Pamukkale Univ Muh Bilim Derg, 29(3), 256-263, 2023

Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi

Pamukkale University Journal of Engineering Sciences

Özellik seçim algoritmaları ve derin öğrenme tabanlı mimarilerin hibrit kullanımıyla akut lösemilerin sınıflandırılması

Classification of acute leukaemias with a hybrid use of feature selection algorithms and deep learning-based architectures

Fatma AKALIN[1*], *Nejat YUMUSAK*[2]

[1]Bilişim Sistemleri Mühendisliği Bölümü, Bilgisayar ve Bilişim Bilimleri Fakültesi, Sakarya Üniversitesi, Sakarya, Türkiye.
fatmaakalin@sakarya.edu.tr

[2]Bilgisayar Mühendisliği Bölümü, Bilgisayar ve Bilişim Bilimleri Fakültesi, Sakarya Üniversitesi, Sakarya, Türkiye.
nyumusak@sakarya.edu.tr

Geliş Tarihi/Received: 02.06.2022 Düzeltme Tarihi/Revision: 28.08.2022 doi: 10.5505/pajes.2022.62282
Kabul Tarihi/Accepted: 31.08.2022 Araştırma Makalesi/Research Article

Öz

Tıp ve biyoloji alanlarında tercih edilen mikrodizi teknolojisi, kantitatif veya niteliksel veriler üreten bir analiz yöntemidir. Genler arasındaki örüntülerin açığa çıkartılarak yorumlanabilmesi için güçlü bir potansiyel barındırmaktadır. Bu potansiyeli ortaya çıkarmak için genler ile ilişkili kanser hastalıkları üzerinde moleküler değerlendirme sağlamak mümkündür. Ancak mikrodizi veri kümeleri, yüksek boyutlu bir yapıya sahiptir. Bu durum makine öğrenmesinde boyutluluğun laneti olarak bilinmektedir. Mikrodizi veri kümeleri üzerinde değerlendirme sürecinin kolaylaştırılması için bilgisayar destekli sistemler kullanılarak uzmanlara yardımcı bir fikir verilmesi temel amaçtır. Bu çalışmada akut lösemilerin sınıflandırılabilmesi için yüksek boyut sunan mikrodizi veri kümesi analiz edilmiştir. Çalışmanın ilk aşamasında, hastalıkla ilişkili genlerin veri kümesinden seçilebilmesi için karınca kolonisi, balina ve parçacık sürü optimizasyon algoritmaları kullanılmıştır. Seçilen potansiyel genler klasik makine öğrenmesi algoritmaları ile değerlendirilmiştir. Çalışmanın ikinci aşamasında elde edilen bu genler, dalgacık dönüşümü yöntemi ile spektrogramlar olarak ifade edilmiştir. Çalışmanın üçüncü aşamasında, spektrogramlardaki yerel kontrastın iyileştirilmesi için CLAHE yöntemi kullanılmıştır. Son olarak elde edilen iyileştirilmiş spektrogramlar; aktarım öğrenme mimarileri ve DGCNN(derin graf evrişimsel sinir ağı) yaklaşımı ile sınıflandırılmıştır. Karınca, parçacık sürü ve balina özellik seçim algoritmaları kullanılarak seçilen genlerin spektral yoğunluk bilgisinin ifade edildiği spektrogramların DGCNN yaklaşımı ile sınıflandırılmasının sonucunda elde edilen maksimum başarı oranları sırasıyla %93.33, %86.6 ve %86.6 olarak bulunmuştur.

Anahtar kelimeler: Mikrodizi teknolojisi, Doğadan ilham alan optimizasyon algoritmaları, Sürekli dalgacık dönüşümü tekniği, Derin graf evrişimli sinir ağı yaklaşımı, ALL ve AML malignitelerinin sınıflandırılması.

Abstract

The microarray technology which is preferred in the fields of medicine and biology is an analysis method that produces quantitative or qualitative data. It has a strong potential for revealing and interpreting patterns between genes. To reveal this potential, it is possible to provide a molecular evaluation of cancer diseases associated with genes. However, microarray datasets have a high dimensional structure. This is known as the curse of dimensionality in machine learning. The main aim is to give a helpful idea to the experts by using computer-aided systems to facilitate the evaluation process on microarray datasets. In this study, a high-dimensional microarray dataset is analyzed for the classification of acute leukaemias. In the first phase of the study, ant colony, whale and particle swarm optimization algorithms are used to select disease-related genes from the dataset. Selected potential genes were evaluated with classical machine learning algorithms. These genes obtained in the second stage of the study were expressed as spectrograms by the wavelet transform method. In the third stage of the study, the CLAHE method is used to improve the local contrast in the spectrograms. Finally, the obtained improved spectrograms are classified by transfer learning architectures and DGCNN (deep graph convolutional neural network) approach. The maximum success rates obtained as a result of the classification of the spectral density information of the selected genes using the ant, particle swarm and whale feature selection algorithms with the DGCNN approach are found to be 93.33%, 86.6% and 86.6%, respectively.

Keywords: Microarray technology, Nature-inspired optimization algorithms, Continuous wavelet transform technique, Deep graph convolutional neural network approach, Classification of ALL and AML malignancies.

1 Giriş

Hücre, canlılığın temel birimidir. Bölünebilme özelliği sayesinde yaralanan dokuların onarılmasını ya da ölen hücrelerin yenilenebilmesini sağlar. Sağlıklı bir hücre için bu yeti sınırlıdır. Ancak; genetik ya da çevresel etkenlerden dolayı hücrelerin kontrolsüz bir şekilde bölünebilmesi de muhtemeldir. Böyle bir ihtimalin gerçekleşmesi sonucunda vücudun farklı bölgelerinde malignite oluşumları meydana gelebilmektedir. Kanser hücreleri bu oluşumların başlıca sebebidir[1].

Kanser, teşhis ve tedavi sürecinde maliyeti en yüksek hastalıklar arasında yer alır. Kişinin hayat kalitesini düşürür ve erken teşhisin sağlanamaması sonucunda tedavi sürecini olumsuz etkileyerek ölüme sebep olabilir. Çünkü vücudun kendi kendini düzeltebilen bir mekanizması vardır. Kanser hastalığı ile bozulan bu mekanizma, genetik kodlamanın altında yatan yanlış düzenlemelerin bir dereceye kadar kötüleşmesine ve nihayetinde vücudun bu durumu düzeltemeyecek duruma

*Yazışılan yazar/Corresponding author

gelmesine neden olur. Bu süreç, teşhisin sağlanabilmesi açısından oldukça önemlidir [2].

