



ANKARA ÜNİVERSİTESİ COMPSEG
HİSTOPATOLOJİ GÖRÜNTÜLERİ ARAŞTIRMA GRUBU

TÜBİTAK Proje No: 121E379 ve “Histopatoloji Görüntülerinde Meme Kanseri Tespiti için Derin Öğrenme Tabanlı Metodolojinin Geliştirilmesi” Konulu Proje Kapsamında Oluşturulan

NuSeC ve MiDeSeC Veri Setleri

NuSeC ve MiDeSeC Veri Setlerinin Tanımı

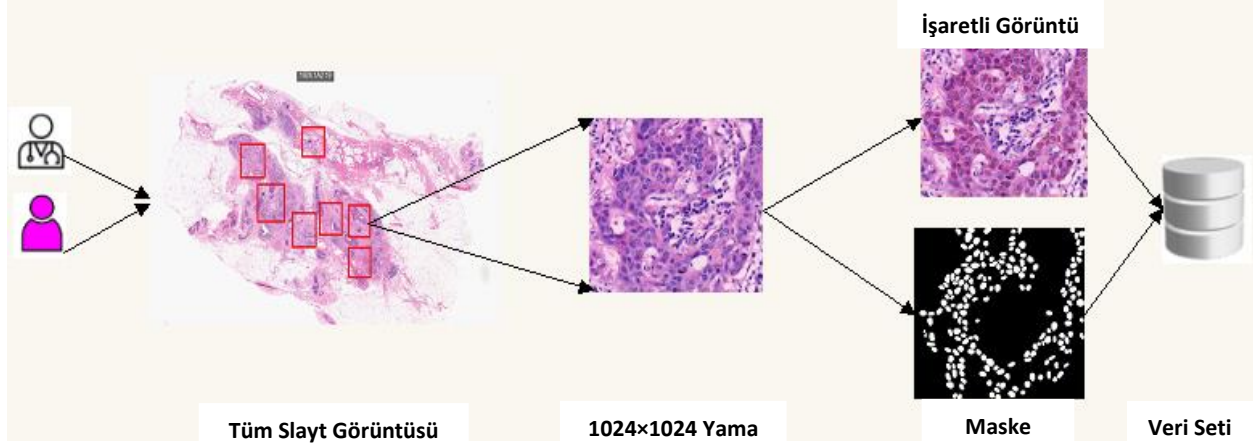
(NuSeC and MiDeSeC Datasets Description)

1. Giriş

“Histopatoloji Görüntülerinde Meme Kanseri Tespiti için Derin Öğrenme Tabanlı Metodolojinin Geliştirilmesi” başlıklı TÜBİTAK 1001 projesi kapsamında meme kanseri tespiti ve derecelendirilmesi için derin öğrenme modellerinin eğitim ve testinde kullanılmak üzere Çekirdek Segmentasyonu ve Sınıflandırılması (**NuSeC – Nuclei Segmentation and Classification**) ve Mitoz Tespit, Segmentasyon ve Sınıflandırma (**MiDeSeC – Mitosis Detection, Segmentation and Classification**) için veri setleri oluşturulmuştur.

Meme kanseri, özellikle kadınlarda olmak üzere dünya çapında önde gelen kansere bağlı ölüm nedenlerinden biridir. Ancak erken tespit, tedavi başarısını önemli ölçüde artırmaktadır. Erken tespit amacıyla, histopatoloji görüntüleri doğru analiz edilmelidir. Spesifik olarak, tespit prosedürü sırasında uzmanlar hem genel hem de yerel doku organizasyonunu tüm slayt ve mikroskopik görüntülerle değerlendirir. Bununla birlikte, büyük miktarda veri ve görüntülerin karmaşıklığı, bu görevi zaman alıcı ve zahmetli kılıyor ve uzmanlar arasında farklı yorumlar yapılmasına neden oluyor. Bu nedenle, bilgisayar destekli (otomatik) tespit için yazılım araçlarının geliştirilmesi gerekiyor.

