



# Öz Dikkat Mekanizması Tabanlı Görü Dönüştürücü Kullanılarak Sıtma Parazit Tespiti

## Malaria Parasite Detection Using Self-Attention Mechanism Based Vision Transformer

İbrahim Tuncel<sup>1\*</sup>, Abdülkadir Albayrak<sup>2</sup>, Mehmet Akın<sup>3</sup>

<sup>1</sup> Dicle Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, [eamibrhmtnc1@gmail.com](mailto:eamibrhmtnc1@gmail.com)  
ORCID: <https://orcid.org/0000-0002-4965-5134>

<sup>2</sup> Dicle Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, [kadir.albayrak@dicle.edu.tr](mailto:kadir.albayrak@dicle.edu.tr)  
ORCID: <https://orcid.org/0000-0002-0738-871X>

<sup>3</sup> Dicle Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, [makin@dicle.edu.tr](mailto:makin@dicle.edu.tr)  
ORCID: <https://orcid.org/0000-0001-5439-4824>

### MAKALE BİLGİLERİ

#### Makale Geçmişi:

Geliş 26 Mayıs 2022  
Revizyon 25 Haziran 2022  
Kabul 25 Haziran 2022  
Online 28 Haziran 2022

#### Anahtar Kelimeler:

Sıtma Teşhisi, bilgisayar destekli teşhis, vision transformer, derin öğrenme

### ÖZ

Sıtma, tedavisiz olgularda ölümlerle sonuçlanabilen ve ciddi ateşli hastalığa yol açan bir enfeksiyon hastalığıdır. Bu yüzden bu hastalığın erken tanı ve tedavisi oldukça kritik öneme sahiptir. Gelişmiş teknolojilerle birlikte sıtma hastalığının teşhisine yönelik birçok klinik yöntem ve test kullanılmaktadır. Bu çalışmada Sıtma hastalığının teşhis edilmesi amacıyla son yıllarda doğal dil işleme alanında oldukça yüksek performans gösteren transformer yöntemlerinden esinlenilerek önerilen Vision Transformer (ViT) yöntemi kullanılmaktadır. Elde edilen sonuçlar değerlendirildiğinde ViT yönteminin %97.22 gibi yüksek bir sınıflandırma performansı elde ettiği gözlemlenmiştir. ViT yöntemi ile elde edilen sonuçlar, literatürde uygulanan geleneksel ve derin öğrenme yöntemleri karşılaştırılmış ve bu sonuçlar karşılaştırmalı olarak tabloda sunulmuştur. Uygulanan ViT modelinin sıtma hastalığı tespitinde başarılı sonuç verdiği gözlemlenmiştir.

### ARTICLE INFO

#### Article history:

Received 26 May 2022  
Received in revised form 25 June 2022  
Accepted 25 June 2022  
Available online 28 June 2022

#### Keywords:

Malaria Diagnosis, computer aided diagnosis, vision transformer, deep learning.

Doi: 10.24012/dumf.1120289

\* Sorumlu Yazar

### ABSTRACT

Your Malaria is an infectious disease that can result in death in untreated cases and causes serious febrile illness. Therefore, early diagnosis and treatment of this disease is of critical importance. Along with advanced technologies, many clinical methods and tests are used for the diagnosis of malaria. In this study, the Vision Transformer (ViT) method, which has been inspired by the transformer methods that have shown very high performance in the field of natural language processing in recent years, is used to diagnose Malaria. When the obtained results were evaluated, it was observed that the ViT method achieved a high classification performance of 97.22%. The results obtained with the ViT method, traditional computer vision algorithms and deep learning methods applied in the literature were compared and the results were presented in a comparative table. It has been observed that the applied ViT model gives successful results in the detection of malaria.

## Giriş

Sıtma, Plasmodium adı verilen parazitlerin, tedavisiz olgularda ölümlerle sonuçlanabilen ve ciddi ateşli hastalığa yol açan bir enfeksiyon hastalığıdır. Parazit, öncelikle insanlara en çok dişi anofel sivrisineklerin ısırıkları vasıtasıyla ulaşır oksijen taşıyan kırmızı kan hücrelerini etkileyecektir. Sıtma ile enfekte kişilerde, üşüme, titreme, ateş yükselmesi ve terleme döngüsü şeklinde görülen sıtma nöbetleri hastalığın karakteristik özelliğidir. Sıtma, dünyada çok büyük bir nüfusu tehdit eden bir hastalık olarak ılıman iklimlerde çok nadir rastlanırken, tropikal ülkelerde sıtma yaygın olarak

görülmektedir. Dünya Sağlık Örgütü (DSÖ) verilerine göre dünya nüfusunun yarısı sıtma riski altında olup bu hastalık nedeniyle çoğu Afrika ve Uzak Asya ülkelerinde olmak üzere yılda yaklaşık 1 milyon kişi hayatını kaybetmektedir. Son yıllarda DSÖ, ölüm sayısının düştüğünü tahmin etse de sıtma dünyada halen önemli bir enfeksiyon hastalığı olmaya devam etmektedir. Dolayısıyla sıtma hastalığının ciddiyetini kavramak ve hastalığı önlemek için erken tanı testi ve tedavisi hayati öneme sahiptir. Sıtma teşhisi için klinik boyutlu birçok yöntem ve test kullanılmaktadır. Sıtmanın tanısında mikroskopik inceleme, hızlı tanı testleri, seroloji ve