Bu çalışmada, genlerle ilişkili kanser hastalıklarının teşhisinde moleküler tanı imkânı sağlayan mikrodizi teknolojisi tercih edilmiştir. Mikrodizi teknolojisi, birçok genin nispi ekspresyon değerlerinin aynı anda izlenmesini sağlayarak genomik ölçekli bir inceleme sunmakta ve kanserin moleküler keşfi üzerinde temellenen bir sınıflandırmayı mümkün kılmaktadır. Ayrıca gen ekspresyon profilinin kullanımı, tümörün ilerleme durumu hakkında bilgi sağlamak ve hastalara daha iyi bir tedavi planı oluşturulabilmek için de tercih edilmektedir [3]-[6]. Ancak mikrodizi teknolojisi ile elde edilen genetik verinin miktarı dikkate alındığında geleneksel sınıflandırma yöntemleri ile veriler üzerinden çıkarım yapılmasının güç olduğu görülmektedir. Gen ekspresyon verilerinin az sayıda hasta örneği barındırmasının yanı sıra genleri temsil eden farklı ve çok sayıda özelliğe sahip olması yüksek bir boyut oluşturmaktadır [6]. Gen ekspresyon verilerinin boyutsal üstünlüğü, makine öğrenmesinde boyutluluğun laneti olarak bilinmektedir [7]. Diğer yandan yüksek boyut, karmaşıklık terimi ile de ilişkilendirilebilmektedir. Çünkü hastalık, belirli genlerin alt kümesinden kaynaklanmaktadır. Dolayısıyla incelenen tüm genler, tümörün tanıtılmasında ayırıcı bir rol oynamamakta ve bunun ile birlikte karmaşıklığa yol açarak yapılan analizlerin doğruluğunu etkilemektedir. Tüm genler içerisinden tümörle ilgili bilgi veren doğru genlerin titizlikle seçilmesi önemli bir ölçüttür. Bu süreçte gerçekleştirilecek optimizasyon, analiz kısmı için önemli bir işlemdir ve sınıflandırma yöntemlerinin uygulanılmasından önce hastalığı ilgilendiren genler arasındaki örüntüleri ortaya çıkarmaktadır. Aynı zamanda veri kümesinden hastalık ile ilişkisi olmayan genlerin uzaklaştırılması da işlem yükünün düşmesini ve zamandan tasarruf edilmesini sağlamaktadır [6],[7].

Mikrodizi veri kümesi üzerinde bilişim destekli birçok çalışma gerçekleştirilmiştir. Bu kapsamda [3] çalışmasında, mikrodizi veri kümesi üzerinde kanserli gen tahmininin sağlanılabilmesi amacıyla özellik seçimi ve destek vektör makineleri yaklaşımlarının birlikte kullanıldığı aktif öğrenme modeli sunan simetrik belirsizlik isimli güçlü bir algoritma geliştirilmiştir. Mikrodizi tabanlı veriler üzerinde genetik algoritma ve destek vektör makinelerinin kombinasyonunun kullanıldığı [8] çalışmasında, kanser teşhisi yapılmıştır. Mikrodizi tabanlı veriler kullanılarak temel bileşenler analizi ile lojistik regresyon yöntemi üzerinde temellenen bir yaklaşım [9] çalışmasında sunulmuştur. Bu çalışmada sınıflandırma işlemi için uygun genlerin seçilmesi sağlandıktan sonra lösemi ile lenfoma veri kümeleri üzerinde deneysel işlemler gerçekleştirilmiştir. Mikrodizi tabanlı veriler üzerinde ilişkili genlerin elde edilmesi amacıyla parçacık sürü optimizasyon tekniğinin kullanıldığı [7] çalışmasında, çok sınıflı kanserlerin ayrımı gerçekleştirilmiştir. Yarı denetimli Ellipsoid ARTMAP sinir ağı yaklaşımı, NCI60 ve ALL veri kümeleri üzerinde test işlemleri gerçekleştirilmiştir. Mikrodizi tabanlı veri kümeleri üzerinden gen ifadelerinin genomik niteliklerini öğrenebilme gücüne sahip yeni bir evrişimsel sinir ağının tasarlandığı [6] çalışmasında, çoklu doku kanseri sınıflandırıcısı önerilmiştir. Lösemi ve prostat kanserine sahip hastaların mikrodizi gen ekspresyon verilerinden kansere ilişkin genlerin tanımlanabilmesinde SVST(Support Vector Sampling Technique) ile RFGS(Random Forest Gene Selection) yöntemleri [10] çalışmasında önerilmiştir. Bu yöntemlerin ilişkili genlerin tanımlanmasında başarılı olduğu ifade edilmiştir. Mikrodizi veri kümeleri üzerinde kansere ilişkin biyobelirteçlerin tanımlanabilmesi için [11] çalışmasında KFRS (Kernelized Fuzzy Rough Set) yaklaşımı önerilmiştir. Seçilen biyobelirteçlerin yardımıyla TSVM yöntemi kapsamında SRBCT, DLBCL, Leukemia ve MicroRNA veri kümelerinde yer alan kategorilerin sınıflandırılması sağlanmıştır. Genlerin birlikte ifade edilmesine imkân sağlayan mikrodizi teknolojisi ile elde edilen veriler üzerinde küme yoğunluğunu ve ayrımını mümkün kılan çok amaçlı kümeleme tekniği, MOCSVMEN (MultiObjective Clustering with Support Vector Machine based ensemble) [12] çalışmasında önerilmiştir. Bu tekniğin biyolojik olarak önemli gen kümelerinin ayırt edilmesinde güçlü bir potansiyel barındırdığı belirtilmiştir. Yüksek miktarda boyutsallık sunan mikrodizi verileri üzerinde optimal genlerin seçilebilmesi amacıyla parçacık sürü algoritması üzerinde temellenen bulanık etkileşimli bir optimizasyon yöntemi [13] çalışmasında önerilmiştir. Uygun sınıflandırıcılar ile gerçekleştirilen işlemler sonucunda yüksek potansiyele sahip tanıma oranlarının üretildiği ifade edilmiştir. Mikrodizi tabanlı gen ekspresyon verilerinin öznitelik sayıları ile örnek sayıları arasındaki mevcut dengesizlik nedeni ile öznitelik seçimine ilişkin bir yaklaşım [14] çalışmasında önerilmiştir. Önerilen yaklaşım kapsamında kolon tümörü, CNS tümörü, DLBCL, Leukemia (3 farklı veri kümesi), akciğer kanseri ve göğüs kanseri veri kümeleri(toplam 7 veri kümesi) üzerinde NB (Naive Bayes), DT (Decision Tree), SVM (Support Vector Machine) ve KNN (K-nearest Neighbour) algoritmaları kullanılarak uygun parametreler doğrultusunda kanser sınıflarının ayrımı sağlanmıştır. [15] çalışmasında, mikrodizi veri kümeleri kullanılarak parametre optimizasyonunun ya da güçlü HBE (Hyper-Box Enclosure) algoritmasının uygulanmasıyla elde edilen az sayıdaki gen biyobelirteci üzerinde Leukemia, Prostate Cancer, Prostate outcome, DLBCL, Lymphoma ve SRBCT veri kümelerindeki kategorilerin ayırt edilmesi gerçekleştirilmiştir. [16] çalışmasında, lösemi hastalığına ait mikrodizi veri kümesi içerisinde belirlenen merkez noktalara en yakın tümörlerin kümelenmesi gerçekleştirilerek kendi kendini organize eden haritalar(SOM) ile sınıflandırma süreci sağlanmıştır.

