

Mustafa Etcil



mustafa.etcil@agu.edu.tr

0000-0002-5573-3418



Thesis Advisor

Assoc.Dr. Burcu
Bakır Güngör

burcu.gungor@agu.edu.tr

Title of the Thesis

abstract According to the World Health Organization (WHO), breast cancer is one of the most prevalent illnesses, with 7.8 million instances recorded in the previous five years. As such, it poses a serious threat to world health. This alarming statistic underscores the urgent necessity for enhanced diagnostic methods. Against this backdrop, the current study proposes a novel diagnostic model, the CSA-PSO-LR classifier, which innovatively combines the clonal selection algorithm (CSA) with particle swarm optimization (PSO) to refine the logistic regression model training process for breast cancer detection. This research employs two extensively recognized datasets: the Wisconsin Diagnostic Breast Cancer (WDBC) and the Wisconsin Breast Cancer Database (WBCD), putting into practice a strict evaluation procedure that assesses performance using Bayesian hyperparameter optimization and 10-fold cross-validation. Furthermore, the study introduces CPU parallelization strategies to significantly curtail the model training time. Comparative analyses against machine learning algorithms, encompassing decision trees, extreme gradient boosting, k-nearest neighbors, logistic regression, random forests, and support vector machines, demonstrate the CSA-PSO-LR classifier's superior performance in detection accuracy and F1-measure. This investigation contributes a groundbreaking approach to the early detection of breast cancer, potentially facilitating more effective treatment plans and enhancing patient survival prospects.

keywords: Clonal selection algorithm, Particle swarm optimization, Logistic regression, Bayesian optimization, Breast cancer diagnosis

özet Dünya Sağlık Örgütü (WHO) tarafından belirlendiği üzere, göğüs kanseri, son beş yılda 7.8 milyon yeni vakayla en yaygın kanser türlerinden biri olarak ön plana çıkmaktadır. Bu çarpıcı istatistik, gelişmiş tanı yöntemlerine olan acil ihtiyacı vurgulamaktadır. Bu bağlamda, mevcut çalışma, göğüs kanseri tespiti için lojistik regresyon modeli eğitim sürecini iyileştirmek amacıyla klonal seçim algoritması (CSA) ile parçacık sürü optimizasyonunu (PSO) yenilikçi bir şekilde birleştiren CSA-PSO-LR sınıflandırıcısını önermektedir. Bu araştırma, geniş çapta tanınan iki veri seti olan Wisconsin Diagnostik Göğüs Kanseri (WDBC) ve Wisconsin Göğüs Kanseri Veritabanı (WBCD) kullanılarak, performans değerlendirmesi için 10 kat çapraz doğrulama ve Bayes hiperparametre optimizasyonunu içeren katı bir değerlendirme protokolü uygulamaktadır. Ayrıca, çalışma, model eğitim süresini önemli ölçüde kısaltmayı amaçlayan CPU paralelleştirme stratejilerini tanıtmaktadır. Karar ağaçları, aşırı gradyan artırma, en yakın komşular, lojistik regresyon, rastgele ormanlar ve destek vektör makineleri gibi makine öğrenimi algoritmalarına karşı yapılan karşılaştırmalı analizler, CSA-PSO-LR sınıflandırıcısının tespit doğruluğu ve F1-ölçütü açısından üstün performans sergilediğini göstermektedir. Bu araştırma, göğüs kanserinin erken tespitine yönelik yenilikçi bir yaklaşım sunarak, daha etkili tedavi planlarının kolaylaştırılmasına ve hastaların hayatta kalma beklentilerinin artırılmasına katkıda bulunmaktadır.

anahtar kelime Klonal seçim algoritması, Parçacık sürü optimizasyonu, Lojistik regresyon, Bayes optimizasyonu, Göğüs kanseri teşhisi