık olmayan optimizasyon problemlerinde sade yapısı, az parametre gereksinimi ve yüksek hesaplama verimliliği ile dikkat çeker (Poli 2007). Bu tez çalışmasında PSO algoritması, derin öğrenme modelinin hiperparametrelerini optimize etmek amacıyla tercih edilmiştir. Parametre arama sürecinde geniş çözüm uzayını etkili bir şekilde tarayabilmesi ve diğer yöntemlere kıyasla daha düşük hesaplama maliyetiyle yüksek doğruluk değerlerine ulaşabilmesi, bu tercihin başlıca gerekçelerini oluşturmaktadır.

2.8 Literatürdeki Çalışmalar

Sağlık bilişimi alanında, yapılandırılmamış tıbbi metinlerden anlamlı ve yapılandırılmış bilgi çıkarımı, günümüzde giderek daha stratejik bir araştırma alanı hâline gelmiştir. ESK, radyoloji raporları, klinik notlar ve hasta özetleri gibi metin tabanlı veriler, tanı destek sistemlerinden epidemiyolojik analizlere kadar pek çok klinik ve bilimsel uygulama için kritik bilgiler barındırmaktadır. Ancak bu veriler, serbest metin formatında, standardize edilmemiş yapılar ve çeşitli uzmanlık alanlarına özgü terminolojik farklılıklar içerdiğinden, otomatik olarak işlenmeleri son derece zorlu bir süreçtir (Parameshwari vd. 2022).

NLP yöntemleri, özellikle NER görevlerinde, tıbbi kavramların (ör. hastalık, semptom, ilaç, test, prosedür) metin içerisinden otomatik olarak belirlenmesini ve etiketlenmesini sağlayarak kritik bir rol üstlenmektedir (Sheikhalishahi vd. 2019). Literatürde, bu hedefe yönelik olarak geliştirilen yöntemler zaman içinde önemli bir evrim geçirmiştir. Başlangıçta kural tabanlı yaklaşımlar ve sözlük temelli yöntemler ön planda iken; sonrasında destek vektör makineleri ve karar ağaçları gibi makine öğrenimi algoritmalarına dayalı modeller geliştirilmiştir. Son yıllarda ise, özellikle derin öğrenme mimarilerinin ve Transformer tabanlı önceden eğitilmiş dil modellerinin (BERT, BioBERT, ClinicalBERT) kullanımıyla, tıbbi NER görevlerinde bağlamsal anlam çıkarma kabiliyeti ve genel doğruluk oranlarında kayda değer artışlar sağlanmıştır.

2.8.1 Geleneksel Makine Öğrenimi Yöntemleri ile Yapılan Klinik NER Çalışmaları

Tıbbi metin madenciliği alanında gerçekleştirilen ilk çalışmaların önemli bir bölümü, kural tabanlı yaklaşımlar ve istatistiksel öğrenme tekniklerine dayanmaktadır. Bu yöntemler, belirli terminolojik kalıpları ve alan bilgisini temel alarak metinlerdeki tıbbi varlıkları tanımlamayı ve sınıflandırmayı hedeflemiştir. Özellikle radyoloji raporları gibi serbest metin formatındaki belgelerde bilgi çıkarımı amacıyla yapılan erken dönem uygulamalar, bu geleneksel yöntemlerin etkinliğini ortaya koymuştur.

2017 yılında Castro ve arkadaşları, mamografi raporlarını sınıflandırmak amacıyla kural tabanlı BROK algoritmasını kullanmışlardır. Çalışmada, öncelikle BI-RADS kategorileri mamografi raporlarından çıkarılmış, ardından bu veriler Bayes ağına giriş olarak verilerek radyoloji raporları malign ve benign olarak sınıflandırılmıştır (Castro vd. 2017). Bu çalışma, kural tabanlı sistemler ile olasılıksal modellerin hibrit kullanımına örnek teşkil etmektedir.

Benzer şekilde Nguyen ve ekibi, mamografi raporlarını otomatik özetlemek ve BI-RADS skorlarını sınıflandırmak amacıyla bir dil modeli ve sınıflayıcıdan oluşan hibrit bir sistem geliştirmiştir. Çalışmalarında farklı makine öğrenimi algoritmaları test edilmiş, en yüksek doğruluk oranı (%83,3) Destek Vektör Makineleri (SVM) ile elde edilmiştir. Bu bulgu, klasik makine öğrenmesi algoritmalarının meme kanseri sınıflandırmasında etkili bir çözüm sunduğunu göstermektedir (Nguyen vd. 2020).

Geleneksel makine öğrenimi yöntemlerinin sistematik analizine yönelik çalışmalardan biri, Casey ve arkadaşları tarafından gerçekleştirilmiştir. 2015–2019 yılları arasında yayımlanan ve radyoloji raporlarında doğal dil işlemeyi kullanan çalışmaları PRISMA ilkeleri doğrultusunda inceleyen bu sistematik tarama, literatürdeki yöntemlerin kapsamlı bir dökümünü sunmuştur (Casey vd. 2021).

Soomro ve arkadaşları ise biyomedikal metinlerde hastalık adları, ilaçlar, genler ve diğer tıbbi varlıkları otomatik olarak tanımlamak için hem kural tabanlı hem de

istatistiksel öğrenme yöntemlerini kullanmışlardır. Çalışmalarında, CRF başta olmak üzere çeşitli makine öğrenmesi algoritmaları test edilmiş; Bayesian Network, Naïve Bayes, DTNB ve NNGE modellerinin kombinasyonu ile 0,890 F1-skoruna ulaşılmıştır (Soomro vd. 2025). Bu sonuç, geleneksel makine öğrenimi yöntemleri doğru yapılandırıldığında hala rekabetçi performanslar sunabildiğini göstermektedir.

Geleneksel makine öğrenimi yöntemlerinin avantajları arasında yorumlanabilirlik ve düşük veri gereksinimi yer almakla birlikte, bağlamsal anlamı yeterince modelleyememeleri, çok anlamlılık ve terminolojik çeşitlilik karşısında zayıf performans sergilemeleri gibi önemli sınırlılıkları bulunmaktadır. Özellikle farklı hastaneler ya da uzmanlar arasında değişebilen ifade kalıpları, bu sistemlerin genel geçer bir model oluşturmasını zorlaştırmaktadır (Jia vd. 2025, Kamath vd. 2018). Bu nedenlerle, son yıllarda derin öğrenme temelli modeller, geleneksel sistemlerin yerini alarak bağlamsal ilişkileri daha etkin biçimde öğrenen ve daha yüksek doğruluk sunan yapay zekâ tabanlı çözümleri ön plana çıkarmıştır.

2.8.2 Derin Öğrenme Modelleri ile Yapılan Klinik NER Çalışmaları

Son yıllarda, NLP yaşanan ilerlemeler, özellikle derin öğrenme temelli modellerin yükselişiyle birlikte, tıbbi metinlerin daha etkin ve doğru biçimde analiz edilmesini mümkün kılmıştır. Radyoloji raporları gibi serbest metinlerin sınıflandırılması, özetlenmesi ve klinik anlam çıkarımı, bu gelişmelerin en çok uygulama bulduğu alanlardan biri hâline gelmiştir. Geleneksel yöntemlerin sınırlılıkları, bağlamsal ilişkileri yeterince modelleyememeleri ve manuel özelleştirme gereksinimi gibi nedenlerle belirginleşmiş; bunun sonucunda derin öğrenme mimarileri klinik bilgi çıkarımında ön plana çıkmıştır (Jia vd. 2023).

Magna ve arkadaşları, İspanyolca radyoloji raporlarından oluşan veri seti kullanarak gerçekleştirdikleri meme kanseri teşhisine yönelik çalışmalarında, Word2Vec ve çift yönlü LSTM (BiLSTM) tabanlı mimarilerle %98 gibi yüksek bir F1 skoru elde etmişlerdir (Magna vd. 2020). Bu çalışma, kelime gömme ve ardıl yapılarla bağlamsal öğrenmenin başarısını açıkça ortaya koymaktadır.

Benzer biçimde, Nakamura ve arkadaşları, BERT, LSTM ve geleneksel makine öğrenmesi modellerini karşılaştırdıkları çalışmalarında, Transformer temelli BERT modelinin radyoloji raporlarındaki bulguları tespit etmede daha yüksek doğruluk sağladığını rapor etmişlerdir (Nakamura vd. 2021). Bu durum, bağlamsal bilgiye dayalı modelleme kapasitesi sayesinde BERT'in dil anlayışında derin yapıların gücünü vurgulamaktadır.

Jaiswal ve arkadaşları, MIMIC-CXR veri seti üzerinde gerçekleştirdikleri çalışmalarında, RadBERT-CL adını verdikleri özel bir model geliştirerek, kritik tıbbi ifadelerin çıkarımında BERT ve BlueBERT gibi modellerden daha üstün performans göstermiştir (Jaiswal vd. 2021).

Yogarajan ve ekibi ise, MIMIC-III ve eICU veri setleri üzerinde geliştirdikleri Longformer mimarisini kullanarak, uzun sekanslı tıbbi belgelerde Transformer mimarilerinin etkili sonuçlar verdiğini ortaya koymuşlardır (Yogarajan vd. 2021).

Grancharova ve Dalianis, İsveççe elektronik hasta kayıtları üzerinde gerçekleştirdikleri çalışmalarında, İsveç diline özgü BERT modelinin NER ve sınıflandırma görevinde %92'nin üzerinde başarı sağladığını göstermiştir (Grancharova ve Dalianis 2021).

Benzer biçimde, Suárez-Paniagua ve arkadaşları, İspanyolca radyoloji raporları üzerinde çok dilli BERT modelleri ve Google Health NLP API'lerini birleştirdikleri hibrit sistemlerinde, en yüksek hatırlama oranlarına ulaşmışlardır (Suárez-Paniagua vd. 2021).

Yan ve arkadaşları, AB Savunma Bakanlığı Sağlık Sistemlerinden elde edilen raporlar üzerinde RadBERT varyantlarını test etmiş ve BERT-base, ClinicalBERT ve BioMed-RoBERTa gibi modellerle kıyasladıkları bu çalışmada, %10'dan daha az eğitim verisiyle dahi yüksek F1 puanları elde ettiklerini raporlamışlardır (Yan vd. 2022). Bu bulgu, Transformer tabanlı modellerin küçük veri kümelerinde dahi genelleme kapasitesine sahip olduğunu göstermektedir.

Fransız Radyoloji Departmanına ait 10 bin BT raporunu kullanan Delevaux ve ekibi, CamemBERT modelini pulmoner emboli tespiti için incelemiş ve yüksek doğruluk, özgüllük ve duyarlılık oranları elde etmişlerdir (Delevaux vd. 2023).

Gupta ve arkadaşları ise genetik mutasyonları sınıflandırmak için metin açıklamalarını kullanmış ve Word2Vec ile dönüştürdükleri vektörleri RNN ve diğer sınıflayıcılarla test etmişlerdir. Bu çalışmada RNN, %70 daha yüksek doğruluk oranı ile diğer algoritmalardan ayrışmıştır (Gupta vd. 2023).

Lopez-Ubeda ve arkadaşları, XLM-RoBERTa, BETO ve BioBERT-Spanish modellerini karşılaştırmalı biçimde kullanarak, İspanyolca mamografi raporlarının BI-RADS kategorisine göre sınıflandırılmasında BETO modelinin %77,5 doğruluk ile en iyi sonucu verdiğini, önceliklendirme görevinde ise BBS modelinin %91,02 doğruluk sağladığını rapor etmişlerdir (Lopez-Ubeda vd.2024).

Derin öğrenme yöntemlerinin bu denli etkili olmasının temelinde, yüksek boyutlu temsilleri öğrenebilme ve bağlamsal bilgiye dayalı karar verebilme yetenekleri yer almaktadır. Özellikle Transformer tabanlı mimariler, pozisyonel kodlama ve çok başlıklı dikkat mekanizmaları sayesinde uzun mesafeli bağımlılıkları modelleyebilmekte ve tıbbi metinlerin heterojen yapısını daha doğru biçimde çözümleyebilmektedir (Vaswani vd. 2017).

Dong ve arkadaşları tarafından önerilen çok görevli BiLSTM-CRF mimarisi, Çin tıbbi kayıtları üzerinde %3,2'lik F1 skoru artışı ile tıbbi varlıkların (hastalık, semptom, test) çıkarımında başarılı bir çözüm sunmuştur (Dong vd. 2019).

Yine Abbas ve arkadaşları, aktif öğrenme ve bağlamsal kelime temsillerini birleştirerek ClinicalBERT, SCIBERT ve DistilBERT gibi modellerin performansını karşılaştırmış ve ClinicalBERT'in %96 doğrulukla en iyi sonucu verdiğini ortaya koymuştur (Abbas vd. 2024). Bu alanda gerçekleştirilen diğer önemli derin öğrenme çalışmaları Çizelge 2.5'te özetlenmiştir.

Çizelge 2.5 Literatürde yer alan diğer çalışmalar.

Çalışmayı Yapan	Yıl	Veri Seti	Veri Sayısı	Model	Sonuçlar
Chen vd. (Chen vd. 2018)	2018	Göğüs BT (Özel Veri Seri)	4361	CNN	F1 skor 0,89
Trivedi vd. (Trivedi vd. 2018)	2018	İskelet Sistemi MR (Özel Veri Seti)	1544	IBM Watson	Doğruluk 0,85
Banerjee vd. (Banerjee vd. 2019)	2019	Göğüs Röntgen (Halka Açık Veri Seti: MIMIC-III)	10.000	RNN	F1 skor 0,70
Wang vd. (Wang vd. 2019)	2019	Femur Kırığı Röntgen (Halka Açık Veri Seti)	3955	CNN	F1 skor 0,97
Carrodeguas vd. (Carrodeguas vd. 2019)	2019	MIMIC-III (Halka Açık Veri Seti)	10.000	LSTM	F1 skor 0,60,
Yuan vd. (Yuan vd. 2019)	2019	Göğüs Röntgen (Halka Açık Veri Seti: IU-RR)	2549	CNN, LSTM	CNN: F1 skor 0,90 LSTM: F1 skor 0, 90
Huang vd. (Huang vd. 2021)	2021	Göğüs Röntgen (Halka Açık Veri Seti: MIMIC-III)	10.000	BERT	Hassasiyet 88.1
Dai vd. (Dai vd. 2021)	2021	Kemik Kırığı Röntgen (Özel veri seti)	12.712	BoneBERT	F1 skor 0.92

Farklı anatomik bölgelerden elde edilen radyolojik görüntüler ve metinler üzerinde geliştirilen bu modeller; CNN, RNN, LSTM ve Transformer tabanlı mimariler kullanarak çeşitli başarı metrikleriyle değerlendirilmiştir. Özellikle göğüs ve iskelet sistemi görüntüleri üzerinde yapılan çalışmalarda, %90'ın üzerinde F1 skoru ve doğruluk oranları elde edilmiştir. Huang ve arkadaşlarının MIMIC-III veri seti üzerinde BERT kullanarak gerçekleştirdikleri çalışma, bağlamsal öğrenme mimarilerinin tıbbi metinlerdeki potansiyelini ortaya koyarken (Huang vd. 2021); Dai ve arkadaşları tarafından önerilen BoneBERT modeli, kemik kırıklarının sınıflandırılmasında %0,92 F1 skoru ile dikkat çekici bir başarı göstermiştir (Dai vd. 2021).

Literatürdeki çalışmalar, Transformer temelli derin öğrenme modellerinin özellikle bağlamsal bilgiye dayalı NER görevlerinde önemli başarılar elde ettiğini göstermektedir. Klinik metinlerin yapısal karmaşıklığı, veri hacmi ve çokanlamlı doğası

göz önünde bulundurulduğunda, bu modellerin sunduğu yüksek doğruluk oranları hem bilimsel hem de klinik uygulamalar açısından büyük bir potansiyel barındırmaktadır.

2.8.3 Hibrit Modeller ile Yapılan Klinik NER Çalışmaları

Klinik metin madenciliğinde yüksek doğruluk oranı ve bağlamsal tutarlılık elde etme amacıyla son yıllarda hibrit model yaklaşımları giderek daha fazla ilgi görmektedir. Bu modeller, genellikle önceden eğitilmiş Transformer tabanlı mimarilerin sunduğu bağlamsal temsil gücünü, sıralı bağımlılıkları yakalama konusunda başarılı olan BiLSTM gibi tekrarlayan sinir ağı yapılarıyla ve etiket geçişlerini optimize eden CRF katmanıyla birleştirmektedir. Bu birleşim, NER görevlerinde hem semantik hem de yapısal bütünlüğü sağlama açısından önemli avantajlar sunmaktadır.

Literatürde bu tür hibrit mimarilerle gerçekleştirilen çalışmalar, BERT'in güçlü bağlam yakalama yeteneği ile BiLSTM'in sıralı veri üzerindeki başarısını ve CRF'nin etiketleme doğruluğunu birleştirerek etkileyici sonuçlar ortaya koymuştur. Dai ve arkadaşları, Çin ESK üzerinde geliştirdikleri BERT-BiLSTM-CRF modeliyle, CCKS 2019 ve SAHSU veri kümelerinde yaklaşık %75 F1-skoru elde ederek bu yapının başarısını ortaya koymuştur (Dai vd. 2019).