Bu çerçevede yapılan çalışmalar incelendiğinde, genom dizileme teknolojisindeki gelişmeler ile birlikte hastalıkların erken teşhisine ve prognozuna yönelik yapılan araştırmalar için biyolojik sistemlere ilişkin inceliklerin keşfedilmesinin önem taşıdığı görülmektedir. Yapılan analizler, biyolojik süreçlerdeki ilişkilerin ortaya çıkarılmasında bilgisayar destekli, verimli ve uygun maliyetli yaklaşımların geliştirilmesine ortam hazırlamaktadır. Bu çalışma kapsamında temel hedef, mikrodizi veri kümesindeki hastalıkla ilişkili genlerin seçilmesinin ardından akut lösemi hastalığının alt türlerine ilişkin malignitelerin sınıflandırılmasında derin öğrenme tabanlı mimarilerin özellik seçim algoritmaları ile hibrit kullanımı sağlanarak güçlü bir çıkarım sürecinin gerçekleştirilmesidir.

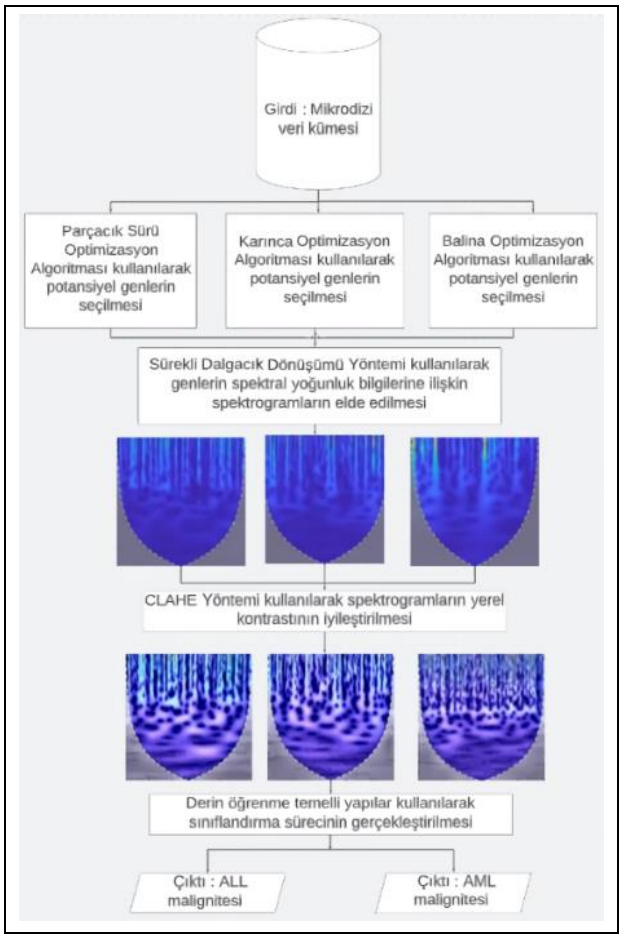
2 Ön işleme aşaması

2.1. Özellik seçim yöntemleri ile optimizaston

Optimizasyon, karşılaşılan problemler karşısında daha doğru, daha hızlı ve daha düşük maliyetler sunan sonuçlar üretmek amacıyla tercih edilen bir yaklaşımdır. Karar verme kalitesinin yükseltilmesinde yardımcı olan bu yaklaşım, gerçek hayat problemlerinde sıklıkla tercih edilmektedir [17].

Bu çalışmada, farklı birçok genin nispi ekspresyon seviyesi üzerinden moleküler tanı imkânı sunan mikrodizi teknolojisi ile elde edilen yüksek boyutlu veri kümeleri analiz edilmiştir. Temel amaç, veri kümeleri içerisinde yer alan farklı genler arasında hastalıkla ilişkili bilgili genlerin seçilmesidir. Böylece, işlem maliyetinin azaltılması, zamanın verimli kullanılması ve isabetli çıktıların üretilebilmesi sağlanacaktır. Bu doğrultuda, Parçacık Sürü, Karınca ve Balina Optimizasyon Algoritmaları kullanılmıştır. Doğadan ilham alınarak tasarlanmış mevcut 3 algoritma, sürü zekasına dayalı olarak geliştirilmiştir.

Bu kapsamda parçacık sürü optimizasyon algoritması; kuş sürülerinin yiyecek arama davranışından, karınca kolonisi optimizasyon algoritması; karınca kolonilerinde yer alan karıncaların yiyecek arama davranışından, balina optimizasyon algoritması; kambur balinaların kabarcık-ağ avlanma stratejilerinden esinlenerek tasarlanmıştır. Böylece sınırlı yeteneklere sahip canlıların biraraya gelmesi ile oluşan matematiksel model, zor problemlerin çözümünde kolaylık oluşturmaktadır. Öte yandan doğadan esinlenerek tasarlandığı için anlaşılması ve analiz edilmesi kolay yapılardır[17]. Bu yüzden mikrodizi veri kümeleri içerisinden hastalık ile ilişkili potansiyel genlerin seçilmesinde uygun ve uyarlanabilir bir akış gerçekleşmektedir. Kurgulanılan çalışmanın tüm aşamalarını içeren akış şeması Şekil 1'de verilmiştir.



Şekil 1. Çalışmanın akış şeması.

Figure 1. Flow chart of study.

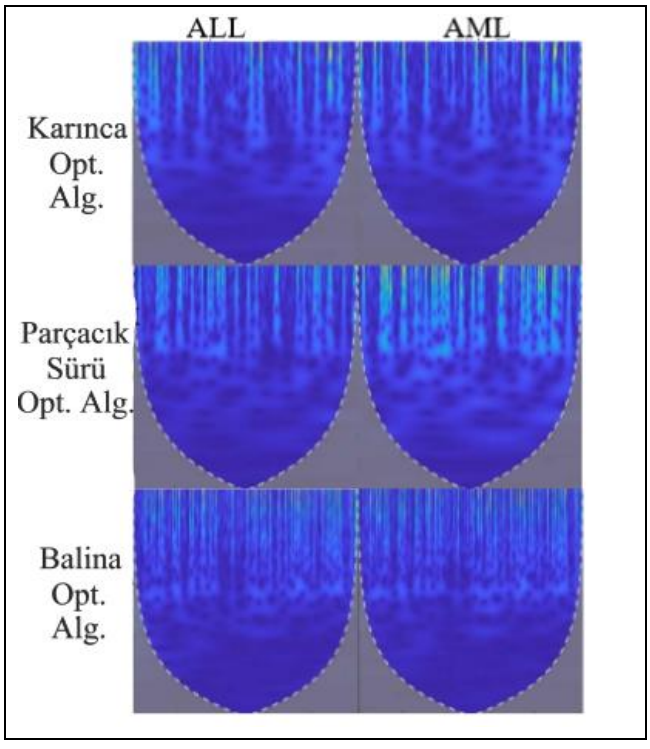
Şekil 1'de verilen akış diyagramı, doğadan ilham alarak tasarlanan sürü tabanlı 3 farklı algoritma ile seçilen genlere ilişkin spektral yoğunluk bilgisinin sayısal sinyal işleme yöntemi olan sürekli dalgacık dönüşümü yöntemi ile analiz edildiğini ve ardından CLAHE yöntemi ile yerel kontrastın iyileştirildiği spektrogramların derin öğrenme tabanlı mimariler ile sınıflandırılmasına ilişkin bir yapının olduğunu tasvir etmektedir.