Bu çalışmada meme kanserinin otomatik tespiti için kullanılacak NuSeC ve MiDeSeC veri setlerinin tanımı, hazırlama ve oluşturma aşamaları açıklanmaktadır. Veri setleri oluşturma aşamaları Şekil 1’de görülmektedir. İlk aşamada, tüm slayt görüntüleri 1024x1024 piksel boyutlarında yamalara bölünmekte ve ikinci aşamada ise görüntüler piksel bazında etiketlenmekte/işaretlenmekte ve maskeleri oluşturulmaktadır.



Şekil 1. Veri Seti Oluşturma Aşamaları

Veri setinde kullanılan görüntüler, Ankara Üniversitesi Tıbbi Patoloji Anabilim Dalı'nda 40x büyütme ile üretilen 25 farklı hastanın H&E boyalı invaziv meme karsinomu, özel tip (NST) slaytlar kullanılarak oluşturulmuştur. Slaytlar 3D Histech Panoramic p250 Flash-3 tarayıcı ve Olympus BX50 mikroskobu ile taranmıştır. Ver setlerinin oluşturulmasında QuPath yazılımı kullanılmıştır.

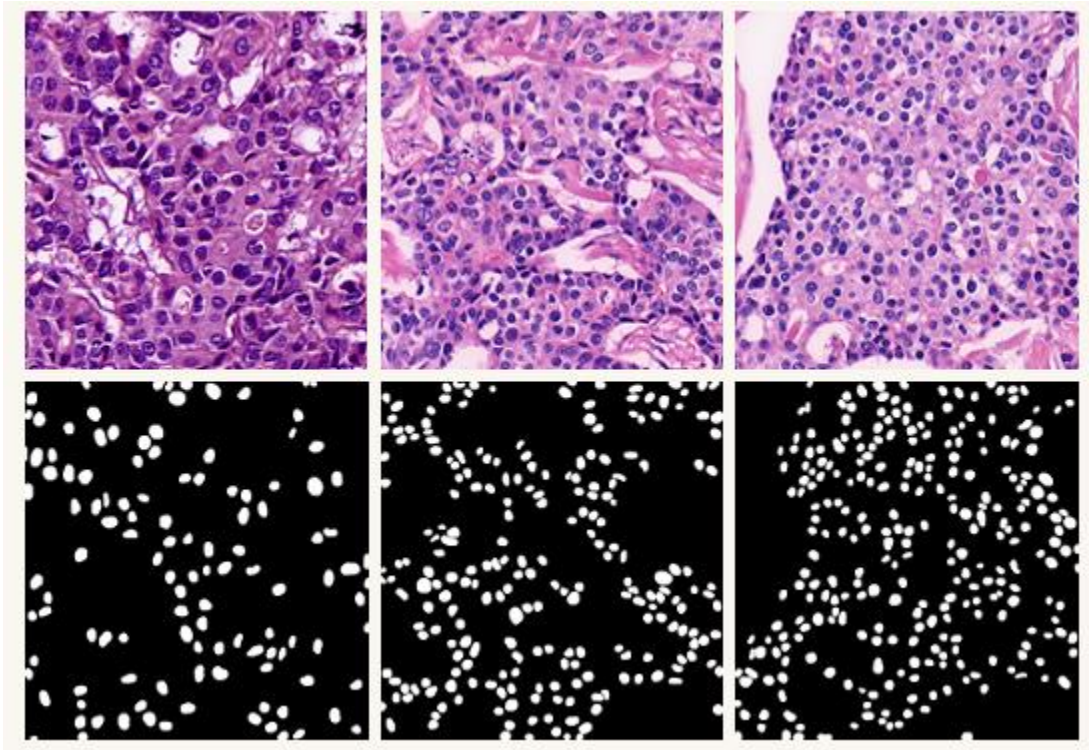
Veri setleri aşağıda listelenen Gayri Patolog (NP) ve Patolog (P) araştırmacılar tarafından oluşturulmuştur (<http://compseg.ankara.edu.tr/arastirmacilar/>):