moleküler yöntemler (PCR) kullanılabilir. Bunlar arasında en yaygın ve referans mikroskopi tanı yöntemi kullanılmaktadır. Bu yöntemde elde edilen periferik kan yayma görüntüleri, sıtma parazitlerinin etkin görsel tanımlamasını gerçekleştirmek için Giemsa lekeleme adı verilen popüler bir leke kullanılarak boyanır. Ancak yapılan bu test ve incelemeler uzmanlık gerektiren manuel işlemler çok zaman alıcıdır, insan hataları nedeniyle hatalı raporlar üretebilir ve kapsamlı teşhisler için de zahmetlidir. Bu, sıtma tespitini hızlı, kolay ve etkili erken teşhise yol açan bir yöntem kullanarak sıtmanın otomatik olarak saptanmasını önermemiz için bizi teşvik ediyor. Bu noktada, bilgisayar destekli sistemler, yapay zeka, açık kaynaklı araçlar, teknikler bu hastalığın tespiti ve değerlendirilmesi için uzmanlar için eşsiz bir araç haline gelmiştir. Bununla birlikte hem daha az maliyetli hem de daha nesnel sonuçlar ortaya koymaktadır. Sıtma hastalığının sınıflandırılmasında ve tespitinde literatürde birçok yöntem kullanılmıştır. Bunlar; evrişimsel sinir ağları temelli derin öğrenme algoritmaları, transfer öğrenme, derin inanç ağı (DBN), adaptif eşikleme ve morfolojik görüntü işleme algoritmaları, gerçek zamanlı nesne algılama algoritmaları (YOLOV3 ve YOLOV4),Yapay Sinir Ağları (YSA), Destek Vektör Makine (DVM), k- En Yakın Komşuluk algoritması (k-Ek), AdaBoost algoritması, Otsu metodu, Çok Katmanlı Algılayıcı , Radyal tabanlı fonksiyon sınıflandırıcı ve Evrişimsel sinir ağları(ESA) ve birçok alanda kullanılan derin öğrenme algoritmalarıdır. **Vijayalakshmi ve arkadaşları**, yaptıkları çalışmada transfer öğrenme kullanarak enfekte olmuş falciparum sıtma parazitinin tanımlanması için yeni bir derin sinir ağı modeli önermişlerdir. Önerilen transfer öğrenme yapısı VGG ağı ve Destek Vektör Makinesi (DVM) birleştirilerek elde edilmiştir. Dijital sıtma görüntüleri, VGG19-DVM'nin performansını analiz etmek için kullanılmıştır, bu da enfekte olmuş falciparum sıtma parazitinin tanımlanmasında % 93.1 oranında sınıflandırma doğruluğu ile sonuçlanmıştır [1]. **Rajaraman ve arkadaşları**, çalışmalarında hastalık taramasının iyileştirilmesine yardımcı olmak için parazitli ve parazit bulaşmamış hücrelerin sınıflandırılmasında önceden eğitilmiş evrişimsel sinir ağı (ESA) temelli derin öğrenme (DÖ) modellerini özellik çıkarıcı olarak kullanmışlardır. Özellik çıkarımı için en uygun model katmanlarını belirlemeye çalışmışlardır [2]. **Dong ve çalışma arkadaşları**, yaptıkları çalışma ile sıtma bulaşmış hücrelerin derin öğrenme yöntemleri kullanılarak otomatik olarak tanımlanmasını başarmışlardır. Çalışmalarında LeNet, AlexNet ve GoogLeNet olmak üzere en yaygın evrişimsel sinir ağlarını kullanmışlardır. Tüm ESA'ların, DVM yöntemi kullanılarak % 95'in üzerinde sınıflandırma doğruluğunu sağladığı gözlemlenmiştir [3]. **Bibin ve arkadaşları**, derin inanç ağı (DBN) kullanarak periferik kan yayma görüntülerinde sıtma parazitlerinin tespiti için yeni bir yöntem önermişlerdir. Önerilen yöntemde, parazit veya parazit olmayan toplam 4100 periferik kan yayma görüntüsünü sınıflandırmak için DBN'ye dayanan eğitilmiş bir model sunulmaktadır. Tavsiye edilen yöntem,% 89.66'lık bir F-skoru,% 97.60'lık bir duyarlılık ve % 95.92 özgüllüğü ile diğer modern yöntemlerden önemli ölçüde daha iyi bir performans göstermiştir [4]. **Liang ve arkadaşları**, yaptıkları çalışmada, standart mikroskop lamları üzerindeki enfekte veya enfekte olmamış ince kan