3 Sayısal sinyal işleme yöntemi

Sinyal; frekans ve genlik bileşenlerinden meydana gelen bir işarettir. Bu işarete matematiksel dönüşüm yöntemleri uygulanarak frekans ve genlik bilgileri elde edilmektedir. Sinyal işleme disiplini ile elde edilen genlik ve frekans bilgileri, seçilen potansiyel genlere ilişkin bir analiz sağlamaktadır. Çünkü sinyalin içerisinde yer alan frekans bileşenlerinin zaman ve genlik bilgilerine ilişkin bir değerlendirme yapmak mümkündür[18]. Bu çalışmada dalgacık disiplini kullanılarak bir araştırma yapılmıştır.

Dalgacık dönüşümü, değişken zaman-frekans çözünürlüğü gösteren yapılardır[18]. Sunulan çalışmada seçilen potansiyel genlere ilişkin özelliklere dalgacık dönüşümü yönteminin sürekli dalgacık dönüşümü türü uygulanmıştır. Böylece bilgi veren kritik genlerin spektral yoğunluk bilgileri spektrogramlar ile ifade edilmiş ve yorumlanabilir bir çıktı elde edilmiştir.

Farklı uygulama alanlarına sahip olan dalgacık dönüşümünün sürekli dalgacık dönüşümü türüne ilişkin matematiksel ifadeleri [19]'da verilmiştir. Bu çalışmada seçilen genlere uygulanılan sürekli dalgacık dönüşümü yöntemi ile elde edilen spektrogramlar Şekil 2'de sunulmuştur.



Şekil 2. Optimizasyon yöntemleri kullanılarak elde edilen potansiyel genlerin spektrogramları.

Figure 2. Spectrograms of potential genes obtained using optimization methods.

Şekil 2'de gösterilen spektrogramlardaki spektral yoğunluk bilgilerinin belirginleştirilmesi için CLAHE yöntemi kullanılmıştır.

4 Görüntü iyileştirme süreci

4.1 CLAHE yöntemi

CLAHE yöntemi; görüntüyü parçalara ayıran, ardından her bir parçanın kontrastını histogram eşitleme yöntemi ile geliştiren ve son olarak tüm parçaları birleştirerek kontrast sınırlı yapı sunan bir yaklaşımdır. CLAHE yöntemi ile görüntünün yerel kontrastı, gürültünün etkisi arttırılmadan iyileştirilmektedir[20].

Tıp alanında farklı çalışmalarda da kullanılan CLAHE yönteminin [21] bu çalışmadaki nihai sonuca sağladığı pozitif katkı değerlendirilmiştir.

5 Metodoloji

5.1 Mikrodizi veri kümesi

Vücudun farklı bölgelerinde oluşum gösteren kanser hastalığının tanısal ve prognostik tahmini için mallignitelerin sınıflandırılması önemlidir[11]. Bu doğrultuda https://file.biolab.si/biolab/supp/bi-cancer/projections/ sitesinden temin edilen veriler kullanılarak akut lösemilerin sınıflandırılması hedeflenmiştir. Mevcut çalışmada kullanılacak veri kümesi aşağıda açıklanmıştır.

Leukemia Veri kümesi: 5147 gen ve 72 farklı örnek barındırmaktadır. 2 farklı lösemi türünün sınıflandırılması amacı ile oluşturulan veri kümesinde 47 adet akut lenfoblastik lösemi örneği ve 25 adet akut miyeloid lösemi örneği mevcuttur [22].

Leukemia veri kümesinde yer alan veriler kullanılarak ALL ve AML malignitelerinin bilgisayar destekli sistemler ile ayırt edilmesi hedeflenmiştir.

5.2 Sınıflandırma aşaması

Yapay zekâ insanların sahip oldukları bilişsel fonksiyonların makineler tarafından gerçekleştirilmesini amaçlayan geniş bir yelpaze sunar. Sağlık alanında gerçekleştirilen çalışmalarda öne çıkan yapay zekâ; klinik tanıların konulmasında, görüntü tabanlı tıbbi teşhislerin sağlanmasında ve karmaşık genomik verilerin işlenmesinde tercih edilmektedir [23].

Bu çalışmada, yapay zekâ çerçevesinde değerlendirilen klasik makine öğrenme algoritmaları, aktarım öğrenme mimarileri ve derin graf evrişimsel sinir ağı yaklaşımının özellik seçim algoritmaları ile hibrit kullanımı sağlanarak moleküler analiz yapılmıştır.

5.2.1 Klasik makine öğrenme algoritmaları

Makine öğrenimi, bilgisayar destekli sistemler kullanılarak hedef görevin verimli bir şekilde sağlanılabilmesi için faydalanılan bir çerçevedir. Bu doğrultuda ilk olarak göreve ilişkin mevcut tüm veriler toplanır. Ardından, veriler içerisindeki gizli bilgilerin çıkarılmasında faydalanılan matematiksel modeller ile bilgisayarın öğrenmesi gerçekleştirilir. Bu süreçte, veriler üzerinde yapılan analizler ile performansın geliştirilebilme durumu incelenir. Makine öğrenimi algoritmaları; sağlık hizmetleri, karar verme, tahmin yapma ve performans optimizasyonu gibi farklı amaçlar doğrultusunda kullanılmaktadır [24].

Bu çalışmada klasik makine öğrenmesi teknikleri arasında yer alan en yakın komşu algoritması(KNN), karar ağacı algoritması, rastgele orman algoritması ve destek vektör makineleri(SVM) algoritması kullanılmıştır. Veri kümesinin; %80'i eğitim kümesi ve %20'si test kümesi olarak seçilmiştir. Eğitim sürecinde eğitim ve test kümelerindeki verilerin belirli bir düzende karıştırılması için random_state fonksiyonu 42 olarak belirlenmiştir.

5.2.2 Aktarım öğrenme yöntemi

Bir problemin eğitiminde kazanılan bilgilerin farklı bir problem kapsamında kullanıldığı makine öğrenme tekniğidir[25],[26]. Evrişimsel sinir ağının eğitimi sırasında; zaman, donanım ya da veri miktarının azlığı nedeniyle modelin yeterince eğitilememesi durumuna karşı bir çözüm sunar [27]. Bu işlem sırasında model sıfırdan üretilmez. Eğitilen ağdaki son birkaç katman kaldırılır ve ardından hedef görev için yeni katmanlar eklenir [25].