Hou ve arkadaşları, bu mimariye Dinamik Dikkat Mekanizması ekleyerek modelin performansını daha da artırmış ve %81,68 F1-skoruna ulaşmıştır (Hou vd. 2024).

Benzer şekilde, Tang, Çince tıbbi metinlerde ALBERT-IDCNN-CRF modelini kullanarak %83,37 F1-skoru elde etmiş, modelin özellikle düşük kaynaklı ortamlarda yüksek doğruluk ve işlem verimliliği sunduğunu göstermiştir (Tang 2022).

Pan ve arkadaşları ise Geleneksel Çin Tıbbı EMR verileri üzerinde ALBERT-BiLSTM-CRF modelini uygulamış, ALBERT'in parametre verimliliği ile BiLSTM'in dizisel bağımlılıkları etkili biçimde modelleyebildiğini, CRF katmanının ise etiketleme doğruluğunu artırdığını göstermiştir. Bu hibrit yapı, F1-skorunu geleneksel yöntemlere

kıyasla %2 oranında artırarak, hesaplama maliyetlerinin azaltılmasıyla birlikte yüksek doğruluk sunmuştur (Pan vd. 2022).

Bununla birlikte, hibrit modeller yalnızca sıralı etiketleme görevlerinde değil, aynı zamanda tıbbi ölçek çıkarımı ve hasta durumu değerlendirmesi gibi ileri düzey klinik çıkarım görevlerinde de kullanılmaktadır. Yang ve arkadaşları, ClinicalBERT-BiLSTM-CRF temelli bir model geliştirerek elektronik sağlık kayıtlarından inme ile ilişkili NIHSS ölçeği öğelerini otomatik olarak çıkarmayı hedeflemiştir. Modelin, CART (Sınıflandırma ve Regresyon Ağaçları) ile birlikte kullanılması sonucunda %88,23 F1-skoru ile en iyi performansı gösterdiği bildirilmiştir (Yang vd. 2023).

Biyomedikal metinlerdeki karmaşık terim ilişkilerini ve semantik yapıları anlamada da hibrit mimariler etkili olmuştur. Örneğin, Alamro ve arkadaşları, BioBERT-BiLSTM-CRF temelli BioBBC adlı modeli önermiş; bu model, NCBI-Disease, BC5CDR-Disease ve BC5CDR-Chem veri kümelerinde en yüksek F1-skorlarına ulaşarak bağlamsal temsillerin gücünü açık biçimde ortaya koymuştur (Alamro vd. 2024)

Tüm bu bulgular ışığında, BERT ve türevleriyle geliştirilen hibrit modellerin, yalnızca semantik bilgi yakalamada değil, aynı zamanda dizisel yapıların korunmasında da üstün başarı sergilediği görülmektedir. Bu durum, özellikle klinik karar destek sistemleri ve hasta güvenliği açısından kritik öneme sahip tıbbi varlık tanıma görevlerinde, bu modellerin tercih edilme nedenini açıklamaktadır.

2.8.4 Meta-sezgisel Algoritmalar ile Yapılan Klinik NER Çalışmaları

Son yıllarda, derin öğrenme tabanlı modellerin klinik metin madenciliği görevlerinde sergilediği başarı, bu modellerin daha kararlı ve optimize biçimde çalışmasını sağlayacak tekniklerin araştırılmasını da beraberinde getirmiştir. Bu kapsamda, hiper-parametre optimizasyonu, model performansı üzerinde doğrudan etkili bir faktör olarak öne çıkmaktadır. Ancak, klasik hiper-parametre ayarlama yöntemlerinin yüksek boyutlu parametre alanlarında yetersiz kalması ve hesaplama maliyetinin artması nedeniyle, GA ve PSO gibi meta-sezgisel optimizasyon teknikleri, tıbbi NER görevlerinde yaygın

şekilde uygulanmaya başlanmıştır (Mohammed vd. 2024, Abiodun vd. 2021, Karnatapu vd. 2020).Özellikle BERT tabanlı modellerin hibrit yapılarla birleştirilmesiyle elde edilen yüksek doğruluk oranları, meta-sezgisel algoritmalarla optimize edildiğinde daha da gelişmektedir.

Kollapally ve Geller tarafından gerçekleştirilen bir çalışmada, Clinical BioBERT modeli üzerine uygulanan GA tabanlı hiper-parametre arama yöntemiyle %91,91 doğruluk oranı elde edilmiştir. GA, modelin hiperparametrelerini optimize ederek doğruluk oranının belirgin şekilde artmasına katkı sağlamıştır (Kollapally ve Geller 2023)

Benzer şekilde, Abdennour ve arkadaşları, Genetik Algoritma ile optimize edilen SVM, Random Forest ve Lojistik Regresyon bileşenlerinden oluşan hibrit yapıda %92,22 doğruluk oranına ulaşmış ve GA'nın sınıflandırma başarısını artırmadaki etkinliğini göstermiştir (Abdennour vd. 2024).

Meta-sezgisel optimizasyonun doğrudan NER görevlerine uygulanması açısından dikkate değer bir çalışma da Peng ve arkadaşları tarafından sunulmuştur. Araştırmacılar, BERT-CRF modeline PSO entegre ederek hiper-parametre seçim sürecini otomatikleştirmişlerdir. Deneysel sonuçlar, bu hibrit yapının %96,94 F1-Skoru ile, klasik CRF, BiLSTM-CRF ve CNN-CRF modellerine kıyasla daha yüksek doğruluk sağladığını ortaya koymuştur (Peng vd. 2023).

PSO'nun sınıflandırma görevlerindeki performansı ise Gasmi tarafından değerlendirilmiştir. BERT tabanlı model kullanılarak tıbbi metinlerden özellik çıkarılmış ve hiper-parametreler PSO algoritması ile optimize edilmiştir. Çalışma, Hallmarks veri seti üzerinde %71 doğruluk oranı ile başarılı sonuçlar sunmuştur (Gasmi 2022). Buna ek olarak, çok dilli ve alan-spesifik veri kümeleri için meta-sezgisel tekniklerin uygulanabilirliğini değerlendiren Cuevas ve arkadaşları, MédicoBERT adlı İspanyolca tıbbi NER modeli için Uyarlanabilir PSO (APSO) yöntemini kullanmışlardır. Deneyler sonucunda, APSO uygulamasıyla F1-skorunda %30,21 oranında artış sağlandığı bildirilmiştir (Cuevas vd. 2024).

Bu çalışmalar, meta-sezgisel algoritmaların yalnızca hiper-parametre optimizasyonu sağlamakla kalmayıp, aynı zamanda modelin genel kararlılığını ve genellenebilirliğini de önemli ölçüde artırdığını göstermektedir. Ancak literatür incelendiğinde, bu yöntemlerin çoğunlukla farklı veri kümeleri ve model yapıları üzerinde ayrı ayrı test edildiği ve doğrudan karşılaştırmaya olanak tanımadığı görülmektedir. Ayrıca, hesaplama maliyeti, eğitim süresi ve model kararlılığı gibi açılardan GA ve PSO'nun sistematik biçimde karşılaştırıldığı çalışmaların oldukça sınırlı olduğu dikkat çekmektedir.

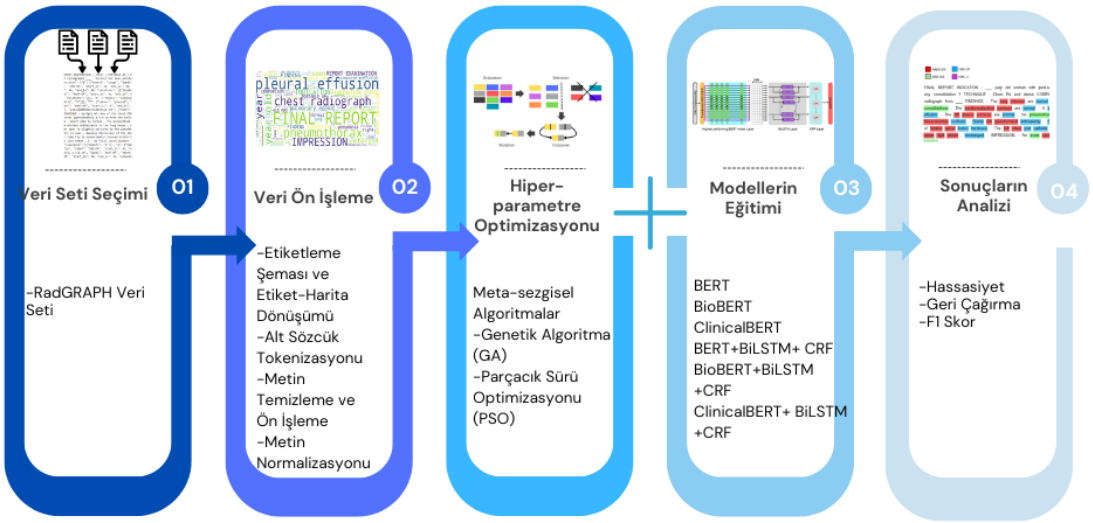
Bu doğrultuda, tez kapsamında yalnızca standart dil modellerinin performansını değerlendirmekle yetinilmemiş, aynı zamanda bu modeller üzerine inşa edilen hibrit mimariler (BiLSTM+CRF) ile birlikte kapsamlı bir optimizasyon stratejisi geliştirilmiştir. BERT, BioBERT ve ClinicalBERT modelleri temel alınarak oluşturulan bu yapılar, klinik metinlerdeki bağlamsal ve sıralı bağımlılıkları daha etkin biçimde modelleyebilmek amacıyla BiLSTM ve CRF katmanlarıyla entegre edilmiştir. Söz konusu modellerin hiper-parametrelerinin en uygun biçimde belirlenebilmesi için GA ve PSO gibi literatürde sık başvurulan meta-sezgisel algoritmalar, aynı veri setleri ve deneysel koşullar altında karşılaştırmalı olarak uygulanmıştır. Bu kapsamda yalnızca doğruluk, kesinlik, duyarlılık ve F1-Skoru gibi klasik başarı ölçütleri değil; aynı zamanda modelin eğitilme süresi, parametre güncelleme kararlılığı ve genel öğrenme verimliliği gibi operasyonel performans göstergeleri de dikkate alınmıştır.

3. MATERYAL ve YÖNTEM

Bu çalışmada, göğüs radyoloji raporları üzerinde NER görevini gerçekleştirmek amacıyla derin öğrenme temelli klinik NLP modelleri tasarlanmış ve bu modellerin başarımı sistematik olarak analiz edilmiştir. Literatürdeki güçlü performansları nedeniyle tercih edilen Transformer tabanlı modeller (BERT, BioBERT, ClinicalBERT), sıralı etiketleme yeteneğine sahip BiLSTM ve CRF katmanları ile entegre edilerek hibrit mimariler oluşturulmuştur. Modelin hassasiyet, geri çağırma ve F1-skor metrikleri üzerinden değerlendirilmesi hedeflenmiş; buna ek olarak GA ve PSO meta-sezgisel algoritmaları ile hiper-parametre ayarlamaları gerçekleştirilmiştir.

Çalışmanın metodolojik çerçevesi, Şekil 3.1'de sunulduğu üzere dört temel aşamadan oluşmaktadır:

(1) Veri kümesi seçimi,
(2) Klinik metinler üzerinde veri ön işleme,
(3) Meta-sezgisel optimizasyon algoritmaları ile hiper-parametre ayarı ve model eğitimi,
(4) Elde edilen sonuçların analizi ve karşılaştırmalı değerlendirme.



Şekil 3.1 Çalışma akışı.

Her bir aşama, modelin hem içsel doğruluğunu hem de klinik uygulamalara uygunluğunu artıracak biçimde yapılandırılmıştır. Bu bağlamda, ilk olarak etiketli bir radyoloji raporu veri kümesi belirlenmiş, ardından bu metinlere özel ön işleme teknikleri (temizleme, etiket ayrıştırma, tokenize etme) uygulanmıştır. Daha sonra, belirlenen mimarilerin hiper-parametreleri meta-sezgisel teknikler kullanılarak optimize edilmiştir. Son aşamada ise, tüm modeller performans açısından değerlendirilmiş, en etkili yapı ve optimizasyon stratejisi belirlenmiştir.

Bu çok katmanlı yaklaşım sayesinde yalnızca bir modelin performansı değil, farklı mimarilerin ve optimizasyon stratejilerinin NER başarımına etkisi de kapsamlı biçimde incelenmiştir. Böylelikle, klinik metin madenciliği alanında bütüncül bir modelleme ve karşılaştırma perspektifi sunulması hedeflenmiştir.

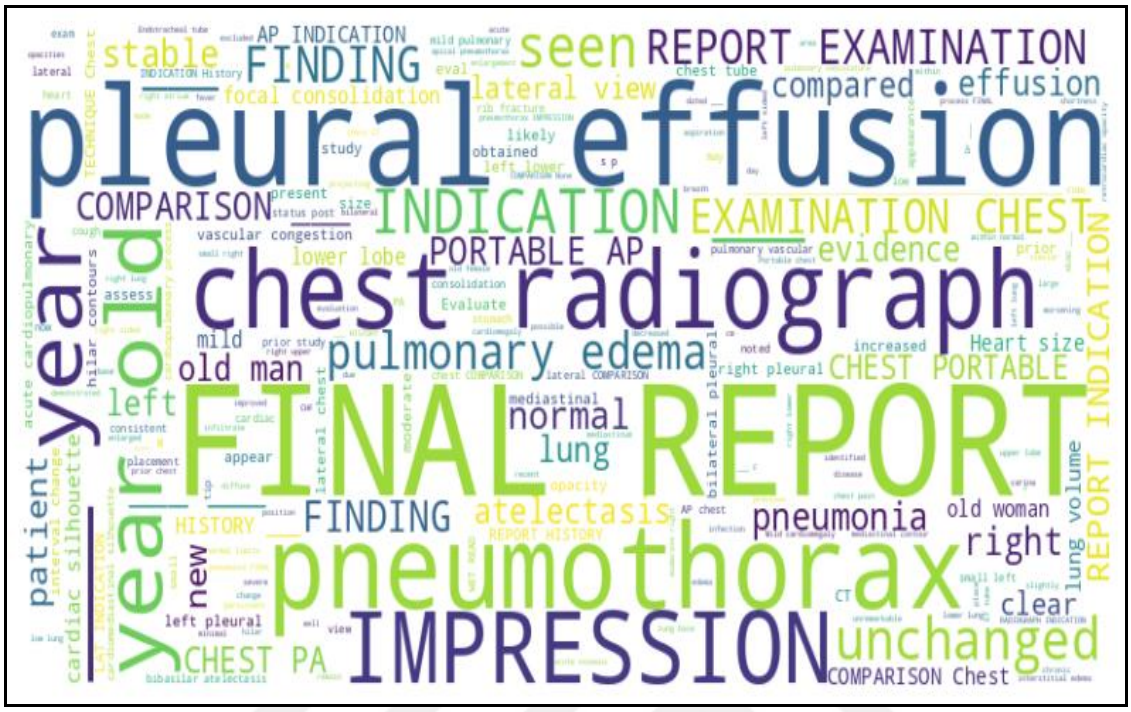
3.1 Veri Seti Seçimi

Bu çalışmada, klinik doğal dil işleme alanında yaygın olarak kullanılan ve açık erişimli bir kaynak olan RadGraph veri kümesi tercih edilmiştir (Jain 2021). RadGraph, PhysioNet platformu üzerinden araştırmacıların kullanımına sunulmuş, etiketli göğüs radyoloji raporları içeren kapsamlı bir veri setidir. Bu veri kümesi, tıbbi varlık tanıma ve ilişki çıkarımı gibi görevlerde kullanılmak üzere, MIMIC-CXR (Johnson vd. 2019) ve CheXpert (Irvin vd. 2019) veri kümelerinden türetilmiştir. Şekil 3.2'de, RadGraph veri setinde en sık karşılaşılan tıbbi kavramlardan bazıları örneklenerek görselleştirilmiştir.

RadGraph veri seti, tıbbi radyoloji raporları içindeki semantik varlıkları dört ana başlık altında sınıflandırmaktadır:

- ANAT-DP (Definite Anatomy): Vücutta tanımlanmış anatomik bölgeleri temsil eder.
- OBS-DP (Definite Observation): Görüntüleme bulgularında belirgin şekilde gözlemlenen patolojik veya fizyolojik durumları ifade eder.
- OBS-DA (Definitely Absent Observation): Radyolojik olarak bulunmadığı açıkça belirtilmiş gözlemleri tanımlar.

- OBS-U (Uncertain Observation): Belirsizlik içeren ya da şüpheli gözlemleri kapsar.



Şekil 3.2 RadGraph veri setinden elde edilen yaygın kullanılan kelimeler.

Veri setinin oluşturulması sürecinde, Datasaur.ai platformu kullanılarak raporlar üç radyolog tarafından Dr. Curt Langlotz'un geliştirdiği anotasyon şemasına uygun şekilde etiketlenmiştir. Etiketleme işlemi sırasında tıbbi doğruluğun korunmasına özen gösterilmiş ve varlık türleriyle bunlar arasındaki ilişkiler manuel olarak işaretlenmiştir.

