**Makine Öğrenim Yöntemleriyle Meme Kanseri Teşhisi için İğne Biyopsisi Verilerinin Sınıflandırılması**

Özer Özaydın\_11, Özge Demir\_22

Dünyada her beş kişiden biri hayatları boyunca kanserle yüz yüze gelmektedir ve sekiz erkekten biri ve on bir kadından biri kanser sebebiyle yaşamları sonlanmaktadır (GLOBOCAN veri tabanı, 2020). Kadınlarda, en yaygın görülen 5 kanser türünden biri meme kanseridir. Tüm kanser türlerinde olduğu gibi meme kanserinde de erken tanı ve teşhis hayati önem arz etmektedir. Bu çalışmada meme kanserinin bir meme kütlesinin ince iğne aspirasyonunun yani biyopsi ile alınan örneklerden elde edilen verilerin teşhisi için sınıflandırma modelleri oluşturulacaktır.

Çalışmada son zamanlarda sınıflandırmada popüler olan yapay zekâ makine öğrenim algoritmalarından Yapay Sinir Ağları (YSA), Destek Vektör Makineleri (DVM) ve klasik istatistiksel sınıflandırma yöntemlerinden k-En Yakın Komşu (KNN) algoritması kullanılarak sınıflandırma performansları karşılaştırılacak ve en uygun model sunulacaktır. Bu çalışmada kullanılacak veri seti UCI (The University of California, Irvine) tarafından makine öğrenimi için depolanan verilerden (https://archive.ics.uci.edu/ml/index.php) oluşturulmuştur.

Makine öğrenim yöntemleri literatürde tıp alanında özellikle kanser teşhis ve ön tanısı konusunda yenilikçi ve oldukça iyi çözümler sunmaktadır. Söz konusu çalışmalar hem tıbbi teşhis koyuculara kolaylık hem de hayati önem arz eden durumların önceden fark edilmesini sağlamaktadır. Makine öğrenim yöntemleri istatistiksel temele dayalı optimizasyon algoritmalarıdır. Bu çalışmada kullanılacak YSA beynin sinirsel iletim ve çalışmasını hayata geçirerek karmaşık problemleri makine aracılığıyla çözmeyi sağlamaktadır. YSA klasik sınıflandırmada ağaçlandırma yöntemine benzemektedir. DVM, Vladimir Vapnik tarafından geliştirilen yine istatiksel optimizasyon temelli bir makine öğrenim algoritmasıdır. Söz konusu algoritma, büyük verilerde ve ikili sınıflandırmalarda yaygın olarak kullanılmaktadır. DVM, veri uygulanabilirliğini optimum performans için çekirdek seçimleri sunmakta, aynı zamanda bir hiper düzlem oluşturarak uzaklık parametresini maksimum hale getirerek sınıflandırma yapmaktadır. Klasik istatistiksel sınıflandırma yöntemlerinden KNN ise, tüm mevcut verileri depolayan ve bir benzerlik ölçüsüne (ör. mesafe fonksiyonları) dayalı yeni verileri sınıflandıran basit bir algoritmadır. KNN, 1970’lerden itibaren parametrik olmayan bir yöntem olarak istatistiksel tahmin ve örüntü tanımlamada kullanılmaktadır.

**Anahtar Kelimeler:** Destek Vektör Makineleri, Yapay Sinir Ağları, K – en yakın komşu