yayma görüntülerinde tek hücreleri otomatik olarak sınıflandırmak için ESA temelli yeni bir makine öğrenme modeli önermektedir. 27.578 tek hücreli görüntüye dayanan on kat çapraz doğrulamada, yeni 16 katmanlı ESA modelinin ortalama doğruluğu % 97.37'dir [5]. **Hung ve arkadaşları**, yaptıkları çalışmada sıtma paraziti bulaşmış hücreleri tespit etmek ve sınıflandırmak için iki aşamalı yeni bir derin öğrenme modeli önermişlerdir. Hücreleri tespit etmek için daha hızlı bölge bazlı-ESA ve sınıflandırma için AlexNet ağlarını kullanmışlardır. Önerilen yöntemin doğruluk oranı % 98 elde edilmiştir [6]. **Narayanan ve arkadaşları**, yaptıkları çalışmada sıtma paraziti bulaşmış hücreleri tespit etmek ve sınıflandırmak için yeni bir ESA mimarisi modeli önermişlerdir. Bu kullanılan yöntemde %96,7'lik bir genel doğruluk sağlamıştır [7]. **Muneer ve arkadaşları ise**, çalışmalarında sıtma paraziti bulaşmış hücreleri tespit etmek için görüntü işleme teknikleri ve sınıflandırma için ise karar ağacı kullanarak yeni model önermişlerdir. Önerilen yöntemde %88'lik bir genel doğruluk sağlamıştır [8].

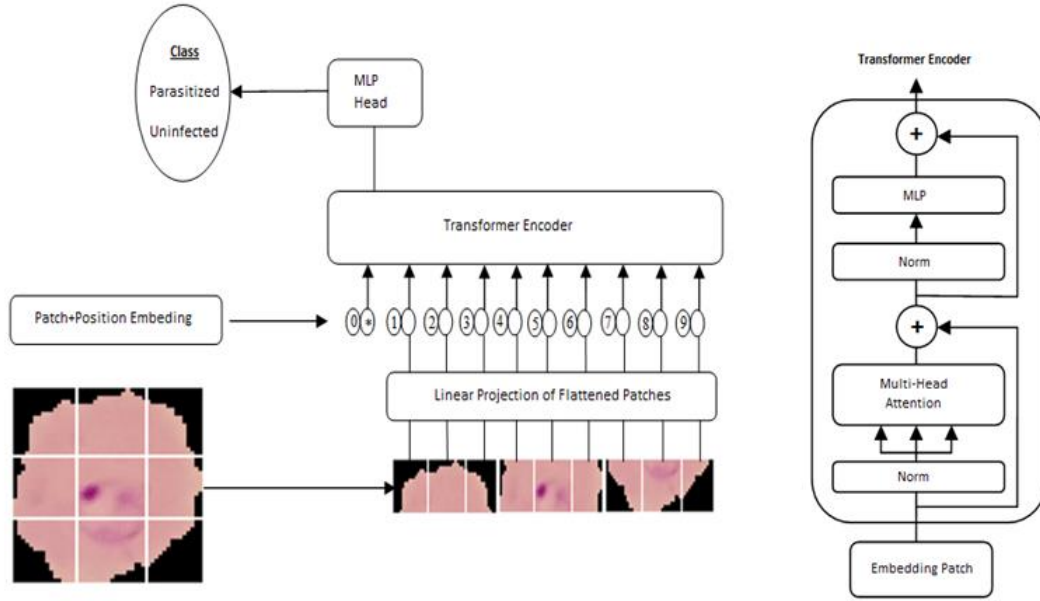
Bu çalışmada, sıtma hastalığının teşhisi ve sınıflandırılması için Vision Transformer (ViT) modeli kullanılmıştır. Amaç, kan hücresi görüntülerini "Parazit" veya "Enfekte olmayan" olarak sınıflandırmaktır. Yapılan çalışmada, Lister Hill Ulusal Biyomedikal İletişim Merkezi (LHNCBC) Ulusal Tıp Kütüphanesinden alınan sıtma veri seti kullanılmıştır. Veri seti 13779'u sıtma virüsü taşıyan ve 13779'u ise sağlıklı kırmızı kan hücre görüntüsü içermektedir [9]. Veri seti önerilen modelimize uygulandığında güncel metotlarla karşılaştırılabilir bir performans göstermiştir.

## Materyal ve Metod

### Vision Transformer (ViT) Modeli

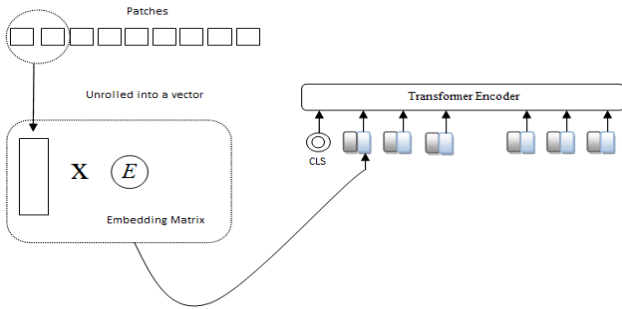
Transformer, giriş verilerinin her bir bölümünün önemini farklı şekilde ağırlıklandıran, kendi kendine dikkat mekanizmasını benimseyen bir derin öğrenme modelidir. Öncelikle doğal dil işleme (NLP) ve bilgisayarla görme alanlarında kullanılmaktadır. Transformatörler, çeviri ve metin özetleme gibi görevler için doğal dil gibi sıralı giriş verilerini işlemek üzere tasarlanmıştır [10]. Ancak, transformatörler verileri mutlaka sırayla işlemez. Bunun yerine, dikkat mekanizması girdi dizisindeki herhangi bir konum için bağlam sağlar. NLP'deki Transformer ölçekleme başarılarından esinlenerek, mümkün olan en az değişiklikle standart bir Transformer'ı doğrudan görüntülere uygulamayı denedik. Bunu yapmak için, bir görüntüyü yamalar halinde böleriz ve bir Transformer'a bir girdi olarak bu yamaları doğrusal gömme sırasını sağlarız. Görüntü yamaları, bir NLP uygulamasında belirteçler (kelimeler) ile aynı şekilde ele alınır. Modeli denetimli bir şekilde görüntü sınıflandırma konusunda eğitiyoruz.