RadGraph veri setine erişim, HIPAA yönetmeliklerine uygun biçimde kişisel sağlık bilgileri (PHI) kimliksizleştirilmiş olan MIMIC-CXR ve CheXpert veri setlerini kapsayan etik ve teknik eğitimlerin başarıyla tamamlanması şartıyla erişilebilir olmuştur. Veri erişimi için iki bölümden ve 16 modülden oluşan bir eğitim programı tamamlanmış, sınavlarda %90 başarı oranı elde edilmiştir

Çalışmada modelin eğitimi ve değerlendirilmesi üç aşamalı bir yapı üzerinden gerçekleştirilmiştir:

- Eğitim Aşaması: 425 adet MIMIC-CXR radyoloji raporundan oluşan alt küme, modelin öğrenmesi için kullanılmıştır.

- Doğrulama (Validasyon) Aşaması: 75 adet MIMIC-CXR raporundan oluşan alt küme, hiper-parametre ayarı ve erken durdurma kriterleri için değerlendirilmiştir.
- Test Aşaması: Modelin genelleme yeteneğini değerlendirmek amacıyla, 50 MIMIC-CXR ve 50 CheXpert raporundan oluşan test alt kümeleri kullanılmıştır. Ek olarak, modelin daha geniş bir popülasyona uygulanabilirliğini test etmek amacıyla, tam MIMIC-CXR test kümesi ve CheXpert test kümesi üzerinde ayrı ayrı testler gerçekleştirilmiştir.

Bu yapı sayesinde, modelin hem eğitim verisi üzerindeki başarımı hem de dış veri kümeleri üzerindeki genellenebilirliği bütüncül bir şekilde analiz edilmiştir.

3.2 Veri Ön İşleme

Makine öğrenmesi ve derin öğrenme tabanlı NLP modellerinde elde edilecek başarım, büyük ölçüde ön işleme sürecinin kalitesine bağlıdır. Özellikle tıbbi metinler gibi yapısal olmayan ve yüksek terminolojik çeşitlilik içeren veri türlerinde, etkili bir ön işleme adımı, modelin öğrenme sürecini hem anlamlı hem de verimli hâle getirmede kritik rol oynamaktadır. Bu kapsamda, çalışmada kullanılan RadGraph (Jain 2021) veri setine yönelik ön işleme süreci, üç temel aşamadan oluşmaktadır: (1) etiketleme şemasının standardizasyonu, (2) alt kelime düzeyinde tokenizasyon, ve (3) metin temizleme ve normalizasyon işlemleri. Çalışmada gerçekleştirilen veri ön işleme adımları Şekil 3.3'te sunulmuştur.

Şekil 3.3 Çalışmada gerçekleştirilen veri ön işleme adımları.

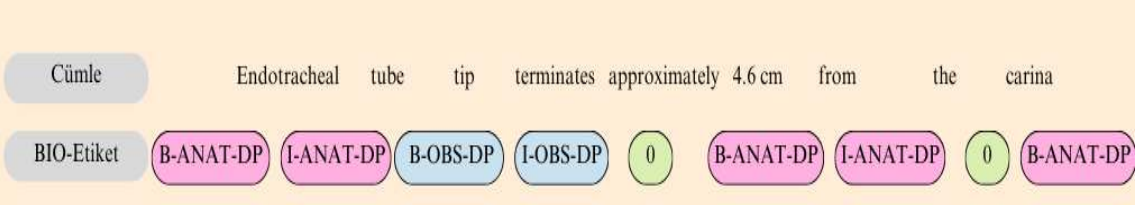
3.2.1 Etiketleme Şeması ve Haritalama İşlemi

Tıbbi NER görevlerinde, modelin sıralı ve yapısal bilgiye duyarlılığını artırmak için

literatürde yaygın olarak kullanılan BIO (Beginning, Inside, Outside) etiketleme şeması benimsenmiştir (Alamro vd. 2024). Bu şema, bir varlık ifadesinin sınırlarını açıkça belirleyebilmekte ve modelin dizisel bağımlılıkları daha etkili biçimde öğrenmesini sağlamaktadır.

- B- (Beginning): Varlık sınıfına ait bir ifadenin ilk kelimesini belirtir.
- I- (Inside): Varlığın devam eden öğelerini temsil eder.
- O (Outside): Herhangi bir varlık sınıfına ait olmayan kelimeleri ifade eder.

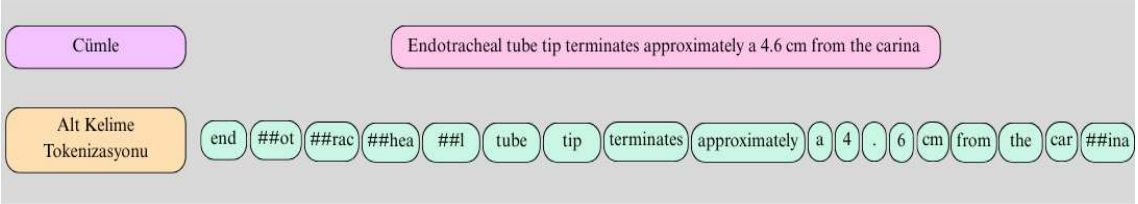
Bu yapı, çok sözcüklü tıbbi terimlerin doğru biçimde modellenmesini sağlarken, aynı zamanda etiketleme hatalarını azaltarak yanlış pozitif oranlarını minimize etmektedir. Bu çalışmada, RadGraph veri setinde yer alan dört temel tıbbi varlık türü olan ANAT-DP (Anatomik Yapı), OBS-DP (Kesin Gözlem), OBS-DA (Olumsuz Gözlem), ve OBS-U (Belirsiz Gözlem) etiketleri, BIO formatına uygun şekilde işaretlenmiş ve etiketleme süreci standardize edilmiştir. Bu etiketleme yaklaşımı, modelin anatomik yapıları ve gözlemsel bulguları bağlamsal doğrulukla tanımasını mümkün kılmaktadır. Şekil 3.4'te örnek bir radyoloji raporu üzerinden yapılan BIO etiketleme işlemi, kelime bazlı sınıf ayrımıyla görselleştirilmiştir.



Şekil 3.4 BIO etiket haritalama örneği.

3.2.2 Alt Kelime Düzeyinde Tokenizasyon

Tıbbi metinlerin karakteristik özelliklerinden biri, nadir görülen terimlerin ve bileşik sözcüklerin yaygın olarak kullanılmasıdır. Bu durum, kelime düzeyinde yapılan vektörleştirme işlemlerinin yetersiz kalmasına neden olmaktadır. Bu nedenle, çalışmada alt kelime düzeyinde tokenizasyon yöntemi uygulanmış ve özellikle Transformer tabanlı modellerle tam uyumlu olan WordPiece algoritması tercih edilmiştir (Guo vd. 2022). Şekil 3.5'te örnek bir tıbbi ifadede kelimelerin WordPiece algoritması ile hangi alt bileşenlere bölündüğü gösterilmiştir.



Şekil 3.5 Alt kelime tokenizasyon örneği.

WordPiece algoritması, bir kelimeyi anlamlı alt bileşenlerine ayırarak modelin morfolojik özellikleri ve dilin iç yapısını daha etkin biçimde öğrenmesine olanak tanır. Bu yöntem, sözlük dışı (out-of-vocabulary) terimlerle karşılaşıldığında dahi anlam taşımayan rastgele temsiller üretmek yerine, tanıdığı parçalardan anlamlı bir bütün oluşturarak modelin genelleme kabiliyetini artırır. Özellikle tıbbi terminolojideki çoklu bileşik kelimelerin analizinde yüksek başarı sağlar.

3.2.3 Metin Temizleme ve Normalizasyon

Alt kelime düzeyinde ayrıştırılan metinlerin model girdisine uygun hâle getirilebilmesi için kapsamlı bir metin temizleme ve normalizasyon süreci uygulanmıştır. Bu sürecin temel amacı, veri kümesindeki yapısal ve dilsel tutarsızlıkları gidererek, modelin hem öğrenme etkinliğini artırmak hem de genel başarı oranını yükseltmektir (Pinto vd. 2016).

Öncelikle, kelime temsillerinde gereksiz varyasyonlara neden olabilecek harf duyarlılığı ortadan kaldırılmış ve tüm metinler küçük harfe dönüştürülmüştür (Mielke vd. 2021). Bu dönüşüm, aynı anlama gelen ancak farklı yazım biçimlerine sahip kelimelerin model tarafından farklı kavramlar olarak algılanmasının önüne geçmiştir. Diğer yandan metin içerisindeki noktalama işaretleri ele alınmış, sözdizimsel bağlamın korunabilmesi amacıyla bu karakterler kelimelerden ayrı olarak birer token şeklinde işlenmiştir. Böylece, noktalama işaretlerinin cümle yapısındaki işlevleri model tarafından açıkça öğrenilebilir hâle gelmiştir.

Buna ek olarak, veri setinde yer alan fazladan boşluklar, model açısından anlamsız olan özel simgeler ve biçimlendirme bozuklukları da temizlenerek metinlerin yapısal

bütünlüğü sağlanmıştır. Bu işlemler sonucunda, modelin bağlamsal ilişkileri daha etkin bir biçimde öğrenmesi mümkün olmuş; böylece NER görevindeki başarım önemli ölçüde iyileştirilmiştir (Paiş ve Tufiş 2022). Aynı zamanda, bu normalizasyon süreci eğitim verisindeki gürültüyü azaltarak modelin daha kararlı, güvenilir ve genellenebilir çıktılar üretmesini desteklemiştir.

3.3 Modeller

Bu çalışmada, göğüs radyoloji raporlarında yer alan tıbbi varlıkların otomatik olarak tanımlanması amacıyla derin öğrenme tabanlı çeşitli model mimarileri kullanılmıştır. Özellikle NER görevlerinde yüksek başarı sağlamasıyla bilinen Transformer tabanlı önceden eğitilmiş dil modelleri, bu çalışmanın temelini oluşturmaktadır. Bu kapsamda, genel dil modeli olan BERT, biyomedikal makalelerle eğitilmiş BioBERT ve klinik notlara özgü olarak ince ayar yapılmış ClinicalBERT modelleri değerlendirilmiştir.

Söz konusu modeller, bağlamsal dil temsillerini etkili biçimde öğrenebilme yetenekleriyle öne çıkarken, bu temsillerin daha doğru biçimde etiketlenmesini sağlamak amacıyla BiLSTM ve CRF katmanları ile zenginleştirilmiştir. BiLSTM katmanı, çift yönlü bilgi akışı sayesinde kelimeler arası dizisel ilişkileri modelleyerek bağlamın hem geçmişini hem de geleceğini öğrenebilmekte; CRF katmanı ise etiket geçiş olasılıklarını dikkate alarak sıralı etiketleme görevlerinde tahmin tutarlılığını artırmaktadır.

Bu yaklaşımla oluşturulan hibrit mimariler (BERT+BiLSTM+CRF, ClinicalBERT +BiLSTM+CRF, BioBERT+BiLSTM+CRF), yalnızca bağlamsal anlam bilgisini değil, aynı zamanda tıbbi varlıklar arasındaki yapısal ve semantik ilişkileri de etkili bir şekilde modelleyebilmektedir. Böylece, elde edilen modeller yalnızca doğruluk açısından değil, aynı zamanda genelleme kabiliyeti, hesaplama verimliliği ve etiketleme tutarlılığı bakımından da karşılaştırmalı olarak analiz edilmiştir.

3.3.1 BERT

Bu çalışmada kullanılan temel modellerden biri olan BERT, çift yönlü bağlamsal öğrenme kabiliyeti sayesinde, tıbbi metinlerdeki karmaşık dil yapılarının daha doğru

analiz edilmesine olanak tanımaktadır (Devlin vd. 2019). BERT modeli, yalnızca encoder katmanlarından oluşan bir Transformer mimarisi üzerine kuruludur ve kelimelerin anlamını hem öncesindeki hem de sonrasındaki bağlamdan eşzamanlı olarak öğrenmektedir.

Modelin girişine verilen her bir kelime, token'lara ayrılarak sabit boyutta vektörlere dönüştürülmekte; ardından konumsal ve segment bilgileriyle birleştirilerek modelin işlem kapasitesi artırılmaktadır. Bu yapı, tıbbi metinlerde sık karşılaşılan çok anlamlı veya bağlama duyarlı terimlerin daha doğru temsil edilmesini sağlamaktadır. Modelin çıktısı, her token için bağlamsal bir temsil üretmekte; bu temsiller sırasıyla BiLSTM ve CRF katmanlarına aktarılmak üzere kullanılmaktadır. Bu sayede, modelin yalnızca kelime düzeyinde anlam çıkarımı değil, aynı zamanda dizisel bağımlılıkları da etkin biçimde modelleyebilmesi mümkün hale gelmiştir.

Bu çalışmada kullanılan BERT modeli, genel dil verileri üzerinde önceden eğitilmiş bir versiyon olup, biyomedikal terminolojiye özgü semantik örüntüler konusunda sınırlı bir yetkinliğe sahiptir. Bu nedenle, klinik bağlamda daha başarılı sonuçlar elde etmek amacıyla, BioBERT ve ClinicalBERT gibi alan uyarlanmış sürümler de değerlendirmeye alınmıştır. Ancak BERT, karşılaştırmalı analizler için temel yapı olarak konumlandırılmış, hibrit mimarideki etkisi izole biçimde test edilerek diğer modellerle karşılaştırılmıştır.

3.3.2 BioBERT

BioBERT, tıbbi ve biyomedikal metinleri anlamak için BERT modelinin tıbbi verilerle yeniden eğitilmiş bir sürümüdür (Lee vd. 2020). Bu sayede, tıbbi terminolojinin daha iyi anlaşılmasını sağlayarak genel amaçlı BERT modeline kıyasla biyomedikal metinlerde daha iyi performans göstermektedir.

Bu çalışmada BioBERT, tıbbi varlıkların bağlamsal anlamlarını doğru şekilde öğrenmek ve göğüs radyoloji raporlarında geçen hastalık, anatomik yapı ve klinik bulguların tanımlanmasını sağlamak için kullanılmıştır. Ancak BioBERT, doğrudan

klinik metinlerle eğitilmediği için klinik bağlamdaki kullanım alanlarında belirli sınırlamalara sahiptir. Bu nedenle, elektronik sağlık kayıtları ve klinik raporlar üzerinde eğitilmiş ClinicalBERT modeli de bu çalışmada değerlendirilmiştir.

3.3.3 ClinicalBERT

ClinicalBERT, tıbbi ve klinik bağlamda kelime anlamlarını daha iyi yakalamak için MIMIC-III (Medical Information Mart for Intensive Care III) (Johnson vd. 2016) verileriyle önceden eğitilmiş bir BERT türevidir (Alsentzer vd. 2019). Elektronik sağlık kayıtları içeren bu veri seti, hastane ortamında doktorlar tarafından yazılmış klinik notları ve hasta geçmişlerini içermektedir.

ClinicalBERT'in BioBERT'e kıyasla avantajı, klinik dil yapısına daha iyi adapte olması ve tıbbi terminolojinin hastane ortamındaki kullanımını daha iyi öğrenmesidir. Göğüs radyoloji raporlarında sıkça rastlanan kısaltmalar, terim kullanımları ve spesifik ifadeler, ClinicalBERT'in bağlamsal anlama yeteneği ile daha doğru bir şekilde modellenebilir.

3.3.4 BiLSTM Katmanı

BiLSTM katmanı, dizisel veriler üzerinde uzun vadeli bağımlılıkları yakalamak ve sıralı bilgiyi daha etkin bir şekilde modellemek için kullanılan bir sinir ağı katmanıdır (Schuster ve Paliwal 1997). Çift yönlü yapıya sahip olan BiLSTM, yalnızca kelimenin geçmiş bağlamını değil, gelecekteki bağlamını da dikkate alarak daha kapsamlı ve tutarlı tahminler yapmaktadır. Bu sayede, model her kelimenin bağlamını hem önceki hem de sonraki kelimelerle birlikte değerlendirerek sıralı bağımlılıkları daha iyi öğrenebilmekte ve uzun mesafeli bağlam ilişkilerini etkili bir şekilde yakalayabilmektedir. Özellikle tıbbi metinlerde, bir terimin anlamının çevresindeki kelimelerle şekillendiği göz önüne alındığında, BiLSTM'in sağladığı bu avantajlar varlık tanıma görevlerinde doğruluk oranını önemli ölçüde artırmaktadır.

3.3.5 CRF Katmanı

CRF, sıralı etiketleme görevlerinde, özellikle NER gibi problemlerde sıklıkla tercih edilen ayrımcı olasılık temelli bir modeldir (Lafferty vd. 2001). Transformer tabanlı modeller ve BiLSTM gibi derin öğrenme yapıları, her kelime için bireysel tahminler üretebilmekte; ancak bu yaklaşımlar, kelimeler arasındaki etiket geçiş ilişkilerini doğrudan modelleyememektedir. CRF katmanı, bu eksikliği gidermek üzere sekans düzeyinde tahmin yaparak tüm dizideki etiket geçişlerini birlikte değerlendirir ve böylece daha tutarlı etiketleme sonuçları elde edilmesini sağlar.