Bu çalışmada VGG19, Inceptionv3, MobilenetV2, DenseNet201 ve Resnet150 aktarım öğrenme mimarilerinden faydalanılmıştır. Imagenet veri kümesi ile eğitilen bu mimarilerden elde edilen özellikler ve ağırlıklar akut lösemilere ilişkin malignitelerin sınıflandırılması aşamasında kullanılmıştır. Ancak derin öğrenme yaparken parametreler çoğaldıkça tamamen bağlı katmanda çok fazla ağırlık meydana gelmektedir. Bilgisayarın hesaplama gücünü zorlamasının yanı sıra modelin aşırı öğrenmeye neden olması nedeni ile ağa Global AveragePooling katmanı eklenerek her kanalın ortalama değerinin alınması sağlanmıştır. Böylece verinin boyutunu azaltan güncel bir mimari oluşturulmuştur. Aynı zamanda aşırı öğrenme problemini önlemek amacıyla dropout katmanı da eklenmiştir. 75 iterasyon boyunca eğitilen bu ağda kullanılan veri kümesinin %80'i eğitim kümesi ve %20'si test kümesi olarak seçilmiştir. Her bir eğitim sürecinde eğitim ve test kümelerindeki verilerin belirli bir düzende karıştırılması için random_state fonksiyonu 42 olarak seçilmiştir.

Ek olarak, eğitim sırasında parametreler üzerinde ince ayar yapılarak her bir yineleme sırasında kayıp değeri iyileştirilmiştir. Böylece minimum kayıp değerine ve maksimum başarı oranına ulaşılması hedeflenmiştir. Kayıp fonksiyonu, ikili çapraz entropi olarak seçilmiştir. Fonksiyonun matematiksel ifadesi Denklem 1'de verilmektedir[28].

$$BCE(h,t) = -(h * \log(t) + (1-h) * \log(1-t)) \quad (1)$$

Denklem 1'de h, hedef vektörünü; t, tahmini yapılan sınıf vektörünü göstermektedir. Faydalanılan kayıp fonsiyonunun yanı sıra kullanılan tüm aktarım öğrenme mimarileri hata oranının minimuma indirgenmesi amacıyla yapılan ince ayar ile 0.00007 öğrenme değerine sahip RMSprop optimize edicisi, 0.7 momentum ve 1e-07 epsilon değeri tercih edilerek oluşturulmuştur. ALL ve AML şeklinde 2 farklı çıktı arasında nihai karar verileceği için aktivasyon fonksiyonu olarak sigmoid kullanılmıştır. Yeni modelin eğitilmesi için gerçekleştirilen ince ayar ile daha başarılı ve tutarlı bir eğitim süreci gerçekleştirilmiştir.

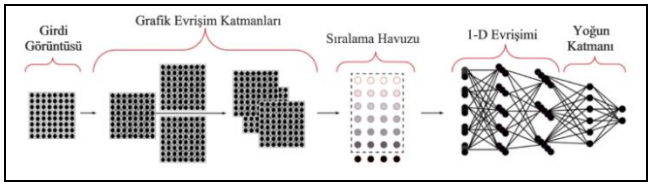
5.2.3 Derin graf evrişimsel sinir ağı

Graf temelli sinir ağları, tensör formları üzerinde yüksek seviyeli özellikleri çıkartabilen dinamik bir yapıya sahiptir. Biyolojik ağlar, sosyal ağlar ya da moleküler yapılar gibi birçok çalışmada tercih edilmektedir [29].

Bu çalışmada geleneksel sinir ağı teknolojisi(CNN) ile graf temelli sinir ağı teknolosinin(GNN) hibrit kullanımına imkân tanıyan DGCNN yaklaşımı kullanılmıştır. DGCNN yapısı 3 temel kısımdan meydana gelmektedir. İlk kısımda graf evrişim katmanları vasıtasıyla yerel altyapı özelliklerinin çıkarıldığı köşeler için sıralama yapılır. Ardından geleneksel sinir ağı katmanları ile graf evrişim katmanlarının birleştirilmesinde

önemli rol oynayan sıralama havuzu katmanı kullanılarak önceden tanımlanan sırada köşe özelliklerinin sıralanması gerçekleştirilir. Son adımda, geleneksel evrişimsel sinir ağlarının sıralı tensör görüntüleri üzerinden çıkarım yapması ile yeni tahminlerin üretilmesi sağlanır. Çalışmada sunulan mimarinin temsili yapısı Şekil 3'te tasvir edilmektedir [29].

Önerilen DGCNN modeli, spatial-temporal graf temelli bir yapıdır. Graf evrişim katmanları ile 1 boyutlu CNN yapılarının ard arda modellenmesi ile oluşur. Burada amaçlanan graf evrişim katmanları ile uzaysal bağımlılığın ortaya çıkartılabilmesini sağlamak ve görüntü üzerinde ilerlerken 1 boyutlu CNN katmanları ile zamansal bağımlılığı yakalamaktır. [30].



Şekil 3. DGCNN mimarisinin temsili yapısı[29].

Figure 3. Representative structure of the DGCNN architecture.

Yapay zekâ temelli DGCNN yaklaşımının kullanımı ile gerçekleştirilen sınıflandırma sürecinin ilk adımında uygulanılan GCN yapısı ile ilişkiye duyarlı temsil(relation-aware representation (RAR)) elde edilir. Ardından sıralama havuzu katmanına iletilen girdi bir sonraki adımda kullanılacak geleneksel evrişimli sinir ağlarının köşelerini tutarlı bir sırada okuyabilmesi ve eğitebilmesi için çıktı sağlamaktadır. Son olarak geleneksel sinir ağı yöntemlerinin gerçekleştirildiği otomatik özellik çıkarma süreci sayesinde görüntü seviyesinde temsil (image-level representation (IIR)) bilgisi elde edilmektedir. RAR, IIR bilgilerinin açığa çıkarılmasının yanı sıra köşe düğümlerin temsil edildiği bir sıralamanın mevcut olması algoritmanın güçlü yanlarını oluşturmaktadır [29].

Bu çalışmada, 4 katmanlı GNN ve 2 katmanlı CNN sinir ağı mimarisi kullanılarak graf temsilleri üzerinden tahminler yapılmıştır. Aşırı öğrenme probleminin önlenmesi amacıyla dropout katmanı eklenerek 100 iterasyon boyunca modelin eğitilmesi sağlanmıştır.

6 Tartışma ve Sonuçlar

Bu makalede akut lösemi hastalığının temel türlerine ilişkin malignitelerin tespit edilmesi için gerçekleştirilen çalışma sonucunda 7 önemli nokta güçlendirilmiştir.