- 1) Prof.Dr.Refik Samet (NP), Ankara Üniversitesi Bilgisayar Mühendisliği Bölümü, Bilgisayar Mühendisi;
- 2) Prof.Dr. Serpil Sak (P), Ankara Üniversitesi Tıbbi Patoloji Anabilim Dalı, Patolog;
- 3) Doç.Dr.Emrah Hancer, Burdur Mehmet Akif Ersoy Üniversitesi Yazılım Mühendisliği Bölümü, Bilgisayar ve Yazılım Mühendisi;
- 4) Doç.Dr.Bilge Ayça Kırmızı, Ankara Üniversitesi Tıbbi Patoloji Anabilim Dalı, Patolog;
- 5) Doktora Öğrencisi Zeynep Yıldırım, Ankara Üniversitesi Bilgisayar Mühendisliği Bölümü, Bilgisayar Mühendisi;
- 6) Doktora Öğrencisi Nuşin Nemati, Ankara Üniversitesi Bilgisayar Mühendisliği Bölümü, Bilgisayar Mühendisi;
- 7) Yüksek Lisans Öğrencisi Mohamed Traoré, Ankara Üniversitesi Bilgisayar Mühendisliği Bölümü, Bilgisayar Mühendisi.

Veri setlerinde kullanılan tam slayt görüntüleri Ankara Üniversitesi Tıbbi Patoloji Anabilim Dalı tarafından temin edilmiştir. Önce, tüm slayt görüntüleri NP araştırmacılar tarafından işlenerek veri setlerinin taslakları oluşturulmuştur. Sonra, oluşturulan veri setleri taslakları Ankara Üniversitesi Tıbbi Patoloji Anabilim Dalı P araştırmacıları tarafından kontrol edilmiş, düzeltilmiş ve onaylanmıştır. Son olarak, proje kapsamında geliştirilen ve literatürde bilinen derin öğrenme modelleri kullanılarak gerçekleştirilen denemeler sonucunda gerekli düzeltmeler ve değişiklikler yapılarak veri setlerinin son hali oluşturulmuş ve web sayfamızdan paylaşılmıştır (<http://compseg.ankara.edu.tr/veri-setleri/>).