Bu çalışmada kullanılan CRF yapısı, bölüm 2.5.4'de tanımlanan toplam skor fonksiyonuna *(s (h, y))* dayalıdır ve bu skorların normalize edilmesiyle koşullu olasılık fonksiyonu *P(Y/X)* elde edilmektedir. Dolayısıyla model, yalnızca bireysel kelime temsillerini değil, aynı zamanda bu temsillerin ardışık etiket dizisi içerisindeki geçiş tutarlılığını da göz önünde bulundurur. CRF'in olasılık temelli formülasyonu verilmiştir (Denklem 3.1).

$$P(Y/X) = \frac{\exp\left(\sum_{t=1}^{T} \alpha_{yt} . \emptyset(y_t, y_{t-1}, x_t)\right)}{Z(X)} \quad (3.1)$$

Burada:

P(Y/X): Verilen giriş dizisi X için en olası etiket dizisi Y'nin olasılığını temsil eder.
α_{yt}: Her bir etiketin önem derecesini belirleyen ağırlık fonksiyonudur. Nadiren görülen etiketlerin kaybını artırarak dengesiz veri setlerinde daha doğru tahminler yapmayı sağlar.
$\Phi\ (y_t,\ y_{t-1},\ x_t)$*:* her bir kelimenin belirli bir etikete sahip olma olasılığı ile ardışık etiketler arasındaki geçiş olasılığını içeren skor fonksiyonudur.
Z(X): Normalizasyon faktörüdür ve olasılıkların toplamını 1 yapmak için kullanılır.

Bu yapı sayesinde, örneğin bir kelimeye "B-OBS-DP" etiketi verildiğinde, CRF modeli, bir sonraki kelime için "I-OBS-DP" etiketini yüksek olasılıkla önererek etiket

geçişlerinde semantik uyumu sağlar. Böylece yalnızca doğruluk oranları değil, çıktıların bağlamsal tutarlılığı da artırılmış olur.

CRF katmanının bu özellikleri, özellikle BIO etiketleme şemasıyla birlikte kullanıldığında daha belirgin hale gelir. Model, nadir görülen ancak klinik açıdan kritik öneme sahip etiketlerin doğru tahmin edilebilmesi için α_{yt} ağırlıkları üzerinden bu sınıflara daha fazla dikkat verir. Ayrıca, veri setinde sıklık dengesizliği bulunan senaryolarda bile CRF, bu ağırlıklandırma mekanizması sayesinde genel performansı koruyarak dengesiz sınıf dağılımlarının etkisini azaltır.

3.3.6 Önerilen Model

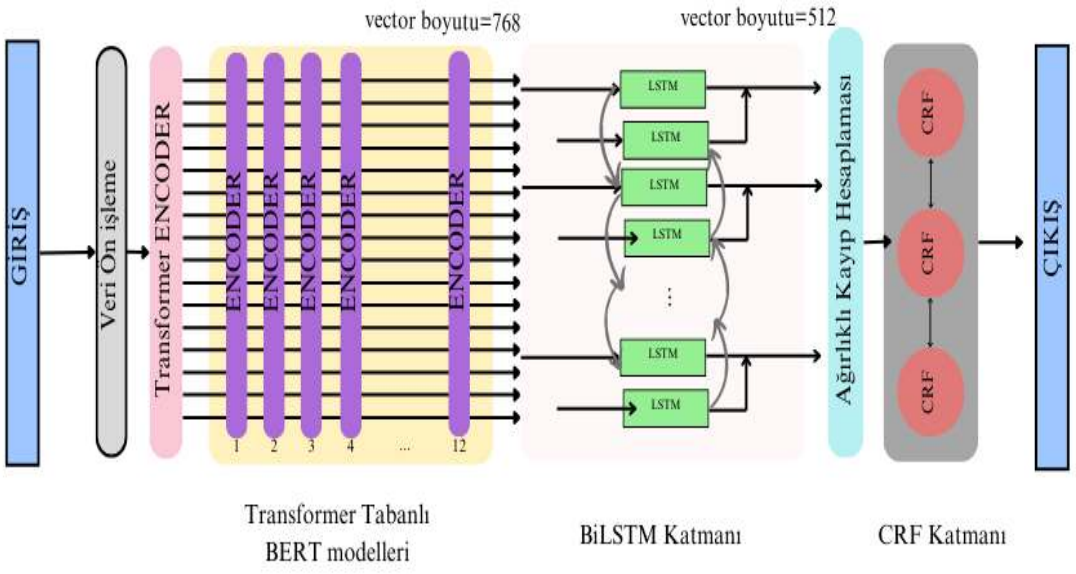
Transformer tabanlı dil modelleri olan BERT, BioBERT ve ClinicalBERT, doğal dil işleme alanında bağlamsal kelime temsillerinin öğrenilmesinde çığır açıcı bir rol üstlenmiştir. Bu modeller, kelimeleri yalnızca kendi anlamlarıyla değil, aynı zamanda içinde bulundukları bağlama göre çift yönlü olarak temsil edebilmekte; bu sayede özellikle tıbbi metinlerde sıkça karşılaşılan çok anlamlı ifadelerin ve karmaşık terminolojilerin daha doğru biçimde çözümlenmesini sağlamaktadır (Grancharova ve Dalianis 2021, Abbas vd. 2024, Fu vd. 2024). Özellikle BioBERT ve ClinicalBERT modelleri, biyomedikal yayınlar ve elektronik sağlık kayıtları üzerinde önceden eğitilerek, domain-özgü bağlamları genel modellerden çok daha yüksek doğrulukla modelleyebilme yetisine sahiptir.

Ancak, bu modellerin sahip olduğu güçlü bağlamsal temsil gücüne rağmen, dil içindeki sıralı bağımlılıkları ve etiket geçiş tutarlılıklarını doğrudan modelleyememesi, etiketleme görevlerinde bazı sınırlamalara yol açabilmektedir. Bu sorunun üstesinden gelmek amacıyla, Transformer tabanlı modellerin çıktıları, BiLSTM katmanı ve CRF katmanı ile zenginleştirilmiştir (Ma vd. 2019).

BiLSTM katmanı, her kelime için hem önceki hem de sonraki bağlamları dikkate alarak uzun vadeli bağımlılıkları modellemektedir. Bu yapı, radyolojik raporlar gibi uzun ve teknik metinlerde bağlamın sürdürülebilirliğini güçlendirerek modelin kavramsal

bütünlük oluşturmasına yardımcı olur. CRF katmanı ise, etiketleme aşamasında yalnızca kelime bazlı tahminlerle yetinmeyip, etiketler arasındaki geçiş olasılıklarını da dikkate alarak tüm sekans boyunca en olası etiket dizisini üretir. Bu sayede, örneğin bir "B-OBS-DP" etiketinden sonra "I-OBS-DP" gelme ihtimalini artırarak, çıktılarda yapısal tutarlılığı temin eder.

Bu üçlü yapı birlikte çalışarak hem kelime düzeyinde derin bağlamsal temsiller üretmekte hem de sıralı bağımlılıkları ve etiket geçiş tutarlılıklarını optimize etmektedir. Böylelikle, göğüs radyoloji raporlarında yer alan tıbbi varlıkların doğru biçimde tanımlanması ve sınıflandırılması mümkün hale gelmekte; klinik doğal dil işleme sistemlerinin güvenilirliği ve etkinliği artırılmaktadır. Modelin genel yapısı, Şekil 3.6'da şematik olarak sunulmuştur.



Şekil 3.6 Önerilen model yapısı.

3.4 Meta-sezgisel Algoritmalar

Derin öğrenme tabanlı modellerin başarımı, yalnızca mimari yapılarına değil, aynı zamanda bu yapıların doğru hiper-parametrelerle yapılandırılmasına da doğrudan bağlıdır. Hiper-parametreler; öğrenme oranı, katman sayısı, gizli birim boyutu, dropout oranı gibi modelin eğitim sürecini belirleyen temel yapıtaşlarıdır. Ancak bu

parametrelerin manuel veya ızgara arama gibi klasik yöntemlerle belirlenmesi, yüksek hesaplama maliyeti ve yerel optimumlara saplanma riski gibi dezavantajlar taşımaktadır. Özellikle büyük ve karmaşık çözüm uzaylarında, bu geleneksel yöntemler çoğu zaman yetersiz kalmaktadır.

Bu nedenle, son yıllarda hiper-parametre optimizasyonunda meta-sezgisel algoritmalar ön plana çıkmaktadır. Bu algoritmalar, doğadan ilham alan sezgisel yöntemler olup, geniş arama uzaylarında küresel optimuma ulaşma yetenekleriyle dikkat çekmektedir. Popülasyon tabanlı yaklaşımları sayesinde, çözüm uzayını çok boyutlu olarak tarayabilmekte ve yerel maksimumlardan kaçınarak daha kararlı sonuçlar üretebilmektedirler.

Bu çalışmada, önerilen modellerin hiper-parametrelerinin optimize edilmesi amacıyla literatürde etkinliği kanıtlanmış iki farklı meta-sezgisel yöntem kullanılmıştır: GA ve PSO. Bu algoritmaların uygulanmasındaki temel amaç, model doğruluğunu ve genelleme yeteneğini maksimize edecek hiper-parametre kombinasyonlarını otomatik olarak keşfetmektir. GA, evrimsel biyolojiden ilham alarak genetik çaprazlama ve mutasyon operatörleri ile çözümleri geliştirirken; PSO, sürü zekasına dayalı bir yaklaşım ile bireysel ve kolektif bilgi akışını kullanarak çözüm uzayında gezinmektedir. Bu iki algoritmanın tercih edilme nedeni, farklı arama stratejileriyle çözüm uzayını tamamlayıcı biçimde keşfetmeleri ve derin öğrenme modelleri için hesaplama açısından verimli ve etkili optimizasyon performansı sunmalarıdır.

Her iki algoritma da modelin öğrenme oranı, epok sayısı, BiLSTM gizli katman boyutu, dropout oranı ve CRF geçiş skorları gibi hiper-parametrelerin optimize edilmesinde kullanılmıştır. Böylece, önerilen hibrit modeller yalnızca mimari düzeyde değil, aynı zamanda parametre düzeyinde de en verimli yapılandırmaya kavuşturulmuş ve daha tutarlı sonuçlar elde edilmiştir.

3.4.1 GA ile Hiper-parametre Optimizasyonu

Bu çalışmada kullanılan hibrit modellerin başarımını artırmak amacıyla, hiper-parametre optimizasyon süreci için GA tercih edilmiştir (Deb 1998). Geleneksel

yöntemlerin aksine, GA'nin geniş çözüm uzaylarını verimli biçimde tarayabilme ve yerel minimumlara sıkışmadan küresel optimuma yakın sonuçlara ulaşma yeteneği, model eğitiminin performansını doğrudan artırmıştır.

Uygulamada, modelin öğrenme oranı, BiLSTM katmanlarının gizli boyutu, dropout oranı, CRF ağırlıkları gibi hiper-parametreler GA ile optimize edilmiştir. Bu süreçte kullanılan genetik yapı, her bireyin bir hiper-parametre kombinasyonunu temsil ettiği şekilde tanımlanmış ve uygunluk fonksiyonu olarak modelin doğruluk ve F1-Skoru esas alınmıştır.

Bu çalışmada kullanılan Genetik Algoritma, 6 bireyden oluşan bir başlangıç popülasyonu ile yapılandırılmıştır. Her birey, farklı hiper-parametre kombinasyonlarını temsil edecek şekilde tanımlanmış ve bu bireyler GA süreci boyunca uygunluk fonksiyonuna göre değerlendirilmiştir. Optimizasyon süreci, 3 ardışık nesil boyunca yürütülmüş ve her nesilde yeni bireyler çaprazlama ve mutasyon işlemleriyle türetilmiştir. Popülasyonun çeşitliliğini korumak ve çözüm uzayının daha geniş bir bölümünün keşfedilmesini sağlamak amacıyla, mutasyon oranı %30 olarak belirlenmiştir. Bu yapılandırma hem hesaplama verimliliğini hem de model başarımını maksimize edecek şekilde optimize edilmiştir. GA sürecinde elde edilen en iyi birey, nihai modelin eğitiminde kullanılan hiper-parametre seti olarak seçilmiştir. Çizelge 3.1'de çalışmada kullanılan parametre ayarları sunulmuştur.

Çizelge 3.1 Çalışmada kullanılan GA parametreleri.

Popülasyon büyüklüğü	6
Nesil sayısı	3
Mutasyon oranı	0.3

Bu yapılandırma sayesinde, manuel ayarlama gereksinimi ortadan kaldırılarak sistematik bir hiper-parametre optimizasyonu gerçekleştirilmiş ve modellerin doğruluk oranlarında önemli artış sağlanmıştır. Aynı zamanda hesaplama maliyeti açısından dengeli bir optimizasyon süreci yürütülmüştür.

3.4.2 PSO ile Hiper-parametre Optimizasyonu

Bu çalışmada, derin öğrenme modellerinin başarımını artırmak amacıyla hiper-parametrelerin sistematik ve otomatik şekilde ayarlanması için PSO algoritması uygulanmıştır (Kennedy ve Eberhart 1995). PSO'nun tercih edilmesindeki temel etken, klasik arama yöntemlerine kıyasla düşük hesaplama maliyetiyle geniş çözüm uzaylarında etkili bir şekilde çalışabilmesi ve hızlı yakınsama (convergence) özellikleri göstermesidir. Özellikle yüksek boyutlu hiper-parametre uzaylarında lokal minimumlara sıkışma riskini azaltarak daha küresel çözümler elde edebilme yeteneği, PSO'yu hiper-parametre ayarlama süreçleri için güçlü bir araç hâline getirmiştir.

Optimizasyon sürecinde, her bir parçacık; modelin öğrenme oranı, BiLSTM katman boyutu (hidden size) ve CRF geçiş skorlarına ilişkin ağırlıklar gibi temel hiper-parametre kombinasyonlarını temsil edecek şekilde kodlanmıştır. Başlangıçta rastgele belirlenen parçacık konumları ve hızları, deneysel olarak tanımlanan uygunluk (fitness) fonksiyonuna göre değerlendirilmiş ve her iterasyonda hem bireysel en iyi konum (pBest) hem de küresel en iyi konum (gBest) bilgisi dikkate alınarak güncellenmiştir

PSO algoritmasının uygulanmasında kullanılan parametre değerleri atalet katsayısı (inertia weight), bilişsel katsayı (c1), sosyal katsayı (c2), iterasyon sayısı ve parçacık popülasyon büyüklüğü gibi literatürde önerilen aralıklarda kalacak şekilde ayarlanmış ve Çizelge 3.2'de sunulmuştur. Deneysel bulgular, bu parametre konfigürasyonlarının modelin genelleme yeteneği üzerinde doğrudan etkili olduğunu ve optimizasyonun, özellikle temel modellerin performansını belirgin şekilde iyileştirdiğini göstermiştir.

Çizelge 3.2 Çalışmada kullanılan PSO parametreleri.

Popülasyon büyüklüğü	6
İterasyon sayısı	3
Eylemsizlik katsayısı	0.7
Bilişsel katsayı	1.5
Sosyal katsayı	1.5

Eylemsizlik (inertia) katsayısı, parçacıkların mevcut hızlarına olan bağlılık düzeyini belirlerken; bilişsel ve sosyal katsayılar, sırasıyla parçacığın kendi geçmişinden ve topluluktaki en iyi çözümden ne ölçüde etkileneceğini tanımlamaktadır. Her iterasyonda parçacıkların hem kendi optimum geçmiş konumları hem de tüm popülasyonun en iyi çözümü göz önünde bulundurularak, hiper-parametre alanında yönlendirilmiş bir arama gerçekleştirilmiştir.

Bu yapı sayesinde, parametreler önceden sabit değerler yerine veri ve görev özelliklerine göre otomatik biçimde belirlenmiş, bu da modelin genel doğruluk ve genelleme kapasitesinde kayda değer iyileşmeler sağlamıştır.

3.5 Performans Değerlendirme Metrikleri

Bu çalışmada geliştirilen derin öğrenme ve hibrit modellerin başarımı, klinik metin madenciliği bağlamında önemli bir görev olan NER üzerinde ölçülmüştür. Model değerlendirmesi, NER görevlerinde standardizasyon sağlaması açısından literatürde yaygın olarak kullanılan Kesinlik (Precision), Duyarlılık (Recall) ve F1 Skoru gibi temel performans metrikleri çerçevesinde gerçekleştirilmiştir (Derczynski 2016).

Bu metrikler, yalnızca modelin doğru sınıflandırma yapabilme becerisini değil, aynı zamanda hatalı veya eksik tahminlerini de dikkate alması yönüyle bütüncül bir değerlendirme imkânı sunmaktadır. Özellikle dengesiz dağılıma sahip klinik veri kümelerinde, yalnızca doğruluk (accuracy) metriğine dayalı bir değerlendirme, yanıltıcı sonuçlar doğurabileceği için daha hassas ölçütlerin kullanılması tercih edilmiştir.