Modele genel bir bakış Şekil 1'de gösterilmektedir. Standart bir Transformer, giriş olarak 1D jeton yerleştirme dizisini alır. Bu nedenle, bir 2D görüntü için, görüntüyü bir dizi düzleştirilmiş 2D yamalar halinde yeniden şekillendirmemiz gerekir. Görüntü boyutları, yama boyutuna göre bölünebilir olmalıdır. Piksel eksikliğinin üstesinden gelmek için dolgu kullanılabilir. 1D'ye düzleştirmek hayati önem taşır.



Şekil 1. Önerilen ViT modelinin genel işlem adımları

Bunu çözmek için, Şekil 2'de görüntü yamaları, öğrenilmiş bir gömme matrisi  $E$  kullanılarak bir vektöre doğrusal olarak yansıtılır. Görüntü yamalarının konum bilgilerini korumak için transformatör kodlayıcıya geçilir. Bunu yapmak için, görüntünün sıra numarası kodlayıcıya aktarılır. Konumsal gömme vektörlerinin her biri parametreleştirilmiştir ve öğrenilebilir bir konumsal gömme tablosu oluşturacaktır. Gömme yama dizisi elde edildikten sonra, bunlar transformatör kodlayıcıya geçirilir [11]. Bir transformatör kodlayıcı,  $L$  özdeş katmanlardan oluşur ve her katmanda, çok başlı öz-dikkat (MSA) bloğu ve çok katmanlı algılayıcı (MLP) bloğu olan iki ana bileşene sahiptir (Denklem 2,3). Katman normalizasyonu, her bloktan önce ve her bloktan sonra kalan bağlantılara birlikte uygulanır. [13].



Şekil 2. ViT modelinin bir aşaması olan kesit kodlayıcı aşaması

Transformer, tüm katmanları boyunca sabit gizli vektör boyutu  $D$  kullanır, bu yüzden yamaları düzleştiririz ve eğitilebilir bir doğrusal izdüşümle (Denklem 1)  $D$  boyutlarına eşler. Bu izdüşümün çıktısını yama yerleştirmeleri olarak adlandırılır.

BERT' in [sınıf] belirtecine benzer şekilde, Transformer kodlayıcısının  $(z_L^0)$  çıkışındaki durumu  $y$  görüntü temsili olarak hizmet eden gömülü yamalar dizisine ( $z_0^0 = x_{class}$ ) öğrenilebilir bir yerleştirme hazırlıyoruz (Denklem 4). Hem

ön eğitim hem de ince ayar sırasında,  $(z_L^0)$ 'ye bir sınıflandırma başlığı eklenir. Sınıflandırma başlığı, ön eğitim zamanında bir gizli katmana ve ince ayar zamanında tek bir doğrusal katmana sahip bir MLP tarafından uygulanır [11].

MLP, arada bir GeLU doğrusal olmayan aktivasyon ile iki tam bağlı katmandan oluşur.

$$z_0 = [x_{class}; x_p^1 E; x_p^2 E; \dots; x_p^N E] + E_{pos},$$

$$E \in R^{(P^2 \cdot C) \times D}, E_{pos} \in R^{(N+1) \times D} \quad (1)$$

$$z'_l = MSA(LN(z_{l-1})) + z_{l-1}, \quad l = 1 \dots L \quad (2)$$

$$z_l = MLP(LN(z'_l)) + z'_l, \quad l = 1 \dots L \quad (3)$$

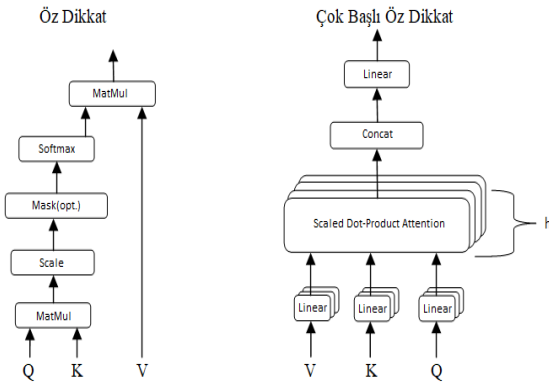
$$y_l = LN(z_L^0) \quad (4)$$

Transformer modelinin ortaya çıkmasında en etkili kısım dikkat mekanizmasıdır. Dikkat mekanizması, girdi dizisine bakar ve dizinin her adımda önemini koruyan bölümlerini seçerek, dizinin hangi bölümlerinin önemli olduğu bilgisini korur. Dikkat mekanizması, anlık olarak birkaç diğer girdi verisini de hesaba katar ve bu girdilere farklı ağırlıklar atayarak önceliklendirme sağlar. Tüm kodlayıcı katmanları, her bir girdi için diğer tüm girdilerin uygunluğunu ölçer ve çıktıyı üretmek için uygun bilgileri alan bir dikkat mekanizması kullanır. Ardından dikkat mekanizması sonucunda çıktı olarak gönderilen ağırlıkları ve kodlanmış