1. Kanser, hücrelerin kontrollü bölünme yeteneklerini kaybetmeleri sonucunda ortaya çıkan bir malignite olduğu için hastalığın tespitinde geleneksel yöntemler kullanılarak kansere neden olan tümörler üzerinde klinik gözlemler ve morfolojik görünümler vasıtasıyla elde edilen parametreler bazen belirsizlik içermektedir. Doktorlar tarafından verilecek nihai kararın net çıktılar ile desteklenmesi amacıyla genlerle ilişkili kanser hastalıkları için moleküler tanı imkânı sunan mikrodizi teknolojisi kullanılmıştır,
2. Kanser hastalıklarının tanı sürecinde umut kaynağı olan mikrodizi teknolojisi [31] genleri temsil eden özellikler üzerinden analiz imkanı sağlamaktadır. Bu durum mikrodizi veri kümelerinin yüksek boyutlu olmasına neden olmuştur. Mikrodizi veri kümelerindeki hastalıkla ilişkili genlerin hastalıkla ilişkisiz genler içerisinden titizlikle ayırt edilebilmesi amacıyla özellik seçim algoritmaları ile uygun genlerin seçilmesi sağlanmıştır. Böylece işlem maliyeti ve çalışma zamanı azaltılmıştır,
3. Uygun genlerin seçilmesinde nitelikli kararlar alabilmek ve efektif bir çıktı üretebilmek için farklı metasezgisel algoritmalar kullanılmış ve karşılaştırmalar yapılmıştır,
4. Seçilen potansiyel genlerin spektral yoğunluk bilgisinden faydalanılmıştır. Bunun için sürekli dalgacık dönüşümü yöntemi kullanılmıştır. Böylece elde edilen spektrogramlar ile frekans ölçeğinde ortalama gücün ifade edilmesi sağlanmıştır,
5. Seçilen genlerin spektral yoğunlukları CLAHE yöntemi ile belirginleştirilmiştir. Bu yöntem ile görüntünün yerel kontrastı, gürültünün etkisi arttırılmadan iyileştirilmiştir[20],
6. Başarımların güçlü sınıflandırma teknikleri kapsamında değerlendirilmesi sürecinde ilk olarak aktarım öğrenme mimarisi tercih edilmiştir. Evrişimsel sinir ağının eğitimi sırasında zaman, donanım ya da veri miktarının azlığı nedeniyle modelin yeterince eğitilememesi durumuna karşı bir çözüm sunan bu yaklaşım ile modelin sıfırdan üretilmesine gerek kalmadan malignitelerin sınıflandırılması amaçlanmıştır[27]. Sınıflandırmada tercih edilen diğer yöntem ise DGCNN yaklaşımıdır. İlişkiye duyarlı temsil ve görüntü seviyesinde temsil bilgilerinin açığa çıkartıldığı bu yaklaşım, hem uzaysal hem de zamansal bilgi çıkarımı ile güçlü bir eğitim süreci gerçekleştirmektedir [29],
7. Uzmanlara fikir vermek ve efektif çıktılar üretebilme gayesi ile bilgisayar destekli bir sistemin inşa edilmesi sağlanmıştır.

Özetle, bu çalışmada lösemi veri kümesi üzerinde sırasıyla karınca, parçacık sürü ve balina optimizasyon algoritmaları kullanılarak uygun genler seçilmiştir. İlk safhada seçilen genler, klasik makine öğrenmesi algoritmaları ile değerlendirilmiştir. İkinci safhada seçilen genler, sürekli dalgacık dönüşümü metodu ile spektrogramlar olarak ifade edilmiştir. Ardından CLAHE yöntemi ile yerel kontrastı iyileştirilen spektrogramlar, aktarım öğrenme mimarileri ve DGCNN yaklaşımı ile sınıflandırılmıştır. Klasik makine öğrenme algoritmaları ile aktarım öğrenme mimarilerine göre geleneksel sinir ağı teknolojisi (CNN) ve graf temelli sinir ağı teknolosinin (GNN) hibrit kullanımına imkân tanıyan DGCNN yaklaşımı, daha başarılı çıktılar üretmiştir. Sınıflandırma mimarileri kullanılarak ulaşılan deneysel ölçütler Tablo 1, 2 ve 3'te sunulmuştur.

Lösemi veri kümesi üzerinde elde edilen deneysel ölçütler incelenmiştir. İlk aşamada makine öğrenme çerçevesinde değerlendirilen KNN Algoritması, Karar Ağacı Algoritması, Rastgele Orman Algoritması ve SVM Algoritması kullanılmıştır. Etiketli veriler üzerinde gerçekleştirilen eğitim süreci sonucunda her iki sınıf kapsamında da doğru tanımlamalar yapılarak ortalama başarı oranı elde edilmiştir. İkinci safhada derin öğrenme temelli yaklaşımlar değerlendirilmiştir. Bu doğrultuda VGG19 ve Inceptionv3 aktarım öğrenme mimarileri kullanılarak 75 iterasyon ile gerçekleştirilen eğitim süreci boyunca ulaşılan başarı oranları incelenmiş ve mimarilerin nitelikli bir çıktı üretemediği görülmüştür.

Tablo 1. Karınca optimizasyon algoritması kullanılarak seçilen potansiyel genlerin sınıflandırma başarımına etkisini gösteren deneysel kriterler.

Table 1. Experimental criteria showing the effect on classification performance.of potential genes selected using the ant optimization algorithm.

Yöntemler	BO.	Dyr.	Özg.	Ksn.	F Kr.
VGG19 Mimarisi	%73.3	0	%100	0	0
Inceptionv3 Mimarisi	%73.3	0	%100	0	0
MobilenetV2 Mimarisi	%66.6	%75	%63.6	%42.8	%54.5
DenseNet201 Mimarisi	%73.3	%100	%63.6	%50	%66.6
Resnet150 Mimarisi	%80	%75	%81.8	%60	%66.6
DGCNN Yaklaşımı	**%93.3**	%80	%100	%100	%88.8
KNN Algoritması	%80	%100	%72.7	%57.1	%72.7
Karar Ağacı Algoritması	%80	%50	%90.9	%66.6	%57.1
Rastgele Orman Algoritması	%73.3	%25	%90.9	%50	%33.3
SVM Algoritması	%86.6	%75	%90.9	%75	%75

BO: Başarı oranı, Dyr: Duyarlılık, Özg: Özgüllük, Ksn: Kesinlik, F Kr: F Kriteri.

Tablo 2. Parçacık sürü optimizasyon algoritması kullanılarak seçilen potansiyel genlerin sınıflandırma başarımına etkisini gösteren deneysel kriterler.

Table 2. Experimental criteria showing the effect on classification performance.of potential genes selected using the particle swarm optimization algorithm.

Yöntemler	BO.	Dyr.	Özg.	Ksn.	F Kr.
VGG19 Mimarisi	%73.3	0	%100	0	0
Inceptionv3 Mimarisi	%33.3	%100	%09.0	%28.5	%44.4
MobilenetV2 Mimarisi	%66.6	%100	%54.5	%44.4	%61.5
DenseNet201 Mimarisi	%33.3	%100	%09.0	%28.5	%44.4
Resnet150 Mimarisi	%73.3	%50	%81.8	%50	%50
DGCNN Yaklaşımı	**%86.6**	%80	%90	%80	%80
KNN Algoritması	%80	%75	%81.8	%60	%66.6
Karar Ağacı Algoritması	%80	%50	%90.9	%66.6	%57.1
Rastgele Orman Algoritması	%73.3	%25	%90.9	%50	%33.3
SVM Algoritması	%86.6	%75	%90.9	%75	%75

BO: Başarı oranı, Dyr: Duyarlılık, Özg:Özgüllük, Ksn: Kesinlik, F Kr: F Kriteri

Tablo 3. Balina optimizasyon algoritması kullanılarak seçilen potansiyel genlerin sınıflandırma başarımına etkisini gösteren deneysel kriterler.

Table 3. Experimental criteria showing the effect on classification performance.of potential genes selected using the whale optimization algorithm.