Kesinlik, modelin pozitif sınıf olarak tahmin ettiği örneklerin ne kadarının gerçekten doğru olduğunu ifade eder. Özellikle yanlış pozitif oranlarının önemli olduğu uygulamalarda, bu metriğin yüksek olması kritik kabul edilir. Kesinlik metriğinin hesaplama formülü Eşitlik 3.2'de sunulmuştur.

$$Kesinlik = \frac{TP}{TP + FP} \tag{3.2}$$

Duyarlılık ise, modelin gerçek pozitif örnekleri ne dereceyle yakalayabildiğini ölçer. Bu metrik, yanlış negatif tahminleri azaltmayı hedefler ve genellikle tıbbi tanılama gibi hata maliyeti yüksek alanlarda öne çıkar. Duyarlılık metriğinin hesaplama formülü Eşitlik 3.3'de sunulmuştur.

$$Duyarlılık = \frac{TP}{TP + FN} \tag{3.3}$$

F1 Skoru, kesinlik ve duyarlılık metriklerinin harmonik ortalamasıdır. Bu metrik, iki değerin dengeli bir şekilde optimize edilmesini sağlamaktadır Özellikle sınıflar arası dengesizlik içeren görevlerde F1-Skoru, model başarımını daha adil bir şekilde yansıtma özelliğine sahiptir. F1 Skor metriğinin hesaplama formülü Eşitlik 3.4'de sunulmuştur.

$$F1\ Skor = 2x\frac{Kesinlik\ x\ Duyarlılık}{Kesinlik\ + Duyarlılık} \tag{3.4}$$

Bu metriklerin hesaplanmasında kullanılan değişkenler aşağıdaki şekilde tanımlanmaktadır: TP (True Positive – Doğru Pozitif), modelin pozitif sınıfa ait olduğunu doğru bir şekilde tahmin ettiği örnekleri ifade eder. Örneğin, bir varlık ismini doğru şekilde tanımlaması bu kategoriye girer. FP (False Positive – Yanlış Pozitif), modelin pozitif olarak sınıflandırdığı ancak gerçekte pozitif olmayan, yani yanlış bir şekilde etiketlenen örnekleri temsil eder. Bu durum, modelin bir kelimeyi yanlışlıkla tıbbi varlık olarak tanımlaması halinde ortaya çıkar. FN (False Negative – Yanlış Negatif) ise modelin negatif (varlık değil) olarak tahmin ettiği fakat aslında pozitif (gerçekte bir varlık) olan örneklerdir. Başka bir ifadeyle, modelin gözden kaçırdığı ve etiketlemediği gerçek varlıklar bu kategoriye dâhildir.Bu değişkenler aracılığıyla modelin yalnızca doğru tahmin kapasitesi değil, aynı zamanda eksik ya da hatalı tahminlerindeki eğilimleri de değerlendirilmekte, böylece modelin genel başarımı daha bütüncül bir biçimde analiz edilebilmektedir.

4. BULGULAR

Bu bölümde, geliştirilen modellerin performans analizleri sunulmuştur. BERT, BioBERT ve ClinicalBERT tabanlı modeller; BiLSTM ve CRF katmanlarıyla entegre edilerek oluşturulan hibrit yapılarla birlikte değerlendirilmiştir. Model başarımı, CheXpert ve MIMIC-CXR test veri setlerinde Precision, Recall ve F1-Skoru metrikleri kullanılarak ölçülmüştür.

İlk olarak, herhangi bir optimizasyon veya ağırlıklandırma uygulanmadan elde edilen temel sonuçlar sunulmuş; ardından manuel alfa ağırlıklandırma yöntemi ile performans gelişimi gözlemlenmiştir. Son aşamada ise hiper-parametrelerin GA ve PSO ile optimize edilmesiyle elde edilen sonuçlar karşılaştırmalı olarak analiz edilmiştir. Deneysel çalışmalar, Google Colab Pro+ platformunda GPU desteği ile gerçekleştirilmiştir.

4.1 Ağırlıklandırma ve Optimizasyon Uygulanmamış Ana Modellerin Sonuçları

Bu bölümde, herhangi bir ağırlıklandırma stratejisi ya da hiper-parametre optimizasyonu uygulanmadan elde edilen temel model çıktıları değerlendirilmiştir. Çalışmada kullanılan BERT, BioBERT ve ClinicalBERT tabanlı modeller, doğrudan NER görevine entegre edilerek test edilmiştir. Model başarımı, yaygın olarak kullanılan CheXpert ve MIMIC-CXR test veri setleri üzerinde, Kesinlik, Duyarlılık ve F1-Skoru gibi performans metrikleri aracılığıyla analiz edilmiştir.

Çizelge 4.1'de sunulan sonuçlar, yalnızca önceden eğitilmiş dil modeli katmanları kullanılarak elde edilen temel performansı yansıtmaktadır. Bu değerlendirme, hibrit mimarilerle kıyaslama yapılabilmesi açısından referans niteliği taşımaktadır. Temel modellerin tek başına ne ölçüde etkili olabildiği bu veriler üzerinden analiz edilebilirken, aynı zamanda daha karmaşık yapıların getireceği katkının da nesnel biçimde ortaya konmasına olanak tanımaktadır. Böylece hibrit mimarilerin sağlayacağı performans artışının anlamlılığı, bu taban sonuçlar ışığında daha sağlıklı bir şekilde değerlendirilebilmektedir.

Çizelge 4.1 Ana model sonuçları.

Modeller	**CheXpert Test Veri Seti**			**MIMIC-CXR Test Veri Seti**		
	Kesinlik	Duyarlılık	F1 Skor	Kesinlik	Duyarlılık	F1 Skor
BERT	0,7339	0,7266	0,7310	0,7479	0,7449	0,7460
BioBERT	0,7902	0,7550	0,7720	0,7923	0,7765	0,7840
ClinicalBERT	0,7733	0,7628	0,7680	0,7933	0,7828	0,7880
BERT+BiLSTM+CRF	**0,9505**	**0,9428**	**0,9470**	**0,9496**	**0,9479**	**0,9490**

Çizelge 4.1'de yer alan bulgular, yalnızca Transformer tabanlı ana modellerin sınırlı düzeyde başarı sağladığını; buna karşılık BiLSTM ve CRF katmanlarının entegrasyonu ile elde edilen hibrit yapının, model performansında kayda değer bir iyileşme sağladığını ortaya koymaktadır.

4.2 Sabit Ağırlıklandırma Uygulanmış Modellerin Sonuçları

Bu çalışmada, modelin sınıflar arasındaki dengesiz veri dağılımına karşı daha dirençli hale gelmesini sağlamak amacıyla sabit ağırlıklandırma stratejisi uygulanmıştır. Bu strateji kapsamında, her bir etiket için manuel olarak belirlenen alfa değerleri kullanılarak, nadir görülen varlık türlerinin model tarafından daha etkin biçimde öğrenilmesi hedeflenmiştir. Özellikle sınıf sıklığı düşük olan ANAT-DP ve OBS-U gibi etiketlerin eğitim sürecinde daha fazla önem taşıması sağlanarak, modelin bu tür varlıkları eksiksiz ve tutarlı şekilde tanıması amaçlanmıştır. Bu yaklaşım, varlık ismi tanıma görevinde sınıf dengesizliğinden kaynaklanan tahmin hatalarının azaltılmasına katkı sağlamıştır. Modelin bu sabit ağırlıklandırma stratejisiyle elde ettiği performans değerleri Çizelge 4.2'de sunulmuştur.

Çizelge 4.2 Sabit ağırlıklandırma tekniği kullanılan model sonuçları.

Modeller	**CheXpert Test Veri Seti**			**MIMIC-CXR Test Veri Seti**		
	Kesinlik	Duyarlılık	F1 Skor	Kesinlik	Duyarlılık	F1 Skor
BERT	0,8132	0,7968	0,8040	0,8164	0,8112	0,8130
BioBERT	0,8082	0,7947	0,8010	0,8121	0,7901	0,7990
ClinicalBERT	0,8193	0,8018	0,8110	0,8483	0,8409	0,8450
BERT+BiLSTM+ CRF	**0,9525**	**0,9513**	**0,9520**	**0,9570**	**0,9564**	**0,9570**

Sabit ağırlıklandırma stratejisinin uygulanması sonucunda, özellikle BERT+BiLSTM+ CRF hibrit modelinde yaklaşık %0,8 oranında bir performans artışı elde edilmiştir. Bu artış, hibrit yapının zaten yüksek bir başarı düzeyine sahip olmasına rağmen, sınıf ağırlıklarının dikkatli bir şekilde ayarlanmasının modele ek katkılar sağlayabildiğini göstermektedir. Öte yandan, temel modellerde (BERT, BioBERT, ClinicalBERT) sabit ağırlıklandırma sonrası %7,3'e varan anlamlı performans artışları gözlemlenmiştir. Bu bulgular, özellikle düşük sıklığa sahip etiketlerin model tarafından yeterince öğrenilebilmesi için örnek ağırlıklandırmasının kritik bir faktör olduğunu ortaya koymaktadır. Sonuçlar, etiket dağılımındaki homojenliğin artırılmasının, modelin genelleme kabiliyeti ve sınıflar arası ayırt ediciliği açısından belirleyici rol oynadığını açıkça göstermektedir.

4.3 Meta-sezgisel Algoritmalar ile Optimize Edilen Model Sonuçları

Çalışmanın bu aşamasında, sabit ağırlıklandırma stratejisinin ötesine geçilerek, model başarımını daha sistematik ve optimize bir biçimde iyileştirmek amacıyla GA ve PSO gibi yaygın olarak kullanılan meta-sezgisel optimizasyon teknikleri uygulanmıştır. Doğal dil işleme alanında sıklıkla tercih edilen bu algoritmalar, yüksek boyutlu hiper-parametre uzaylarında etkin çözüm arama yetenekleriyle öne çıkmakta olup, bu çalışmada model performansını en üst düzeye taşımak üzere değerlendirilmiştir.

Model yapılandırmasında, yalnızca geleneksel BERT+BiLSTM+CRF hibrit modeliyle sınırlı kalınmamış; klinik metin madenciliğine özgü olarak önceden eğitilmiş BioBERT ve ClinicalBERT modellerinin, göğüs radyoloji raporlarında daha yüksek bağlamsal temsil gücü sunacağı öngörülerek bu modeller de analiz kapsamına dâhil edilmiştir. Bu doğrultuda, BioBERT+BiLSTM+CRF ve ClinicalBERT+BiLSTM+CRF mimarileri hem GA hem de PSO algoritmaları ile ayrı ayrı optimize edilerek detaylı biçimde değerlendirilmiştir.

Her iki optimizasyon yaklaşımı ile modelin öğrenme oranı, maksimum dizi uzunluğu, batch boyutu, epoch sayısı ve alfa ağırlıkları gibi kritik hiper-parametreleri en uygun değerlerine ulaşacak şekilde ayarlanmış ve sonuçta elde edilen modeller, CHEXPERT

ve MIMIC-CXR test veri setleri üzerinde Kesinlik, Duyarlılık ve F1-Skoru ölçütleriyle karşılaştırmalı olarak analiz edilmiştir. Optimizasyon sonrası elde edilen tüm model çıktıları Çizelge 4.3'te detaylı biçimde sunulmuştur.

Çizelge 4.3 GA ve PSO ile optimize edilen model sonuçları.

Optimizasyon Algoritmaları	**Modeller**	**CheXpert Test Veri Seti**			**MIMIC-CXR Test Veri Seti**		
		Kesinlik	Duyarlılık	F1 Skor	Kesinlik	Duyarlılık	F1 Skor
GA	BERT	0,8894	0,8865	0,8889	0,9168	0,9182	0,9180
	BioBERT	0,8966	0,8912	0,8930	0,9446	0,9428	0,9440
	ClinicalBERT	0,8914	0,8921	0,8917	0,9200	0,9200	0,9200
	BERT +BiLSTM+CRF	0,9669	0,9664	0,9662	0,9741	0,9731	0,9740
	BioBERT+BiLSTM+CRF	**0,9797**	0,9780	0,9788	0,9890	0,9884	0,9887
	ClinicalBERT+BiLSTM+CRF	0,9793	**0,9794**	**0,9792**	**0,9894**	**0,9889**	**0,9890**
PSO	BERT	0,8874	0,8751	0,8810	0,9163	0,9168	0,9160
	BioBERT	0,8924	0,8923	0,8920	0,9171	0,9157	0,9170
	ClinicalBERT	0,8952	0,8864	0,8910	0,9008	0,9017	0,9010
	BERT +BiLSTM+CRF	0,9667	0,9652	0,9660	0,9734	0,9724	0,9730
	BioBERT+BiLSTM+CRF	**0,9790**	**0,9785**	**0,9786**	**0,9886**	**0,9881**	**0,9882**
	ClinicalBERT+BiLSTM+CRF	0,9779	0,9774	0,9776	0,9870	0,9872	0,9871

Çizelge 4.3'te sunulan bulgular detaylı olarak incelendiğinde, meta-sezgisel algoritmalarla gerçekleştirilen hiper-parametre optimizasyonunun, modellerin genel başarımı üzerinde anlamlı ve dikkat çekici iyileşmeler sağladığı görülmektedir. Özellikle, BERT+BiLSTM+CRF hibrit mimarisi üzerinde yapılan optimizasyon sonucunda yaklaşık %2 oranında F1-Skoru artışı elde edilmiştir. Bu artış, halihazırda yüksek bir performansa sahip olan hibrit mimaride sağlanan marjinal kazanımları ifade ederken; ana modellerde %9'a kadar ulaşan ciddi performans iyileşmeleri kaydedilmiştir. Bu durum, hibrit modellerin yapısal olarak zaten güçlü bir mimariye sahip olması nedeniyle, optimizasyondan elde edilebilecek ek kazanımların görece sınırlı kaldığını, buna karşılık yalın modellerin hiper-parametre seçimlerine daha yüksek derecede duyarlılık gösterdiğini ortaya koymaktadır. Başka bir ifadeyle, BiLSTM ve

CRF gibi ek katmanlar, modelin sıralı bağımlılıkları öğrenme ve çıktı tutarlılığını sağlama yönünde doğal bir stabilizasyon etkisi sunarken, bu yapıdan yoksun olan yalın Transformer modelleri, ancak doğru hiper-parametre kombinasyonları ile benzer düzeyde başarıya ulaşabilmektedir.

Ayrıca, GA ve PSO algoritmalarının ana model ve hibrit model yapıları üzerinde benzer başarı düzeylerine ulaştığı gözlemlenmiştir. Her iki algoritma da hem arama verimliliği hem de çözüm kalitesi açısından birbirine yakın sonuçlar üretmiş; bu durum, farklı optimizasyon stratejilerinin klinik metin madenciliği bağlamında birbirini tamamlayıcı nitelikte olabileceğini göstermiştir.

Elde edilen bulgular, GA ve PSO algoritmalarının modellerin performansı üzerinde benzer etkilere sahip olduğunu göstermektedir. Ayrıca, hibrit model tasarımında çalışmaya dahil edilen ClinicalBERT ve BioBERT modellerinin, standart BERT modeline göre daha üstün sonuçlar verdiği gözlemlenmiştir. Bu bulgu, alan uzmanlığına yönelik ön eğitim süreçlerinin klinik metin işleme görevlerindeki etkinliğini net bir şekilde ortaya koymaktadır.

4.4 Genel Değerlendirme

Gerçekleştirilen deneysel değerlendirmeler sonucunda, Genetik Algoritma ile optimize edilen ClinicalBERT+BiLSTM+CRF modeli her iki test veri setinde de en yüksek F1 skoruna ulaşmıştır. Çizelge 4.4'de sunulan eğitim süreleri dikkate alındığında, Clinical BERT+ BiLSTM+ CRF ve BioBERT+BiLSTM+CRF modelleri arasında performans açısından istatistiksel olarak anlamlı bir fark gözlemlenmemiştir. Ancak, eğitim süresi bakımından incelendiğinde ClinicalBERT+BiLSTM+CRF modelinin daha kısa sürede eğitildiği ve bu yönüyle daha verimli olduğu tespit edilmiştir. Dolayısıyla, model seçimi yalnızca doğruluk oranlarına değil, aynı zamanda hesaplama süresi gibi pratik kriterlere de dayandırılmış; sonuç olarak ClinicalBERT+BiLSTM+CRF, performans-maliyet dengesi açısından en uygun model olarak belirlenmiştir.

Çizelge 4.4 GA ve PSO ile optimize edilen en başarılı modellerin eğitim sürelerinin karşılaştırılması.

	Meta-sezgisel Algoritmalar	
Modeller	GA eğitim süresi (saniye)	PSO eğitim süresi (saniye)
BioBERT+BiLSTM+CRF	229,95	225,46
ClinicalBERT+BiLSTM+CRF	223,27	223,55

Çizelge 4.4'de görüldüğü üzere, ClinicalBERT+BiLSTM+CRF modeli hem GA hem de PSO optimizasyonlarında daha düşük eğitim süresi ile benzer doğruluk oranları sağlamış, bu nedenle en başarılı model olarak belirlenmiştir. Bu model için belirlenen optimum hiper-parametre değerleri Çizelge 4.5'de verilmiştir.

Çizelge 4.5 ClinicalBERT+BiLSTM+CRF modeli için en optimum hiper-parametreler.

Parametre	**Optimum Değer**
Öğrenme Oranı	8.6484e-05
Maksimum Giriş Uzunluğu	64
Yığın Boyutu	8
Optimizer	ADAM
Epok	27
Alfa Değerleri	
•B-ANAT-DP	1,0816971022756987
•I- ANAT-DP	1,1143041821739001
•B-OBS-DP	1,3233669639699912
•I-OBS-DP	1,3140522266004446
•B-OBS-DA	1,0224642026042052
•I-OBS-DA	1,642436361519199
•B-OBS- U	1,785022815972897
•B-OBS-U	1,763358744489493
•O	1,06943908187

Çizelge 4.5'te sunulan hiperparametre konfigürasyonları incelendiğinde, ClinicalBERT +BiLSTM+CRF modelinin üstün performansında bu yapılandırmaların doğrudan etkili olduğu açıkça görülmektedir. Özellikle, oldukça düşük olarak belirlenen öğrenme oranı

(8.6484e-05), modelin ağırlık güncellemelerini daha yavaş, dikkatli ve dengeli şekilde gerçekleştirmesine imkân tanımış; böylece modelin öğrenme süreci boyunca aşırı öğrenme (overfitting) eğilimi azaltılmıştır. Bu durum, eğitim sürecinde kararlılığın artmasına ve genel doğruluk oranlarının iyileştirilmesine katkı sağlamıştır.