şekilde olan diziyi girdi olarak alır. Bu ağlar, tekrarlanan çok başlı dikkat bloklarından ve ileri beslemeli katmanlardan oluşur. Çok başlı dikkat, dikkat mekanizmasındaki işlemleri paralel olarak çalıştırır ve sonuçları birleştirir. Böylece, farklı ilişkilerin öğrenilmesi sağlanmış olur [12]. Girdiye ait özelliklerden doğrusal dönüşümler ile sorgu (Q), anahtar (K) ve değer (V) matrisi oluşturulur [14]. Çıktı, sorgu, anahtar matrislerinin nokta çarpımının anahtar vektörün boyutunun köküne bölünmesiyle hesaplanır daha sonra Softmax işlevi uygulandıktan sonra değer matrisi ile çarpılır ve dikkat hesaplanır (Denklem 4). Bu hesaplama aşağıdaki denkleme dayanmaktadır. Bu nokta çarpımında kullanılan Q, K ve V, hepsi aynı verilerden türetildiklerinde Öz-Dikkat olarak adlandırılır [12].

$$Dikkat(Q, K, V) = softmax\left(\frac{Q^T K}{\sqrt{d_k}}\right)V \quad (5)$$

Çok başlı öz dikkat (MSA), girdiyi h küçük parçalara böler, girdilere paralel olarak birkaç kez dikkat uygular. Bunun nedeni, Q, K ve V'yi birden çok kez yansıtmanın sisteme bir bütün olarak fayda sağlayabilmesidir. Böylece MSA, farklı konumlardaki farklı alt uzaylardan gelen bilgilere katılabilir. Bir katmandaki tüm başların çıktıları birleştirilir, normalize edilir ve bir sonraki katman için girdi olarak ileri besleme ağı tarafından işlenir [12].



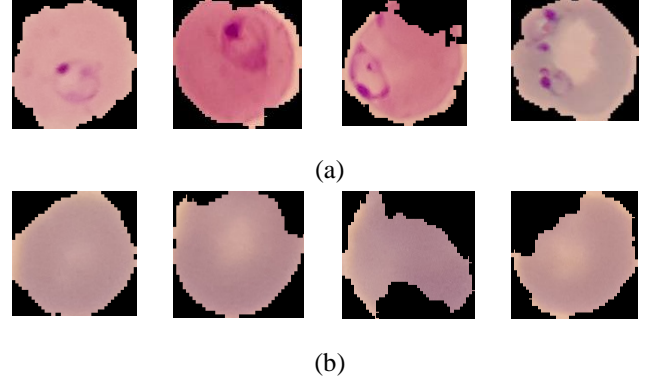
Şekil 3. ViT modelinin temelini oluşturan adımlardan biri olan öz dikkat mekanizması

## Deneysel Sonuçlar

### Veri kümeleri

Bu çalışma, Ulusal Sağlık Enstitüsü (NIH) tarafından oluşturulan veri setinden Giemsa ile boyanmış ince kan (kırmızı kan hücresi) görüntüleri üzerinde gerçekleştirilmiştir [9]. Akıllı telefonun dahili kamerası ile her bir mikroskopik görüş alanı için slayt görüntüleri alındı. Görüntüler, Tayland'daki Mahidol-Oxford Tropical Tıp Araştırma Birimi'ndeki uzman bir slayt okuyucu tarafından manuel olarak açıklanmıştır. Veri kaynağı Chittagong Medical College

Hospital, Bangladeş arşivlenmiş kan yayma görüntüleri ve orijinal görüntülerden eritositlerin segmentasyonu kullanıldı. Veri seti, parazitlenmiş ve enfekte olmamış hücrelerin eşit örneklerine sahip 27.558 hücre görüntüsünü içerir. Çalışmamızda, sınıf üstünlüğünü önlemek için veri setini her iki sınıfa eşit olarak ayırdık. Veri kümesinden alınan örnek görüntüler aşağıda Şekil 1'de gösterilmiştir.



Şekil 4: Sıtma veri setinden alınan parazitlenmiş hücrelerin (a) ve enfekte olmamış hücrelerin (b) örnek mikroskop görüntüleri