Yöntemler	BO.	Dyr.	Özg.	Ksn.	F Kr.
VGG19 Mimarisi	%73.3	0	%100	0	0
Inceptionv3 Mimarisi	%26.6	%100	0	%26.6	%42.1
MobilenetV2 Mimarisi	%80	%100	%72.7	%57.1	%72.7
DenseNet201 Mimarisi	%46.6	%100	%27.2	%33.3	%50.01
Resnet150 Mimarisi	%53.3	%100	%36.3	%36.3	%53.3
DGCNN Yaklaşımı	%86.6	%60	%100	%100	%75
KNN Algoritması	%73.3	%50	%81.8	%50	%50
Karar Ağacı Algoritması	%80	%50	%90.9	%66.6	%57.1
Rastgele Orman Algoritması	%73.3	%25	%90.9	%50	%33.3
SVM Algoritması	%80	%75	%90.9	%60	%66.6

BO: Başarı Oranı, Dyr: Duyarlılık, Özg:Özgüllük, Ksn: Kesinlik, F Kr: F Kriteri.

Bu durum, ALL ve AML malignitelerine ilişkin spektrogramlar için özellikleri ezberleyen modelin aşırı öğrenmeye neden olduğunu açıklamaktadır. Bu iki aktarım öğrenme mimarilerinin konfüzyon matrisleri incelendiğinde tüm görüntülerin ya da görüntülerin yaklaşık %90'ının ALL sınıfı olarak ya da AML sınıfı olarak tahmin edildiği görülmektedir. Dolayısıyla VGG19 ve InceptionV3 mimarileri ile gerçekleştirilen çalışmaların sonuçları güven vermemektedir. Bununla birlikte MobileNetV2, DenseNet201 ve Resnet150 aktarım öğrenme mimarileri ile elde edilen sonuçlar incelendiğinde modelin sadece bir sınıfı tanımasının aksine her iki sınıf için de doğru ve yanlış tahminler ürettiği görülmektedir. Bu üç aktarım öğrenme mimarilerinin konfüzyon matrisleri üzerinden hesaplanan duyarlılık, özgüllük, kesinlik ve F kriterleri incelendiğinde ulaşılan sonuçlar, ALL ve AML spektrogramlarının tahmin edilme başarısını şu şekilde açıklamaktadır. Imagenet veri kümesi kullanılarak eğitilen aktarım öğrenme mimarileri ile elde edilen özelliklerin ya da ağırlıkların yeni bir hedef için genel bir kaynak doğrultusunda yeniden kullanıldığında daha başarılı olduğu bilinmekte iken spesifik bir kaynak olan akut lösemilere ilişkin spektrogramların sınıflandırılmasında çok başarılı

çıktılar üretememe ihtimali mevcuttur. Bu nedenle titizlikle yapılan ince ayar sonucunda mevcut 3 mimaride maksimum %80 başarı oranına ulaşıldığı görülmektedir. Dolayısıyla aktarım öğrenme yaklaşımlarının mevcut çalışma için çok uygun olmadığı görülmüştür.

Öte yandan diğer bir derin öğrenme temelli yapı olan DGCNN yaklaşımı doğrultusunda elde edilen sonuçlar ile daha başarılı çıktılara ulaşılmıştır. Geleneksel sinir ağı teknolojisi(CNN) ve graf temelli sinir ağı teknolosinin(GNN) hibrit kullanımına imkân tanıyan DGCNN teknolojisi ile gerçekleştirilen eğitim sonucunda graf evrişim katmanları kullanılarak uzaysal bağımlılığın ortaya çıkartılabilmesi ve görüntü üzerinde ilerlerken 1 boyutlu CNN katmanları kullanılarak zamansal bağımlılığın ortaya çıkartılabilmesi mümkün olmuştur. Aynı zamanda mesaj geçiren sinir ağının varlığı ile bilginin düğümler vasıtasıyla yayılması spesifik bir çalışmaya özgü olan spektrogramlar üzerinde de başarılı bir çıktı üretmiştir.

Sonuç olarak doğrusal karmaşıklık O(m) sunan bir çıktı üreten DGCNN yöntemi ile hesaplama maliyetinin veri kümesindeki artışına bağlı olarak doğrusal yükselmesi ve gerçek başarı oranının değerlendirilebilmesinde önemli olan duyarlılık, özgüllük, kesinlik ve F skoru kriterlerininde kabul edilebilir sonuçlar üretmesi, derin öğrenme temelli bu yöntemin akut lösemiye ilişkin spektrogramların sınıflandırılmasında uygun bir yaklaşım olduğuna karar verilmiştir.

Sınıflandırma sürecinin daha doğru ve daha efektif gerçekleştirilebilmesi için kullanılan özellik seçim yöntemleriyle gerçekleştirilen optimizasyon işlemi ve CLAHE yöntemi kullanılarak yerel kontrastı arttırılan spektrogramların spektral yoğunluk bilgilerinin netleştirilmesi sınıflandırmaya pozitif katkıda bulunmuştur. Bu doğrultuda Tablo 1, Tablo 2 ve Tablo 3 incelendiğinde hem aktarım öğrenme mimarileri hem de DGCNN yaklaşımı kapsamında elde edilen sonuçların karınca özellik seçim algoritması kullanılarak seçilen genler ile oluşturulan spektrogramlar üzerinde daha başarılı olduğu görülmektedir. Bu durum farklı birçok genin nispi ekspresyon seviyesi üzerinden oluşturulan mikrodizi veri kümesindeki doğru, bilgili ve araştırılan hastalıkla ilişkili potansiyel genlerin seçilebilmesi için tercih edilen özelik seçim yaklaşımları içerisinden karınca kolonisi algoritmasının daha başarılı olduğunu göstermektedir.

Leukeamia veri kümesinin kullanıldığı [3],[9],[13],[32]-[34], çalışmaları incelenmiştir. Genomik analiz imkanı sunan mikrodizi veri kümeleri üzerinde optimizasyon, sinyal işleme, görüntü iyileştirme ve hibrit sinir ağı yapıları kullanan bu çalışma ile moleküler analiz alanına farklı bir bakış açısı kazandırılmıştır.