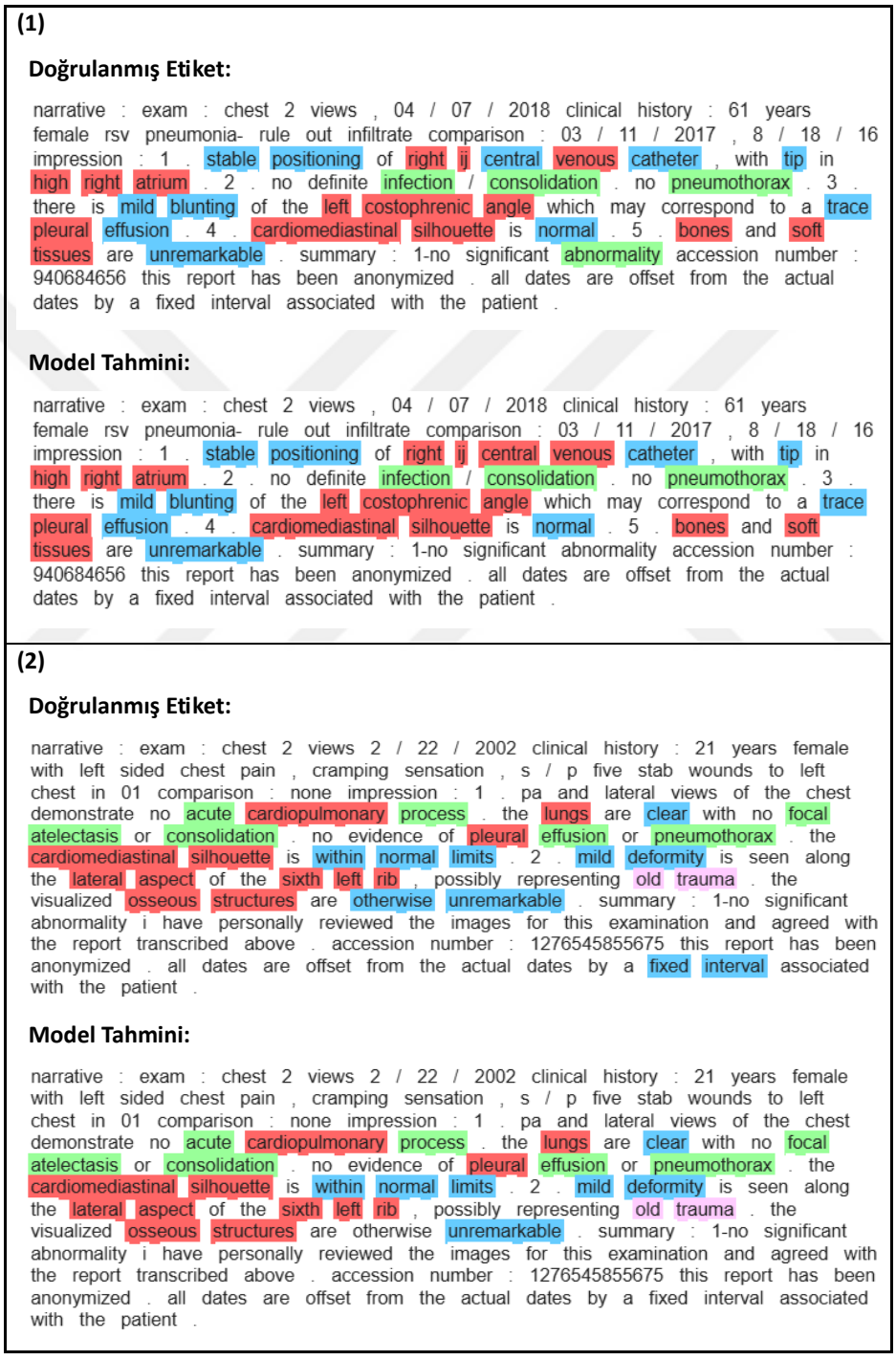
Ayrıca, maksimum giriş uzunluğunun 64 ve yığın boyutunun (batch size) 8 olarak yapılandırılması, modelin hem bağlamsal bilgi taşıma kapasitesini optimize etmiş hem de eğitim sürecinin hesaplama yükünü makul düzeyde tutarak kaynakların verimli kullanılmasına olanak tanımıştır. Epok sayısının 27 olarak seçilmesi, modelin yeterli sayıda yineleme üzerinden öğrenme gerçekleştirmesini sağlayarak hem öğrenme derinliğini hem de genelleme kapasitesini artırmıştır. Öte yandan, ADAM optimizasyon algoritmasının tercih edilmesi, adaptif öğrenme oranı mekanizması sayesinde parametre güncellemelerini daha hızlı ve etkili biçimde gerçekleştirmiş; bu da konverjans sürecinin hızlanmasına ve modelin erken evrelerde istikrarlı bir performans göstermesine katkı sunmuştur.

Şekil 4.1'de ClinicalBERT+BiLSTM+CRF modelinin iki örnek radyoloji raporu üzerinde gerçekleştirdiği varlık ismi tanıma (NER) çıktıları, görsel olarak sunulmuş ve karşılık gelen doğrulanmış etiketlemelerle (ground truth) birlikte karşılaştırmalı biçimde değerlendirilmiştir. Görselleştirmede, model tarafından tanımlanan her bir varlık türü farklı renklerle kodlanmış ve böylece hem model tahminlerinin hem de etiketlerin yorumlanabilirliği artırılmıştır.

Modelde kullanılan alfa ağırlıklandırma katsayıları, özellikle veri setinde düşük sıklıkla temsil edilen varlık türlerinin öğrenme sürecinde göz ardı edilmesini engellemiştir. Bu ağırlıklandırma stratejisi sayesinde, I-OBS-DA ve B-OBS-U gibi belirsiz ve nadir görülen sınıflar için modelin duyarlılığı artırılmış, bu da azınlık sınıflarında dahi yüksek doğruluk oranları elde edilmesine katkı sağlamıştır.

Analiz edilen örneklerde modelin, "right atrium", "sixth left rib" gibi anatomik yapılarla birlikte "pleural effusion", "consolidation", ve "mild deformity" gibi klinik bulguları da başarılı şekilde tanıdığı görülmüştür. Bununla birlikte, "unremarkable" ya da "normal"

gibi normal fizyolojik ifadelerin doğru biçimde etiketlenmesi, modelin yalnızca patolojik varlıkları değil, aynı zamanda normal klinik durumları da yüksek doğrulukla ayırt edebildiğini ortaya koymaktadır. Bu özellik, modelin klinik karar destek sistemlerinde kullanılabilirliğini artıran önemli bir avantaj olarak öne çıkmaktadır.



Şekil 4.1 Önerilen model sonuçları ile doğrulanmış etiketler NER karşılaştırması.

Çizelge 4.6'de çalışmada en iyi performans gösteren ClinicalBERT+BiLSTM+CRF hibrit modeli ile literatürde RadGraph veri setinde yapılan çalışmaların sonuçları sunulmuştur.

Çizelge 4.6 ClinicalBERT+BiLSTM+CRF modeli için en optimum hiper-parametreler.

Çalışma	**Veri Seti**	**Veri Seti Gizliliği**	**Yöntem**	**Sonuç (F1 skor)**
Mou vd. (Mou vd. 2024)	RadGraph	Açık Erişimli	RadLink	0.9580 (MIMIC-CXR and CheXpert)
Liao vd. (Liao vd. 2024)	RadGraph	Açık Erişimli	PURE	0.9662 (MIMIC-CXR) 0.9606 (CheXpert)
Toyoda vd. (Toyoda vd. 2024)	RadGraph	Açık Erişimli	Hierarchical classifier	0.91 (MIMIC-CXR and CheXpert)
Ghosh vd. (Ghosh vd. 2024)	RadGraph	Açık Erişimli	RadLing-MLM	0.98 (MIMIC-CXR) 0.94 (CheXpert)
Dai vd. (Dai vd. 2024)	RadGraph	Açık Erişimli	MaTi + entity type	0.9630 (MIMIC-CXR and CheXpert)
Önerilen Model	**RadGraph**	**Açık Erişimli**	**ClinicalBERT+ BiLSTM+CRF**	**0.9890 (MIMIC-CXR) 0.9792 (CheXpert)**

Çizelge 4.6'te sunulan literatür karşılaştırması, önerilen ClinicalBERT+BiLSTM+CRF modelinin, NER görevinde MIMIC-CXR ve CheXpert veri setleri üzerinde elde ettiği %98,90 ve %97,92 F1 skorlarıyla, literatürdeki en güncel ve başarılı yaklaşımlarla kıyaslandığında üstün performans sergilediğini açıkça ortaya koymaktadır. Bu sonuç, modelin yalnızca yüksek doğruluk sunmakla kalmayıp aynı zamanda iki farklı test ortamında da genelleme yeteneği açısından oldukça sağlam bir yapı sunduğunu göstermektedir.

Mevcut literatür incelendiğinde, Mou ve arkadaşları tarafından geliştirilen RadLink modelinin her iki veri setinden 50 radyoloji raporuna uygulandığı test kümesinde %95,80 F1 skoru elde ettiği görülmektedir (Mou vd. 2024). Liao ve ekibi ise geliştirdikleri PURE modeliyle MIMIC-CXR veri setinde %96,62, CheXpert veri setinde ise %96,06 F1 skoru raporlamış; bu yönüyle önerilen modelin hemen ardından gelen en yüksek başarıyı göstermiştir (Liao vd. 2024). Yine yüksek başarı sunan çalışmalardan biri olan Ghosh ve arkadaşlarının RadLing-MLM modeli, MIMIC-CXR

veri setinde %98,00, CheXpert veri setinde ise %94,00 F1 skoruna ulaşmıştır (Ghosh vd. 2023).

Önerilen model özellikle MIMIC-CXR için rekabetçi bir performans sunmasına rağmen, CheXpert veri seti üzerinde önerilen modelin gerisinde kalmıştır. Buna karşılık, Toyoda ve ekibi tarafından geliştirilen Hierarchical Classifier yaklaşımı, hem MIMIC-CXR hem de CheXpert veri setlerinde yalnızca %91,00 F1 skoru elde edebilmiş ve bu yönüyle önerilen modelin oldukça gerisinde kalmıştır (Toyoda vd. 2024). Dai ve arkadaşları tarafından geliştirilen MaTi + entity type modeli ise %96,30 F1 skoru ile güçlü bir sonuç elde etmiştir; ancak yine de ClinicalBERT+BiLSTM+CRF yapısının ulaştığı seviye ile kıyaslandığında alt düzeyde kaldığı görülmektedir (Dai vd. 2019).

Bu karşılaştırmalar, yalnızca istatistiksel anlamda değil, aynı zamanda potansiyel klinik uygulamalar açısından da önerilen modelin yüksek doğruluk, kararlılık ve güvenilirlik sunabileceğini ortaya koymaktadır. Elde edilen bu başarının temelinde, BiLSTM ve CRF katmanlarının entegrasyonu sayesinde sağlanan dizisel etiketleme tutarlılığı ve bağlamsal bilgi öğreniminin güçlendirilmesi yatmaktadır. Özellikle CRF katmanının kullanımı, etiket geçişlerinin dilsel ve semantik açıdan tutarlı şekilde modellenmesini mümkün kılarak, çıktıların daha anlamlı ve klinik açıdan geçerli hale gelmesini sağlamıştır.

Ek olarak, GA ve PSO gibi meta-sezgisel optimizasyon yöntemleri, yalnızca hiper-parametrelerin el ile belirlenmesinden kaynaklanabilecek öznellikten doğan belirsizlikleri ortadan kaldırmakla kalmamış; aynı zamanda daha iyi genelleme yeteneğine sahip, kararlı ve yüksek doğruluklu modellerin geliştirilmesine olanak tanımıştır. Böylece, modelin hem eğitimi hem de uygulama süreci daha sistematik ve güvenilir bir hale getirilmiştir.

5. TARTIŞMA ve SONUÇ

Klinik karar destek süreçlerinde merkezi bir rol oynayan radyoloji raporları, sağlık çalışanlarının doğru ve zamanında karar verebilmesi açısından kritik bir öneme sahiptir. Ancak bu raporlar genellikle serbest metin formatında ve yüksek düzeyde terminoloji barındıran yapılar olduğundan, özellikle yoğun klinik ortamda çalışan hekim ve sağlık personeli için hızlı ve etkili yorumlama süreci oldukça zorludur. Literatürde, yapılandırılmamış metinlerin otomatik olarak analiz edilememesi, bilgiye erişimi sınırlayarak hem iş yükünü artırdığı hem de tanı ve tedavi süreçlerinde hata riskini büyüttüğü belirtilmektedir. Bu durum, radyoloji raporlarında yer alan tıbbi bilgilerin daha sistematik ve otomatik şekilde çıkarılmasını sağlayacak NLP tabanlı yöntemlerin geliştirilmesini zorunlu hâle getirmiştir.

Bu doğrultuda, çalışmada göğüs radyoloji raporlarından biyomedikal varlıkların otomatik olarak çıkarılmasını hedefleyen bir NER yaklaşımı önerilmiştir. BERT, BioBERT ve ClinicalBERT gibi önceden eğitilmiş Transformer tabanlı dil modelleri temel alınarak bu modellerin sırasıyla BiLSTM ve CRF katmanlarıyla entegre edilmiştir. Bu hibrit yapılar, hem bağlamsal kelime temsillerinin daha derin öğrenilmesini hem de kelimeler arası sıralı bağımlılıkların etkin biçimde modellenmesini mümkün kılmıştır.

"Modelin performansını artırmak amacıyla, az temsil edilen etiketler için alfa ağırlıklandırma stratejisi uygulanarak modelin bu etiketlere karşı duyarlılığı artırılmıştır. Ayrıca, hiper-parametrelerin daha sistematik şekilde optimize edilebilmesi için GA ve PSO gibi meta-sezgisel yöntemler kullanılmıştır. Elde edilen sonuçlar, ClinicalBERT+BiLSTM +CRF mimarisinin her iki test veri setinde de en yüksek F1 skoruna ulaşarak literatürde önerilen yöntemleri geride bıraktığını göstermiştir: MIMIC-CXR veri setinde %98,90, CheXpert test setinde ise %97,92 F1 skoru elde edilmiştir. Bu sonuçlar, modelin hem yüksek doğruluk hem de genelleme kapasitesi açısından güçlü bir başarı sağladığını ortaya koymaktadır.

Literatürde NER görevleri için BERT tabanlı modellerin BiLSTM ve CRF katmanlarıyla birleştirildiği birçok literatür çalışması bulunmaktadır. Ancak, mevcut çalışmaların büyük bir kısmı yalnızca tek bir önceden eğitilmiş modeller üzerinde odaklanmaktadır. Çoğunlukla sabit ya da rastgele değerler ile yapılan hiper-parametre ayarlamaları mevcuttur. Bu bağlamda, çalışmamız literatüre üç önemli katkı sağlamaktadır

1. Birden fazla önceden eğitilmiş klinik modelin karşılaştırmalı analizi yapılmıştır (BERT, BioBERT, ClinicalBERT) ve bu modeller BiLSTM+CRF katmanları ile aynı deneysel çerçevede entegre edilerek test edilmiştir.
2. İlk kez GA ve PSO gibi iki farklı meta-sezgisel algoritma aynı anda uygulanarak tıbbı NER görevinde hiper-parametre optimizasyonlarının model başarımı üzerindeki etkisi derinlemesine karşılaştırılmıştır.
3. Etiket frekansı dengesizliğine karşı alfa ağırlıklandırma stratejisi kullanılarak, az rastlanan klinik varlıkların da etkin biçimde tanınması hedeflenmiş ve bu durum modelin genel doğruluğunu yükseltmiştir.

Önerilen yöntem, benzer nitelikteki klinik metinlerde NER görevlerine yönelik model geliştirme sürecinde araştırmacılar için referans bir yapı sunmakta ve geliştirilecek yeni yaklaşımlar açısından metodolojik bir temel oluşturmaktadır. Ayrıca, farklı dillerde, alanlarda veya veri kaynaklarında çalışacak araştırmalara hem model mimarisi hem de optimizasyon stratejileri açısından önemli bir yönlendirme sağlayarak, sağlık bilişimi alanında daha gelişmiş ve genellenebilir çözümler üretilmesine katkı sunma potansiyeline sahiptir.

Bu tez çalışması yalnızca NER görevine odaklanmakta olup, gelecekte ilişki çıkarımı gibi ileri bilgi çıkarım tekniklerinin entegrasyonu değerlendirilebilir. Belirsizlik içeren etiketlerin diğer kategorilerle karıştırılması, modelin belirsiz ifadeleri doğru şekilde sınıflandırmakta sınırlı kaldığını göstermektedir. Bu durum, çalışmanın önemli bir kısıtı olarak öne çıkmaktadır. İlerleyen araştırmalarda, bu sınırlamaların aşılması adına dikkat mekanizmalarının ve alan uyumlu ön eğitim tekniklerinin kullanılması, modelin belirsiz varlıklara karşı hassasiyetini artırma potansiyeline sahiptir.