### Önerilen Transformer Based Yöntem ile elde edilen sonuçlar

Çalışma kapsamında kullanılan ViT modelinin 5-kat çapraz geçerlilik uygulanmış veri kümesindeki performansı literatürde önerilen diğer yöntemlerle karşılaştırılmış ve elde edilen sonuçlar Tablo 1'de gösterilmiştir. Geleneksel yöntemlerle elde edilen sonuçlar derin öğrenme yöntemlerine nispeten daha düşük performans gösterdiği için karşılaştırma amacıyla eklenmemiştir. Bu çalışmada derin öğrenme yöntemleri ve ViT modeli performansları karşılaştırılmıştır. Sonuçlar değerlendirildiğinde literatürde en yaygın kullanılan derin öğrenme yöntemlerine (LeNet, GoogleNet, AlexNet, EfficientNet) oranla ViT modeli daha yüksek başarı elde etmiştir. Buna ek olarak farklı veri kümeleri ile eğitilip transfer öğrenme ile ağırlıkların bu veri kümesinde kullanıldığı diğer bazı modeller ve tamamiyle bu veri kümesine özel geliştirilmiş veri kümelerine oranla ViT modeli yine daha yüksek başarı göstermiştir. ViT modelinin farklı derin öğrenme tabanlı özelliklerinin birleştirilerek daha ayırt edici özellik vektörlerinin çıkarılmaya çalışıldığı modellere göre nispeten daha düşük başarı gösterdiği gözlemlenmiştir. Burada kullanılan ViT modeli standart model olduğundan bu amaçla özelleştirilecek bir modelin performansı daha yukarı taşıyacağı düşünülmektedir. ViT modellerinde yüksek performans elde etmek amacıyla yüksek donanım kapasitelerinin kullanılması gerekmektedir. Bu çalışmada kullanılan ekran kartı işlemcisinin 8GB bellek kapasitesine sahip olması buna bağlı olan bazı parametrelerin sınırlandırılmasını zorunlu kıldığından daha yüksek başarı elde edilmesinin önüne geçmiştir.

Tablo 1. Uygulanan ViT modeli ile elde edilen sonuçlar ve literatürdeki yöntemlerle karşılaştırılması

Modeller	Veri Seti	Algoritma	Sınıflandırıcı Başarı Oranları
Diker A. [15]	Lister Hill National Center for Biomedical Communications (LHNCBC) [9]	CNN (ShuffleNet)	%96.44
Masud ve diğ. [16]	Lister Hill National Center for Biomedical Communications (LHNCBC) [9]	CLR-triangular2	%97.30
Akilotu ve diğ. [18]	Lister Hill National Center for Biomedical Communications (LHNCBC) [9]	Transfer öğrenme ve VGGNet	%96.02
Liang ve diğ. [5]	Lister Hill National Center for Biomedical Communications (LHNCBC) [9]	CNN	%97.37
Dong ve diğ [3]	PEIR-VM[19]	LeNet, AlexNet, GoogLeNet	%96.18
Rajaraman ve diğ [2]	Lister Hill National Center for Biomedical Communications (LHNCBC) [9]	CNN(cell level)	%98.13
Bibin ve diğ [4]	Lister Hill National Center for Biomedical Communications (LHNCBC) [9]	Deep Belief Networks (DBN)	F-skoru: %89.66 Hassasiyet: %97.60 Özgünlük: %95.92
Montalbo F. J. P. ve Alon A. S. [20]	Lister Hill National Center for Biomedical Communications (LHNCBC) ) [9]	EfficientNetB0	%94.70
Mahmud T. ve Fattah S. A. [17]	Lister Hill National Center for Biomedical Communications (LHNCBC) [9]	DeepFusionNet	<b>%99.50</b>
Irmak E. [21]	Lister Hill National Center for Biomedical Communications (LHNCBC) [9]	A Novel CNN	%95.28
Narayanan ve diğ [7]	Lister Hill National Center for Biomedical Communications (LHNCBC) [9]	Custom CNN	%96.70
Fatima T. ve Farid M. S. [22]	Lister Hill National Center for Biomedical Communications (LHNCBC) [9]	Bilateral filter	%91.80
Vijayalakshmi ve diğ. [1]	Mamic image database[25]	VGGNet ve SVM	%93.10
Abdurahman F. ve diğ. [26]	1182 color microscopic images of thick blood smear malaria slides (publicly malaria dataset)	YOLOV3 YOLOV4	%96.32 %96.14
Saiprasath G. ve diğ. [23]	2703 blood smear images 133 individuals[24]	AdaBoost, Decision Tree, KNN, Linear Regression, Naive Bayes, Random Forest, Extra Tress	%96.20 %94.60 %94.00 %94.30 %85.80 %96.50 %95.60
<b>Önerilen ViT Modeli</b>	<b>Lister Hill National Center for Biomedical Communications (LHNCBC)</b>	<b>Transformer</b>	<b>%97.22±1</b>

ViT modellerinin geliştirilmesinde önemli parametrelerin başında girdi görüntüsünün boyutları ve her biri doğal dil işlemedeki kelimeye karşılık gelen küçük görüntü kesitleri gelmektedir. Girdi görüntüsü toplamda kaç parça görüntü kesitine bölünecekse işlem maliyeti ve elde edilen performans buna göre değişmektedir. Bu çalışmada 128\*128 boyutlarına sahip girdi görüntüsü, 16x16'lık görüntü kesitlerine bölünmüştür. ViT modellerinde normalizasyon amacıyla uygulanan transformer katmanlarının sayısı (transformer layer) parametre olarak kullanıcı tarafından verilmektedir. Bu çalışmada 4 adet transformer katmanı uygulanmıştır. Buna ek olarak her bir dikkat mekanizması transformer kodlama katmanında arka arkaya uygulanagelen işlemlerdir. Bu dikkat mekanizmalarının uçlarına buradan elde edilen çıktılar bir sonraki aşamaya iletmek için başlık (head) kullanılmaktadır. Bu başlıklar tek bir dikkat mekanizmasından elde edilebileceği daha düşük boyutlu bir uzayda çoklu hesaplamalar olarak da yapılabilmektedir. Bu işlemler sistemin performansını arttırmak amacıyla yapılmaktadır. Daha sonra bu başlıkların sonuçları birleştirilmektedir. Çalışma kapsamında 5 adet başlık kullanılmıştır. Standart derin öğrenme yöntemlerinde olduğu gibi ViT modellerinde de yığın boyutu, epoch sayısı ve öğrenme katsayısı mevcuttur. Bu çalışmada epoch sayısı, öğrenme katsayısı ve yığın boyutu için sırasıyla, 500, 0.001 ve 128 değerleri kullanılmıştır.