2. NuSeC Çekirdek Segmentasyonu ve Sınıflandırılması Veri Seti (NuSeC – Nuclei Segmentation and Classification Dataset)

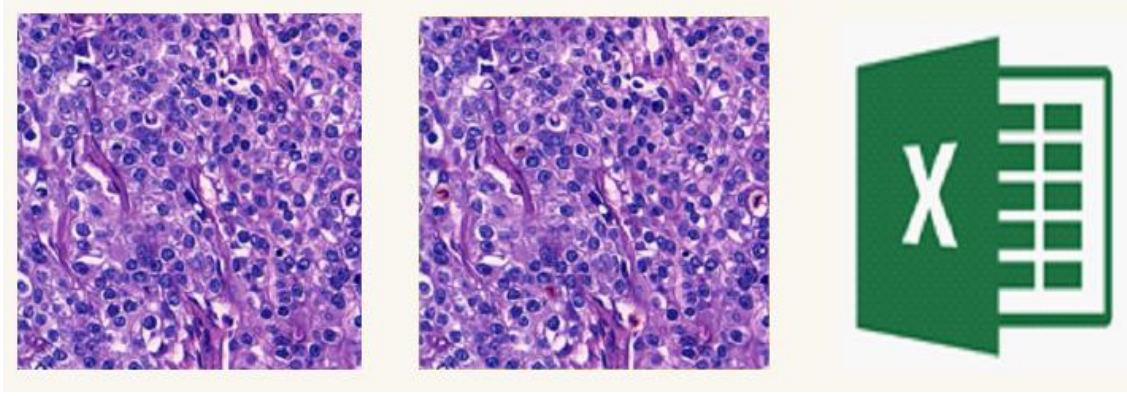
NuSeC veri seti, 25 hastaya ait tam slayt görüntülerinin her birinden 4 adet 1024×1024 piksel boyutunda görüntüler olmak üzere toplam 100 adet görüntüden oluşmaktadır. NuSeC veri seti, %75 (75 adet) eğitim amaçlı ve %25 (25 adet) test amaçlı olmak üzere eğitim ve test alt veri setleri şeklinde oluşturulmuştur. Test veri seti, 25 hastaya ait tam slayt görüntülerinin her birinden üretilen 4 adet görüntülerden birer adet rastgele seçilerek oluşturulmuştur. Eğitim veri seti ise, 25 hastaya ait tam slayt görüntülerinin her birinden üretilen 4 adet görüntülerden test veri seti için rasgele seçilen bir adet hariç kalan üçer adet görüntülerden oluşturulmuştur. Eğitim veri setindeki 75 adet görüntüde yaklaşık 30000 çekirdek yapısı bulunmaktadır. Test veri setindeki 25 adet görüntüde ise yaklaşık 6000 adet çekirdek yapısı bulunmaktadır. Şekil 2’de NuSeC veri setine ait örnek görüntüler görülmektedir.



Şekil 2. NuSeC veri setine ait örnek görüntüler. Birinci satırda 1024x1024 piksel boyutlarında H&E ile boyalı ham histopatoloji görüntülere ait, ikinci satırda ise bu görüntülere için üretilen maske görüntülerine ait örnekler görülmektedir.

3. MiDeSeC Mitoz Tespit, Segmentasyon ve Sınıflandırma Veri Seti (MiDeSeC – Mitosis Detection, Segmentation and Classification Dataset)

MiDeSeC veri seti oluşturulurken farklı mitoz şekillerinin temsil edilmesi için farklı hastalara ait görüntülerin bulunmasına özen gösterilmiştir. Bu amaçla, 25 hastaya ait tam slayt görüntülerinde mitoz yapılarının bulunduğu 50 bölge işaretlenmiş, ve bu bölgeden birer olmak üzere 50 adet 1024×1024 piksel boyutunda görüntüler oluşturulmuştur. Üretilen görüntülerde yaklaşık 500 mitoz bulunmaktadır. MiDeSeC veri seti, %70 (35 adet) eğitim amaçlı ve %30 (15 adet) test amaçlı olmak üzere eğitim ve test alt veri setleri şeklinde oluşturulmuştur. Test veri seti, 25 hastaya ait 50 adet görüntü arasından rastgele seçilerek oluşturulmuştur. Eğitim veri seti ise, test veri seti için rasgele seçilen 15 adet hariç kalan 35 adet görüntülerden oluşturulmuştur. Son olarak mitoz koordinatlarının bulunduğu CSV dosyası üretilmiştir. Şekil 3’de MiDeSeC veri setine ait örnek görüntüler görülmektedir.



Şekil 3. MiDeSeC veri setine ait örnek görüntüler. Solda 1024x1024 piksel boyutlarında H&E ile boyalı ham histopatoloji görüntü, ortada, aynı görüntüdeki işaretlenmiş/etiketlenmiş mitozları içeren görüntü ve sağda mitoz koordinatlarının kaydedildiği excel dosyası görülmektedir.

4. Veri Seti Linkleri

NuSeC Veri seti linki (<http://compseg.ankara.edu.tr/veri-setleri>)

MiDeSeC Veri seti linki (<http://compseg.ankara.edu.tr/veri-setleri>)

Görüş ve öneriler için e-mail adresi: compseg@ankara.edu.tr