6. KAYNAKLAR

Abbas A., Lee M., Shanavas N., Kovatchev V., 2024. Clinical Concept Annotation with Contextual Word Embedding in Active Transfer Learning Environment. Digital Health, 10, Article number 20552076241308987.

Abiodun E. O., Alabdulatif A., Abiodun O. I., Alawida M., Alabdulatif A., Alkhawaldeh R. S., 2021. A Systematic Review of Emerging Feature Selection Optimization Methods for Optimal Text Classification: The Present State and Prospective Opportunities. Neural Computing and Applications, 33(22), 15091–15118.

Alammar J., 2021. Ecco: An Open Source Library for the Explainability of Transformer Language Models. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations, August 2021, pp. 249–257.

Alamro H., Gojobori T., Essack M., Gao X., 2024. BioBBC: A Multi-Feature Model That Enhances the Detection of Biomedical Entities. Scientific Reports, 14(1), 7697.

Alsentzer E., Murphy J. R., Boag W., Weng W. H., Jin D., Naumann T., McDermott M., 2019. Publicly Available Clinical BERT Embeddings. arXiv preprint, arXiv:1904.03323.

Alzubaidi L., Zhang J., Humaidi A. J., Al-Dujaili A., Duan Y., Al-Shamma O. vd., 2021. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. Journal of Big Data, 8, 1–74.

Amari S. I., 1993. Backpropagation and Stochastic Gradient Descent Method. Neurocomputing, 5(4–5), 185–196.

Anandika A., Mishra S. P., 2019. A study on machine learning approaches for named entity recognition. 2019 International Conference on Applied Machine Learning (ICAML), May 2019, pp. 153–159. IEEE.

Antipova K., Horban H., 2024. Positional encoding for transformers. In: Engineering Sciences. Baltija Publishing, Riga, Latvia, DOI: 10.30525/978-9934-26-436-8-1.

Ba J. L., Kiros J. R., Hinton G. E., 2016. Layer normalization. arXiv preprint, arXiv:1607.06450.

Behera L., Kumar S., Patnaik A., 2006. On adaptive learning rate that guarantees convergence in feedforward networks. IEEE Transactions on Neural Networks, 17(5), 1116–1125.

Ben Abdennour G., Gasmi K., Ejbali R., 2024. An optimal model for medical text classification based on adaptive genetic algorithm. Data Science and Engineering, 1–15.

Bengio Y., Simard P., Frasconi P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks, 5(2), 157–166.

Carrodeguas E., Lacson R., Swanson W., Khorasani R., 2019. Use of machine learning to identify follow-up recommendations in radiology reports. Journal of the American College of Radiology, 16(3), 336–343.

Castro S. M., Tseytlin E., Medvedeva O., Mitchell K., Visweswaran S., Bekhuis T., Jacobson R. S., 2017. Automated annotation and classification of BI-RADS assessment from radiology reports. Journal of Biomedical Informatics, 69, 177–187.

Chandra C., Ojima Y., Bendarkar M. V., Mavris D. N., 2024. Aviation-BERT-NER: Named entity recognition for aviation safety reports. Aerospace, 11(11), 890.

Chen P. C., Tsai H., Bhojanapalli S., Chung H. W., Chang Y. W., Ferng C. S., 2021. A simple and effective positional encoding for transformers. arXiv preprint, arXiv:2104.08698.

Chen, M. C., Ball, R. L., Yang, L., Moradzadeh, N., Chapman, B. E., Larson, D. B., ... & Lungren, M. P. (2018). Deep learning to classify radiology free-text reports. Radiology, 286(3), 845-852.

Chinchor N., Marsh E., 1998. Appendix D: MUC-7 information extraction task definition (version 5.1). In: Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax, Virginia, April 29–May 1, 1998.

Dai X., Karimi S., Wan S., 2023. Rethinking the role of entity type in relation classification. In: Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), November 2023, pp. 374–384.

Dai Z., Li Z., Han L., 2021. BoneBERT: A BERT-based automated information extraction system of radiology reports for bone fracture detection and diagnosis. In: Advances in Intelligent Data Analysis XIX: 19th International Symposium on Intelligent Data Analysis, IDA 2021, Porto, Portugal, April 26–28, 2021, Proceedings 19, pp. 263–274. Springer International Publishing.

Dai Z., Wang X., Ni P., Li Y., Li G., Bai X., 2019. Named entity recognition using BERT BiLSTM CRF for Chinese electronic health records. In: 2019 12th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics (CISP-BMEI), October 2019, pp. 1–5. IEEE.

Deb K., 1998. Genetic algorithm in search and optimization: the technique and applications. In: Proceedings of International Workshop on Soft Computing and Intelligent Systems (ISI, Calcutta, India), pp. 58–87.

Derczynski L., 2016. Complementarity, F-score, and NLP evaluation. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), May 2016, pp. 261–266.

Devlin J., Chang M. W., Lee K., Toutanova K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), June 2019, pp. 4171–4186.

Dong X., Chowdhury S., Qian L., Li X., Guan Y., Yang J., Yu Q., 2019. Deep learning for named entity recognition on Chinese electronic medical records: Combining deep transfer learning with multitask bi-directional LSTM RNN. PloS One, 14(5), e0216046.

Eberts M., Ulges A., 2020. Span-based joint entity and relation extraction with transformer pre-training. In: ECAI 2020, pp. 2006–2013. IOS Press.

El Koshiry A. M., Eliwa E. H. I., Omar A., 2024. Improving Arabic spam classification in social media using hyperparameters tuning and particle swarm optimization. Full Length Article, 16, 08–8.

Elman J. L., 1990. Finding structure in time. Cognitive Science, 14(2), 179–211.

Ettaik N., Habib B., 2021. Hyperparameter optimisation in NLP architectures. In: Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning, Vol. 10, No. 0010736600003101.

Fister I. Jr., Yang X. S., Fister I., Brest J., Fister D., 2013. A brief review of nature-inspired algorithms for optimization. arXiv preprint, arXiv:1307.4186.

Franz L., Shrestha Y. R., Paudel B., 2020. A deep learning pipeline for patient diagnosis prediction using electronic health records. arXiv preprint, arXiv:2006.16926.

Fu L., Weng Z., Zhang J., Xie H., Cao Y., 2024. MMBERT: A unified framework for biomedical named entity recognition. Medical & Biological Engineering & Computing, 62(1), 327–341.

Gao W., Zheng X., Zhao S., 2021, Named entity recognition method of Chinese EMR based on BERT-BiLSTM-CRF, Journal of Physics: Conference Series, 1848(1), 012083, IOP Publishing.

Gasmi K., 2022, Improving BERT-based model for medical text classification with an optimization algorithm, International Conference on Computational Collective Intelligence, 21–23 Eylül 2022, Cham, 101–111, Springer International Publishing.

Ghojogh B., Ghodsi A., 2020, Attention mechanism, transformers, BERT, and GPT: tutorial and survey, arXiv preprint, arXiv:2007.00807.

Ghosh R., Karn S. K., Danu M. D., Micu L., Vunikili R., Farri O., 2023, Radling: Towards efficient radiology report understanding, arXiv preprint, arXiv:2306.02492.

Gillioz A., Casas J., Mugellini E., Abou Khaled O., 2020, Overview of the transformer-based models for NLP tasks, 2020 15th Conference on Computer Science and Information Systems (FedCSIS), 6–9 Eylül 2020, Sofia, 179–183, IEEE.

Grancharova M., Dalianis H., 2021, Applying and sharing pre-trained BERT-models for named entity recognition and classification in Swedish electronic patient records, Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa), 31 Mayıs–2 Haziran 2021, Reykjavik, 231–239.

Guo S., Yang W., Han L., Song X., Wang G., 2022, A multi-layer soft lattice based model for Chinese clinical named entity recognition, BMC Medical Informatics and Decision Making, 22(1), 201.

Gupta V., Venkit P. N., Wilson S., Passonneau R. J., 2023, Sociodemographic bias in language models: A survey and forward path, arXiv preprint, arXiv:2306.08158.

Hahn U., Oleynik M., 2020, Medical information extraction in the age of deep learning, Yearbook of Medical Informatics, 29(1), 208–220.

Hinz P., Van de Geer S., 2019, A framework for the construction of upper bounds on the number of affine linear regions of ReLU feed-forward neural networks, IEEE Transactions on Information Theory, 65(11), 7304–7324.

Hochreiter S., Schmidhuber J., 1997, Long short-term memory, Neural Computation, 9(8), 1735–1780.

Holland J. H., 1992, Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence, MIT Press, 211s, Cambridge.

Hossain E., Rana R., Higgins N., Soar J., Barua P. D., Pisani A. R., Turner K., 2023, Natural language processing in electronic health records in relation to healthcare decision-making: A systematic review, Computers in Biology and Medicine, 155, 106649.

Hou J., Saad S., Omar N., 2024, Enhancing traditional Chinese medical named entity recognition with Dyn-Att Net: A dynamic attention approach, PeerJ Computer Science, 10, e2022.

Hu J., Bao R., Lin Y., Zhang H., Xiang Y., 2024, Accurate medical named entity recognition through specialized NLP models, arXiv preprint, arXiv:2412.08255.

Hu Y., Chen Q., Du J., Peng X., Keloth V. K., Zuo X., vd., 2024, Improving large language models for clinical named entity recognition via prompt engineering, Journal of the American Medical Informatics Association, 31(9), 1812–1820.

Huang K., Altosaar J., Ranganath R., 2019, ClinicalBERT: Modeling clinical notes and predicting hospital readmission, arXiv preprint, arXiv:1904.05342.

Huang S. C., Shen L., Lungren M. P., Yeung S., 2021, Gloria: A multimodal global-local representation learning framework for label-efficient medical image recognition, Proceedings of the IEEE/CVF International Conference on Computer Vision, 11–17 Ekim 2021, Montreal, 3942–3951.

Huang Z., Jiang R., Aeron S., Hughes M. C., 2024, Systematic comparison of semi-supervised and self-supervised learning for medical image classification, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 17–21 Haziran 2024, Seattle, 22282–22293.

Huang Z., Xu W., Yu K., 2015, Bidirectional LSTM-CRF models for sequence tagging, arXiv preprint, arXiv:1508.01991.

Huangliang K., Li X., Yin T., Peng B., Zhang H., 2023, Self-adapted positional encoding in the transformer encoder for named entity recognition, International Conference on Artificial Neural Networks, 12–15 Eylül 2023, Cham, 538–549, Springer Nature Switzerland.

Irvin J., Rajpurkar P., Ko M., Yu Y., Ciurea-Ilcus S., Chute C., vd., 2019, CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison, Proceedings of the AAAI Conference on Artificial Intelligence, 27 Ocak – 1 Şubat 2019, Honolulu, 33(1), 590–597.

Jacobs R. A., 1988, Increased rates of convergence through learning rate adaptation, Neural Networks, 1(4), 295–307.

Jain S., Agrawal A., Saporta A., Truong S. Q., Duong D. N., Bui T., vd., 2021, RadGraph: Extracting clinical entities and relations from radiology reports, arXiv preprint, arXiv:2106.14463.

Jaiswal A., Tang L., Ghosh M., Rousseau J. F., Peng Y., Ding Y., 2021, RadBERT-CL: Factually-aware contrastive learning for radiology report classification, Machine Learning for Health, 15–17 Kasım 2021, Virtual Conference, 196–208, PMLR.

Jia J., Liang W., Liang Y., 2023, A review of hybrid and ensemble in deep learning for natural language processing, arXiv preprint, arXiv:2312.05589.

Johnson A. E., Pollard T. J., Berkowitz S. J., Greenbaum N. R., Lungren M. P., Deng C. Y., vd., 2019, MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports, Scientific Data, 6(1), 317.

Johnson A. E., Pollard T. J., Shen L., Lehman L. W. H., Feng M., Ghassemi M., vd., 2016, MIMIC-III, a freely accessible critical care database, Scientific Data, 3(1), 1–9.

Jupin-Delevaux É., Djahnine A., Talbot F., Richard A., Gouttard S., Mansuy A., vd., 2023, BERT-based natural language processing analysis of French CT reports: Application to the measurement of the positivity rate for pulmonary embolism, Research in Diagnostic and Interventional Imaging, 6, 100027.

Kaliyar R. K., 2020, A multi-layer bidirectional transformer encoder for pre-trained word embedding: A survey of BERT, 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 29–31 Ocak 2020, Noida, 336–340, IEEE.

Kamath C. N., Bukhari S. S., Dengel A., 2018, Comparative study between traditional machine learning and deep learning approaches for text classification, Proceedings of the ACM Symposium on Document Engineering 2018, 28–31 Ağustos 2018, Halifax, 1–11.

Kameyama K., 2009, Particle swarm optimization – A survey, IEICE Transactions on Information and Systems, 92(7), 1354–1361.

Kandadi T., Shankarlingam G., 2025, Drawbacks of LSTM algorithm: A case study, SSRN, https://ssrn.com/abstract=5080605, Erişim tarihi: 08.05.2025.

Karnatapu L. K., Annavarapu S. P., Nanduri U. V., 2020, Multi-objective reservoir operating strategies by genetic algorithm and nonlinear programming (GA–NLP) hybrid approach, Journal of The Institution of Engineers (India): Series A, 101(1), 105–115.

Katona M., Farkas R., 2014, Szte-nlp: Clinical text analysis with named entity recognition, Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), 23–24 Ağustos 2014, Dublin, 615–618.

Kaveh M., Mesgari M. S., 2023, Application of meta-heuristic algorithms for training neural networks and deep learning architectures: A comprehensive review, Neural Processing Letters, 55(4), 4519–4622.

Kennedy J., Eberhart R., 1995, Particle swarm optimization, Proceedings of ICNN'95 – International Conference on Neural Networks, 27 Kasım–1 Aralık 1995, Perth, 4, 1942–1948, IEEE.

Kenton J. D. M. W. C., Toutanova L. K., 2019, BERT: Pre-training of deep bidirectional transformers for language understanding, Proceedings of NAACL-HLT, 1(2), 15–20 Haziran 2019, Minneapolis.

Keskar N. S., Mudigere D., Nocedal J., Smelyanskiy M., Tang P. T. P., 2016, On large-batch training for deep learning: Generalization gap and sharp minima, arXiv preprint, arXiv:1609.04836.

Khadhraoui M., Bellaaj H., Ammar M. B., Hamam H., Jmaiel M., 2022, Survey of BERT-base models for scientific text classification: COVID-19 case study, Applied Sciences, 12(6), 2891.

Kingma D. P., Ba J., 2014, Adam: A method for stochastic optimization, arXiv preprint, arXiv:1412.6980.

Kobayashi G., Kuribayashi T., Yokoi S., Inui K., 2021, Incorporating residual and normalization layers into analysis of masked language models, Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 7–11 Kasım 2021, 4547–4568.

Kollapally N. M., Geller J., 2023, Clinical BioBERT hyperparameter optimization using genetic algorithm, arXiv preprint, arXiv:2302.03822.

Komatsuzaki A., 2019, One epoch is all you need, arXiv preprint, arXiv:1906.06669.

Kumar S., Solanki A., 2023, An abstractive text summarization technique using transformer model with self-attention mechanism, Neural Computing and Applications, 35(25), 18603–18622.

Lafferty J., McCallum A., Pereira F., 2001, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, ICML, 1(2), 3, Haziran 2001.

Lample G., Ballesteros M., Subramanian S., Kawakami K., Dyer C., 2016, Neural architectures for named entity recognition, Proceedings of NAACL-HLT 2016, 12–17 Haziran 2016, San Diego, 260–270.

Landolsi M. Y., Hlaoua L., Ben Romdhane L., 2023, Information extraction from electronic medical documents: State of the art and future research directions, Knowledge and Information Systems, 65(2), 463–516.

Lee J., Yoon W., Kim S., Kim D., Kim S., So C. H., Kang J., 2020, BioBERT: A pre-trained biomedical language representation model for biomedical text mining, Bioinformatics, 36(4), 1234–1240.

Li F., Jin Y., Liu W., Rawat B. P. S., Cai P., Yu H., 2019, Fine-tuning bidirectional encoder representations from transformers (BERT)–based models on large-scale electronic health record notes: An empirical study, JMIR Medical Informatics, 7(3), e14830.

Liao Y., Xiang H., Liu H., Spasić I., 2024, Using information extraction to normalize the training data for automatic radiology report generation, IEEE Access, 12, 185116.

López-Úbeda P., Martín-Noguerol T., Luna A., 2024, Automatic classification and prioritisation of actionable BI-RADS categories using natural language processing models, Clinical Radiology, 79(1), e1–e7.

Lu J., Yao J., Zhang J., Zhu X., Xu H., Gao W., vd., 2021, Soft: Softmax-free transformer with linear complexity, Advances in Neural Information Processing Systems, 34, 21297–21309.

Lu W., Li J., Wang J., Qin L., 2021, A CNN-BiLSTM-AM method for stock price prediction, Neural Computing and Applications, 33(10), 4741–4753.

Ma X., Zhang P., Zhang S., Duan N., Hou Y., Zhou M., Song D., 2019, A tensorized transformer for language modeling, Advances in Neural Information Processing Systems, 32.

Magna A. A. R., Allende-Cid H., Taramasco C., Becerra C., Figueroa R. L., 2020, Application of machine learning and word embeddings in the classification of cancer diagnosis using patient anamnesis, IEEE Access, 8, 106198–106213.