Tüm testler, AMD Ryzen 7 2700 8 çekirdekli 3.2 ghz, 32GB RAM ve NVIDIA Geforce RTX 2060 SUPER 8GB içeren bir iş istasyonunda tensorflow kütüphanesi ve Spyder (Python 3.7 ile) ortamı kullanılarak gerçekleştirilmiştir.

## Tartışma

Bu çalışmada Malaria hücrelerinin sınıflandırılmasında son dönemde doğal dil işlemede oldukça başarılı performans veren dönüştürü modellerinin bir türevi kullanılmıştır. Çalışma kapsamında kullanılan ViT modeli standart ViT modelinden esinlenerek geliştirilen ViT modelidir. ViT modelleri doğal dil işleme veya görüntü kesitleri farketmeksizin öbekler arasında ilişki hesapladıklarından yüksek boyutta kaynak gereksinimi duymaktadır. Daha büyük kapasiteli donanımlarda yüksek başarı elde edildiğinden son dönemde birçok alanda oldukça popüler olmuşlardır. Bu çalışmada da malaria hücrelerinin sınıflandırılmasında ViT modellerinden yararlanılması amaçlanmıştır. Deneysel sonuçlar başlığı altında elde edilen sonuçlar değerlendirildiğinde ViT modellerinin literatürde sıkça kullanılan bir çok yöntemden daha yüksek performans gösterdiği gözlemlenmektedir. Geleneksel görüntü işleme yöntemleri malaria tespitinde belirli bir başarının üzerine çıkamazken, derin öğrenme tabanlı yöntemler %90'ların üzerinde performans göstermektedir. Bu çalışmada temel odak noktası da derin öğrenme yöntemleri ile ViT modelinin karşılaştırılması ve avantaj

ve dezavantajların analiz edilmesidir. ViT modeli önceden eğitilmiş LeNet, AlexNet, GoogleNet, EfficientNet gibi modellerin tekil olarak kullanılmasıyla elde edilen sonuçlardan daha iyi performans göstermiştir. Ancak yine derin öğrenme yöntemleri kullanılarak elde edilen öznetelik vektörlerinin birleştirilmesi ile elde edilen sınıflandırma başarısı standart ViT modelinden daha iyi performans göstermiştir. Bunda sınırlı donanım kaynaklarının etkisi olduğu görülmüştür. Çünkü ViT modelleri en az 30-40 GB mertebelerinde bellek boyutuna ihtiyaç duyarken bu çalışmada 8 GB bellek kapasiteli bir ekran kartı işlemcisi kullanılmıştır. Bu durum elde edilen sonuçların sınırlı kalmasına sebep olmuştur. Kaynakların artırılması ile bu performansın daha yukarı çıkması düşünülmektedir.

ViT modellerinde iyileştirilmesi gereken bir diğer nokta, yöntemde kullanılan parametrelerdir. Transformer katmanı, projeksiyon boyutu (projection dimension), başlık sayısı(number of head),

## Sonuç

Bu çalışmada son dönemde doğal dil işleme alanında oldukça yüksek performans gösteren ViT modeli Sıtma hastalığının tespitinde kullanılmıştır. Ayrıca çalışma kapsamında elde edilen sonuçlar literatürde önerilen geleneksel görüntü işleme yöntemleri ve derin öğrenme tabanlı yöntemlerle karşılaştırılmıştır. Sonuçlar değerlendirildiğinde Öneirlen ViT modelinin geleneksel yöntemler ve bazı derin öğrenme yöntemlerinden nispeten daha başarılı olduğu ancak derin öğrenme tabanlı özelliklerin birleştirildiği çalışmalarla yakın sonuçlar verse de nispeten daha az başarılı olduğu gözlemlenmiştir. Bunun ana sebebi ViT modellerinin yüksek performans elde etmek için büyük boyutlu donanım kaynaklarına ihtiyaç duymasındır. Gelecek çalışmalarda bu çalışmada kullanılan donanım kaynaklarından daha büyük boyutlu kaynaklar kullanarak başarıyı yükseltmeyi amaçlamaktayız.

## Kaynaklar

- [1] A. Vijayalakshmi and B. Rajesh Kanna, "Deep learning approach to detect malaria from microscopic images", *Multimedia Tools and Applications*, vol. 79, 2019.
- [2] S. Rajaraman et al., "Understanding the learned behavior of customized convolutional neural networks toward malaria parasite detection in thin blood smear images", *J. Med. Imag.*, vol. 5, no. 3, Jul. 2018.
- [3] Y. Dong et al., "Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells", *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform. (BHI)*, pp. 101-104, 2017.
- [4] D. Bibin, M. S. Nair and P. Punitha, "Malaria Parasite Detection From Peripheral Blood Smear Images

- Using Deep Belief Networks," in IEEE Access, vol. 5, pp. 9099-9108, 2017, doi: 10.1109/ACCESS.2017.2705642.
- [5] Z. Liang et al., "CNN-based image analysis for malaria diagnosis," 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2016, pp. 493-496, doi: 10.1109/BIBM.2016.7822567.
- [6] J. Hung and A. Carpenter, "Applying Faster R-CNN for Object Detection on Malaria Images," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 808-813, doi: 10.1109/CVPRW.2017.112.
- [7] B. N. Narayanan, R. Ali and R. C. Hardie, "Performance analysis of machine learning and deep learning architectures for malaria detection on cell images", *Applications of Machine Learning*, vol. 11139, pp. 11139W, September 2019.
- [8] S. Muneer, S. Jatou, B. Naz, "Detection of Noisy Blood Images Indicating Prodromal Diseases", *International Journal of Latest Trends in Engineering and Technology*, Vol(12), Issue(6), pp. 001-006, 2018.
- [9] lhncbc.nlm.nih.gov, "Lister Hill National Center for Biomedical Communications", 2021. [Çevrimiçi]. <https://ceb.nlm.nih.gov/repositories/malaria-datasets/> [Erişim Tarihi: 15.05.2021]
- [10] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *In NAACL*, 2019.
- [11] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, ..., et al., "An image is worth 16x16 words: Transformers for image recognition at scale", *In ICLR*, 2021.
- [12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., "Attention is all you need", *Advances in Neural Information Processing Systems*, pp. 6000-6010, 2017.
- [13] Q. Wang et al., "Learning deep transformer models for machine translation", *Proc. Annu. Meeting Assoc. Comput. Linguistics*, pp. 1810-1822, 2019.
- [14] T. Czempiel, M. Paschali, D. Ostler, S. T. Kim, B. Busam and N. Navab, "OperA: Attention-regularized transformers for surgical phase recognition", *arXiv:2103.03873*, 2021, [online] Available: <http://arxiv.org/abs/2103.03873>.
- [15] A. Diker, "Sıtma Hastalığının Sınıflandırılmasında Evrişimsel Sinir Ağlarının Performanslarının Karşılaştırılması", *BEÜ Fen Bilim. Derg.*, vol. 9, no. 4, pp. 1825-1835, 2020.
- [16] M. Masud, H. Alhumyani, S. S. Alshamrani, O. Cheikhrouhou, S. Ibrahim, G. Muhammad, et al., "Leveraging deep learning techniques for malaria parasite detection using mobile application", *Wireless Commun. Mobile Comput.*, vol. 2020, pp. 1-15, Jul. 2020.
- [17] T. Mahmud, & S. A. Fattah, "Automatic Diagnosis of Malaria from Thin Blood Smear Images using Deep Convolutional Neural Network with Multi-Resolution Feature Fusion", *arXiv preprint arXiv:2012.05350*. 2020.
- [18] B. N. Akılotu, Z. Kadiroğlu, A. Şengür and M. Kayaoğlu, "Evrişimsel Sinir Ağları ve Transfer Öğrenme Yöntemi Kullanılarak Sıtma Tespiti", *International Engineering and Science Symposium*, Siirt, 2019.
- [19] Whole slide image for malaria infected red blood cells, [Çevrimiçi] Erişim: <http://peir-vm.path.uab.edu/debug.php?slide=IPLab11Malaria> [Erişim Tarihi: 15.05.2021].
- [20] F. Montalbo and A. Alon, "Empirical Analysis of a Fine-Tuned Deep Convolutional Model in Classifying and Detecting Malaria Parasites from Blood Smears", *KSII Transactions on Internet and Information Systems*, vol. 15, no. 1, 2021.
- [21] E. Irmak, "A novel implementation of deep-learning approach on malaria parasite detection from thin blood cell images", *Electrica* 21(2):216-224. 2021. <https://doi.org/10.5152/electrica.2020.21004>.
- [22] T. Fatima and M.S Farid, "Automatic detection of Plasmodium parasites from microscopic blood images", *Journal of Parasitic Diseases*, vol. 44, no. 1, pp. 69-78, 2020.
- [23] G. Saiprasath, N. Babu, J. ArunPriyan, R. Vinayakumar, V. Sowmya and K. Soman, Performance comparison of machine learning algorithms for malaria detection using microscopic images, 2019.
- [24] J. A. Quinn, A. Andama, I. Munabi, F. N. Kiwanuka, "Automated blood smear analysis for mobile malaria diagnosis", *Mobile Point-of-Care Monitors and Diagnostic Device Design*, 31-115, 2014.
- [25] <http://fimm.webmicroscope.net/Research/Momic/mamic> WebMicroscope "Institute for molecular medicine Finland and FIMM". <http://fimm.webmicroscope.net/Research/Momic/mamic>, [Erişim Tarihi: 1 June 2018]
- [26] F. Abdurahman, KF Anlay and M Aliy, "Malaria Parasite Detection in Thick Blood Smear Microscopic Images Using Modified YOLOV3 and YOLOV4 Models", *Research Square*, 2020.