Magnini B., Lavelli A., Magnolini S., 2020, Comparing machine learning and deep learning approaches on NLP tasks for the Italian language, Proceedings of the Twelfth Language Resources and Evaluation Conference, 11–16 Mayıs 2020, Marseille, 2110–2119.

Mahdavi S., Liao R., Thrampoulidis C., 2023, Memorization capacity of multi-head attention in transformers, arXiv preprint, arXiv:2306.02010.

Malik M. K., Sarwar S. M., 2016, Named entity recognition system for postpositional languages: Urdu as a case study, International Journal of Advanced Computer Science and Applications, 7(10).

Marafino B. J., Park M., Davies J. M., Thombley R., Luft H. S., Sing D. C., vd., 2018, Validation of prediction models for critical care outcomes using natural language processing of electronic health record data, JAMA Network Open, 1(8), e185097.

Marquez L., Comas P., Giménez J., Catala N., 2005, Semantic role labeling as sequential tagging, Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005), 15–16 Haziran 2005, Ann Arbor, 193–196.

Masters D., Luschi C., 2018, Revisiting small batch training for deep neural networks, arXiv preprint, arXiv:1804.07612.

McInnes B. T., Tang J., Mahendran D., Nguyen M. H., 2024, BioBERT-based deep learning and merged ChemProt-DrugProt for enhanced biomedical relation extraction, arXiv preprint, arXiv:2405.18605.

Meier-Diedrich E., Lyckblad C., Davidge G., Hägglund M., Kharko A., McMillan B., vd., 2025, Impact of patient online record access on documentation: Scoping review, Journal of Medical Internet Research, 27, e64762.

Meystre S. M., Savova G. K., Kipper-Schuler K. C., Hurdle J. F., 2008, Extracting information from textual documents in the electronic health record: A review of recent research, Yearbook of Medical Informatics, 17(1), 128–144.

Mielke S. J., Alyafeai Z., Salesky E., Raffel C., Dey M., Gallé M., vd., 2021, Between words and characters: A brief history of open-vocabulary modeling and tokenization in NLP, arXiv preprint, arXiv:2112.10508.

Mienye I. D., Swart T. G., Obaido G., 2024, Recurrent neural networks: A comprehensive review of architectures, variants, and applications, Information, 15(9), 517.

Miwa M., Bansal M., 2016, End-to-end relation extraction using LSTMs on sequences and tree structures, arXiv preprint, arXiv:1601.00770.

Mohammed A. H., Ali A. H., 2021, Survey of BERT (Bidirectional Encoder Representation Transformer) types, Journal of Physics: Conference Series, 1963(1), 012173, IOP Publishing, Temmuz 2021.

Moritz N., Hori T., Le J., 2020, Streaming automatic speech recognition with the transformer model, ICASSP 2020 – IEEE International Conference on Acoustics, Speech and Signal Processing, 4–8 Mayıs 2020, Barcelona, 6074–6078, IEEE.

Mou Y., Chen H., Lode G. I., Truhn D., Sowe S., Decker S., 2024, RadLink: Linking clinical entities from radiology reports, 2024 2nd International Conference on Foundation and Large Language Models (FLLM), 4–6 Kasım 2024, Beijing, 443–449, IEEE.

Nakamura Y., Hanaoka S., Nomura Y., Nakao T., Miki S., Watadani T., vd., 2021, Automatic detection of actionable radiology reports using bidirectional encoder representations from transformers, BMC Medical Informatics and Decision Making, 21, 1–19.

Nakerst G., Brennan J., Haque M., 2020, Gradient descent with momentum – to accelerate or to super-accelerate?, arXiv preprint, arXiv:2001.06472.

Nguyen E., Theodorakopoulos D., Pathak S., Geerdink J., Vijlbrief O., Van Keulen M., Seifert C., 2020, A hybrid text classification and language generation model for automated summarization of Dutch breast cancer radiology reports, 2020 IEEE Second International Conference on Cognitive Machine Intelligence (CogMI), 26–27 Ekim 2020, Atlanta, 72–81, IEEE.

Nieminen M., 2023, The Transformer Model and Its Impact on the Field of Natural Language Processing, University of Helsinki, Bachelor's Programme in Computer Science, Lisans Tezi, Helsinki.

Nobel J. M., van Geel K., Robben S. G., 2022, Structured reporting in radiology: A systematic review to explore its potential, European Radiology, 1–18.

Ozyurt I. B., 2020, On the effectiveness of small, discriminatively pre-trained language representation models for biomedical text mining, bioRxiv, 2020-05.

Padilla Cuevas J., Reyes-Ortiz J. A., Cuevas-Rasgado A. D., Mora-Gutiérrez R. A., Bravo M., 2024, MédicoBERT: A medical language model for Spanish natural language processing tasks with a question-answering application using hyperparameter optimization, Applied Sciences, 14(16), 7031.

Păiş V., Tufiş D., 2022, Capitalization and punctuation restoration: A survey, Artificial Intelligence Review, 55(3), 1681–1722.

Pakhale K., 2023, Comprehensive overview of named entity recognition: Models, domain-specific applications and challenges, arXiv preprint, arXiv:2309.14084.

Panchendrarajan R., Amaresan A., 2018, Bidirectional LSTM-CRF for named entity recognition, Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation, 1–3 Kasım 2018, Cebu City, Filipinler.

Parameshwari V., Kumar P., Rai S., 2022, Comparison between retrieval time of manual and electronic medical records – A case study, International Journal of Case Studies in Business, IT, and Education (IJCSBE), 6(2), 1–14.

Patil R., Boit S., Gudivada V., Nandigam J., 2023, A survey of text representation and embedding techniques in NLP, IEEE Access, 11, 36120–36146.

Peng Q., Luo X., Shen H., Huang Z., Chen M., 2023, Collaborative optimization with PSO for named entity recognition-based applications, Intelligent Data Analysis, 27(1), 103–120.

Pinto A., Gonçalo Oliveira H., Oliveira Alves A., 2016, Comparing the performance of different NLP toolkits in formal and social media text, 5th Symposium on Languages, Poli R., 2007, An Analysis of Publications on Particle Swarm Optimisation Applications, Technical Report CSM-469, Department of Computer Science, University of Essex, ISSN: 1744-8050, Essex, Birleşik Krallık.

Poli R., 2007, An analysis of publications on particle swarm optimization applications, Department of Computer Science, University of Essex, Rapor, Essex, Birleşik Krallık.

Preethi M. M., Bhoomadevi A., Amutha A., 2021, Electronic medical records (EMR) over manual documentation of in-patient records: A scientific insight, Turkish Journal of Computer and Mathematics Education, 12(11), 3274–3285.

Qian N., 1999, On the momentum term in gradient descent learning algorithms, Neural Networks, 12(1), 145–151.

Rabiner L. R., 1989, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE, 77(2), 257–286.

Rahman M. H., Islam M. S., Jowel M. M. U., Hasan M. M., Latif M. S., 2021, Classification of book review sentiment in Bangla language using NLP, machine learning and LSTM, 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 6–8 Temmuz 2021, Kharagpur, Hindistan, 1–5, IEEE.

Rezaeenour J., Ahmadi M., Jelodar H., Shahrooei R., 2023, Systematic review of content analysis algorithms based on deep neural networks, Multimedia Tools and Applications, 82(12), 17879–17903.

Rooker T., 1991, Review of genetic algorithms in search, optimization, and machine learning, AI Magazine, 12(1), 102.

Rozova V., Khanina A., Teng J. C., Teh J. S., Worth L. J., Slavin M. A., vd., 2023, Detecting evidence of invasive fungal infections in cytology and histopathology reports enriched with concept-level annotations, Journal of Biomedical Informatics, 139, 104293.

Ruas P., Lamurias A., Couto F. M., 2020, LasigeBioTM Team at CLEF2020 ChEMU Evaluation Lab: Named Entity Recognition and Event Extraction from Chemical Reactions Described in Patents Using BioBERT NER and RE, CLEF 2020, Working Notes of Conference and Labs of the Evaluation Forum, 22–25 Eylül 2020, Thessaloniki, Yunanistan.

Schuster M., Paliwal K. K., 1997, Bidirectional recurrent neural networks, IEEE Transactions on Signal Processing, 45(11), 2673–2681.

Shapiro J., 1999, Genetic algorithms in machine learning, In Advanced Course on Artificial Intelligence (146–168), Springer Berlin Heidelberg, Berlin.

Sheikhalishahi S., Miotto R., Dudley J. T., Lavelli A., Rinaldi F., Osmani V., 2019, Natural language processing of clinical notes on chronic diseases: Systematic review, JMIR Medical Informatics, 7(2), e12239.

Shen K., Kejriwal M., 2023, An experimental study measuring the generalization of fine-tuned language representation models across commonsense reasoning benchmarks, Expert Systems, 40(5), e13243.

Shrestha A., Mahmood A., 2019, Review of deep learning algorithms and architectures, IEEE Access, 7, 53040–53065.

Shwartz-Ziv R., Goldblum M., Li Y., Bruss C. B., Wilson A. G., 2023, Simplifying neural network training under class imbalance, Advances in Neural Information Processing Systems, 36, 35218–35245.

Singh S., 2018, Natural language processing for information extraction, arXiv preprint, arXiv:1807.02383.

Smith L. N., 2017, Cyclical learning rates for training neural networks, 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), 24–31 Mart 2017, Santa Rosa, 464–472, IEEE.

Smith S. L., Kindermans P. J., Ying C., Le Q. V., 2017, Don't decay the learning rate, increase the batch size, arXiv preprint, arXiv:1711.00489.

Sohn G., Zhang N., Olukotun K., 2024, Implementing and optimizing the scaled dot-product attention on streaming dataflow, arXiv preprint, arXiv:2404.16629.

Sonkar S., Baraniuk R. G., 2023, Investigating the role of feed-forward networks in transformers using parallel attention and feed-forward net design, arXiv preprint, arXiv:2305.13297.

Soomro P. D., Kumar S., Banbhrani A. A. S., Shaikh A. A., Raj H., 2017, Bio-NER: Biomedical named entity recognition using rule-based and statistical learners, International Journal of Advanced Computer Science and Applications, 8, 163–170.

Srivastava P., Bej S., Yordanova K., Wolkenhauer O., 2021, Self-attention-based models for the extraction of molecular interactions from biological texts, Biomolecules, 11(11), 1591.

Suárez-Paniagua V., Dong H., Casey A., 2021, A Multi-BERT Hybrid System for Named Entity Recognition in Spanish Radiology Reports, CLEF 2021 – Conference and Labs of the Evaluation Forum, 2021 Working Notes of CLEF (CLEF-WN 2021), 21–24 Eylül 2021, Virtual, Bükreş, CEUR Workshop Proceedings, 2936, 846–856, CEUR-WS, ISSN: 1613-0073.

Sushil M., Kennedy V. E., Mandair D., Miao B. Y., Zack T., Butte A. J., 2024, CORAL: Expert-curated oncology reports to advance language model inference, NEJM AI, 1(4), AIdbp2300110.

Talbi E. G., 2009, Metaheuristics: From Design to Implementation, John Wiley & Sons, 584s, New Jersey.

Tang G., 2022, Named entity recognition in Chinese electronic medical records based on ALBERT-IDCNN-CRF, 2022 IEEE 8th International Conference on

Computer and Communications (ICCC), 9–12 Aralık 2022, Chengdu, Çin, 1753–1757, IEEE.

Tarwani K. M., Edem S., 2017, Survey on recurrent neural network in natural language processing, International Journal of Engineering Trends and Technology, 48(6), 301–304.

Tieleman T., Hinton G., 2012, Lecture 6.5 – RMSprop: Divide the gradient by a running average of its recent magnitude, University of Toronto, Teknik Rapor, Toronto.

Torfi A., Shirvani R. A., Keneshloo Y., Tavaf N., Fox E. A., 2020, Natural language processing advancements by deep learning: A survey, arXiv preprint, arXiv:2003.01200.

Toyoda G., Rojas Y., Colonna J. G., Gama J., 2024, A hierarchical approach for extracting and displaying entities and relations from radiology medical reports, Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS), 12–14 Haziran 2024, Florianópolis, Brezilya, 154–165, SBC.

Trivedi H., Mesterhazy J., Laguna B., Vu T., Sohn J. H., 2018, Automatic determination of the need for intravenous contrast in musculoskeletal MRI examinations using IBM Watson's natural language processing algorithm, Journal of Digital Imaging, 31, 245–251.

Turchin A., Masharsky S., Zitnik M., 2023, Comparison of BERT implementations for natural language processing of narrative medical documents, Informatics in Medicine Unlocked, 36, 101139.

Vakili T., Henriksson A., Dalianis H., 2024, End-to-end pseudonymization of fine-tuned clinical BERT models: Privacy preservation with maintained data utility, BMC Medical Informatics and Decision Making, 24(1), 162.

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., vd., 2017, Attention is all you need, Advances in Neural Information Processing Systems, 30.

Wang Y., Sohn S., Liu S., Shen F., Wang L., Atkinson E. J., vd., 2019, A clinical text classification paradigm using weak supervision and deep representation, BMC Medical Informatics and Decision Making, 19, 1–13.

Wang Y., Wang L., Rastegar-Mojarad M., Moon S., Shen F., Afzal N., vd., 2018, Clinical information extraction applications: A literature review, Journal of Biomedical Informatics, 77, 34–49.

Wierzbicki M. P., Jantos B. A., Tomaszewski M., 2024, A review of approaches to standardizing medical descriptions for clinical entity recognition: Implications for artificial intelligence implementation, Applied Sciences, 14(21), 9903.

Xiong R., Yang Y., He D., Zheng K., Zheng S., Xing C., vd., 2020, On layer normalization in the transformer architecture, International Conference on Machine Learning (ICML), 13–18 Kasım 2020, Virtual, 10524–10533, PMLR.

Xu G., Meng Y., Qiu X., Yu Z., Wu X., 2019, Sentiment analysis of comment texts based on BiLSTM, IEEE Access, 7, 51522–51532.

Xue B., Zhang M., Browne W. N., Yao X., 2015, A survey on evolutionary computation approaches to feature selection, IEEE Transactions on Evolutionary Computation, 20(4), 606–626.

Xuefeng P., Yuanyuan C., Xiaorui H., Wei S., 2022, Named entity recognition of TCM electronic medical records based on the ALBERT-BiLSTM-CRF model, 2022 12th International Conference on Information Technology in Medicine and Education (ITME), 11–13 Kasım 2022, Fuzhou, Çin, 575–582, IEEE

Yadav V., Bethard S., 2019, A survey on recent advances in named entity recognition from deep learning models, arXiv preprint, arXiv:1910.11470.

Yan A., McAuley J., Lu X., Du J., Chang E. Y., Gentili A., Hsu C. N., 2022, RadBERT: Adapting transformer-based language models to radiology, Radiology: Artificial Intelligence, 4(4), e210258.

Yang L., Huang X., Wang J., Yang X., Ding L., Li Z., Li J., 2023, Identifying stroke-related quantified evidence from electronic health records in real-world studies, Artificial Intelligence in Medicine, 140, 102552.

Yang X. S., 2010, Engineering Optimization: An Introduction with Metaheuristic Applications, John Wiley & Sons, 327s, Hoboken.

Yogarajan V., Montiel J., Smith T., Pfahringer B., 2021, Transformers for multi-label classification of medical text: An empirical comparison, International Conference on Artificial Intelligence in Medicine, 15–18 Haziran 2021, Cham, 114–123, Springer International Publishing.

Yoon D., Han C., Kim D. W., Kim S., Bae S., Ryu J. A., Choi Y., 2024, Redefining health care data interoperability: Empirical exploration of large language models in information exchange, Journal of Medical Internet Research, 26, e56614.

Yu X., Hu W., Lu S., Sun X., Yuan Z., 2019, BioBERT-based named entity recognition in electronic medical record, 2019 10th International Conference on Information Technology in Medicine and Education (ITME), 23–25 Ağustos 2019, Qingdao, Çin, 49–52, IEEE.

Yuan J., Liao H., Luo R., Luo J., 2019, Automatic radiology report generation based on multi-view image fusion and medical concept enrichment, Medical Image Computing and Computer Assisted Intervention – MICCAI 2019: 22nd International Conference, 13–17 Ekim 2019, Shenzhen, Çin, Part VI, 721–729, Springer International Publishing.

Zhang Z., Gong Z., Hong Q., 2021, A survey on: Application of transformer in computer vision, Proceedings of the 8th International Conference on Intelligent Systems and Image Processing, 21–24 Eylül 2021, Matsue, Japonya, 21–28.

Zhao J., Huang F., Lv J., Duan Y., Qin Z., Li G., Tian G., 2020, Do RNN and LSTM have long memory? International Conference on Machine Learning (ICML), 12–18 Temmuz 2020, Virtual, 11365–11375, PMLR.

İnternet Kaynakları

1. https://github.com/MIT-LCP/mimic-cxr, Erişim Tarihi: 09.05.2025

2. https://stanfordmlgroup.github.io/competitions/chexpert/, Erişim Tarihi: 09.05.2025

3. https://physionet.org/content/corpus-fungal-infections/1.0.2/, Erişim Tarihi: 09.05.2025
4. https://physionet.org/content/radgraph/1.0.0/, Erişim Tarihi: 09.05.2025
5. https://physionet.org/content/curated-oncology-reports/1.0/, Erişim Tarihi: 09.05.2025