7 Sonuçlar

Bu çalışmada, akut lösemi hastalığına ilişkin malignitelerin ayırt edilebilmesi amacıyla moleküler tanı imkânı sağlayan bir yaklaşım sunulmuştur. İlk olarak mikrodizi veri kümesi üzerinde 3 ayrı özellik seçim algoritması kullanılarak en uygun genlerin seçilmesi sağlanmış ve seçilen genler klasik makine öğrenmesi algoritmaları ile değerlendirilmiştir. İkinci olarak seçilen genlere uygulanılan sürekli dalgacık dönüşümü yöntemi sonucunda elde edilen spektrogramlar, CLAHE yöntemi ile iyileştirilmiştir. Üçüncü olarak yerel kontrastı iyileştirilen görüntüler derin öğrenme temelli yapılar olan aktarım öğrenme mimarileri ve DGCNN yaklaşımı ile sınıflandırılmıştır. DGCNN yaklaşımı ile gerçekleştirilen sınıflandırma işlemi, klasik makine öğrenme algoritmalarına ve aktarım öğrenme mimarilerine kıyasla daha başarılı çıktılar üretmiştir. Çünkü DGCNN yaklaşımı, uzaysal ve zamansal bağımlılığı ortaya çıkartabilen bir yapı olmasının yanı sıra ilişkisel ve görüntü düzeyinde temsiller sağlayarak daha güçlü çıkarımlar elde etmektedir. Aynı zamanda mesaj geçiren sinir ağının varlığı ile bilginin düğümler vasıtasıyla yayılması, spesifik bir çalışmaya özgü olan spektrogramlar üzerinde de başarılı sonuçlar üretmesini sağlamaktadır. Ek olarak DGCNN yaklaşımı, doğrusal karmaşıklık sunmaktadır. Bu durum değerlendirme sürecinin verimli bir şekilde ilerlemesini gerçekleştirmektedir. Genel yapısı itibariyle işlem maliyeti düşük olan ve kısa zamanda verimli bir akışın hedeflendiği bu çalışmanın literatüre katkı sağlayacağı düşünülmektedir. Gelecekte, farklı sezgisel algoritmalara ilişkin hibrit tasarımlar ile potansiyel gen seçiminde daha kaliteli çıktılar üretilmesi planlanmaktadır.

8 Conclusions

This study presents an approach that provides molecular diagnosis to distinguish malignancies related to acute leukaemia disease. First, the most suitable genes are selected by using 3 different feature selection algorithms on microarray dataset, and selected genes were evaluated with classical machine learning algorithms. Secondly, the spectrograms obtained as a result of the continuous wavelet transform method applied to the selected genes are improved by the CLAHE method. Thirdly, images with enhanced local contrast are classified by transfer learning architectures and the DGCNN approach, which are deep learning-based structures. The classification process performed with the DGCNN approach is provided more successful outputs compared to the classical machine learning algorithms and the transfer learning architectures. The DGCNN approach, besides being a structure that can reveal spatial and temporal dependence, provides stronger inferences by providing relational and image-level representations. At the same time, the existence of a neural network that transmits a message and the dissemination of information through nodes enabled it to produce successful results on spectra specific to a specific study. In addition, the DGCNN approach offers linear complexity. This situation will ensure that the evaluation process proceeds efficiently. It is thought that this study, which aims to achieve an efficient flow in a short time with low transaction costs in terms of its general structure, will contribute to the literature. In the future, it is planned to produce better quality outputs in potential gene selection by providing a hybrid design of different heuristic algorithms.

9 Yazar katkı beyanı

Gerçekleştirilen çalışmada Fatma AKALIN, fikrin oluşması, literatür taraması, makalenin yazılması ve sonuçların irdelenmesi hususunda; Nejat YUMUŞAK, fikrin oluşması, literatür taraması, yazım denetimi ve içeriğin kontrol edilmesi hususunda katkıda bulunmuştur.

10 Etik kurul onayı ve çıkar çatışması beyanı

Hazırlanan makalede etik kurul izni alınmasına gerek yoktur. Hazırlanan makalede herhangi bir kişi/kurum ile çıkar çatışması bulunmamaktadır.

11 Kaynaklar

[1] Kocabıyık VB. ALL ve KML'li Hastalarda BCR ve ABL Genlerindeki Mutasyonların İncelenmesi, Yüksek Lisans Tezi, Selçuk Üniversitesi, Konya, Türkiye, 2011.

[2] Jauhari S, Rizvi SAM. "Mining gene expression data focusing cancer therapeutics: A digest". *IEEE/ACM Transactions on Computational Biology and Bioinformatics Bioinforma,* 11(3), 533-547, 2014.
[3] Begum S, Sarkar R, Chakraborty D, Sen S, Maulik U. "Application of active learning in DNA microarray data for cancerous gene identification". *Expert Systems with Applications,* 177, 1-8, 2021.
[4] Yang R, Paparini A, Monis P, Ryan U. "Comparison of next-generation droplet digital PCR (ddPCR) with quantitative PCR (qPCR) for enumeration of Cryptosporidium oocysts in faecal samples". *International Journal for Parasitology*, 44(14), 1105-1113, 2014.
[5] Wang X, Simon R. "Microarray-based cancer prediction using single genes". *BMC Bioinformatics*, 12, 1-9, 2011.
[6] Khorshed T, Moustafa MN, Rafea A. "Learning visualizing genomic signatures of cancer tumors using deep neural networks". *Proceedings of the International Joint Conference on Neural Networks*, Glasgow, UK, 19-24 July, 2020.
[7] Xu R, Anagnostopoulos GC, Wunsch DC. "Multiclass cancer classification using semisupervised ellipsoid ARTMAP and particle swarm optimization with gene expression data". *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 4(1), 65-77, 2007.
[8] Peng S, Xu Q, Ling XB, Peng X, Du W, Chen L. "Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines". *FEBS Letters*, 555(2), 358-362, 2003.
[9] Ocampo-Vega R, Sanchez-Ante G, De Luna MA, Vega R, Falcón-Morales LE, Sossa H. "Improving pattern classification of DNA microarray data by using PCA and logistic regression". *Intelligent Data Analysis*, 20, 53-67, 2016.
[10] Chen AH, Tsau YW, Lin CH. "Novel methods to identify biologically relevant genes for leukemia and prostate cancer from gene expression profiles". *BMC Genomics*, 11, 1-21, 2010.
[11] Chakraborty D, Maulik U. "Identifying cancer biomarkers from microarray data using feature selection and semisupervised learning". *IEEE Journal of Translational Engineering in Health and Medicine*, 2, 1-11, 2014.
[12] Mukhopadhyay A, Maulik U, Bandyopadhyay S. "Gene expression data analysis using multiobjective clustering improved with SVM based ensemble". *In Silico Biology*, 11, 19-27, 2011.
[13] Chen Y, Zhao Y. "A novel ensemble of classifiers for microarray data classification". *Applied Soft Computing Journal,* 8(4), 1664-1669, 2008.
[14] Wang X, Gotoh O. "A robust gene selection method for microarray-based cancer classification". *Cancer Informatics*, 9, 15-30, 2010.
[15] Dagliyan O, Uney-Yuksektepe F, Kavakli IH, Turkay M. "Optimization based tumor classification from microarray gene expression data". *PLoS One*, 6(2), 1-10, 2011.
[16] Golub T, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. "Molecular classification of cancer: class discovery". *Science*, 286, 531-537, 1999.
[17] Doğan C. Balina Optimizasyon Algoritması ve Gri Kurt Optimizasyonu Algoritmaları Kullanılarak Yeni Hibrit Optimizasyon Algoritmalarının Geliştirilmesi, Yüksek Lisans Tezi, Erciyes Üniversitesi, Kayseri, Türkiye, 2019.
[18] Fidan H. Dalgacık Dönüşümü Tekniği ile Motor Arıza Tespiti, Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi, Isparta, Türkiye, 2006.
[19] Öner İV, Yeşilyurt K, Yılmaz EÇ. "Wavelet analiz tekniği ve uygulama alanları". *Ordu Üniversitesi Bilim ve Teknoloji Dergisi*, 7(1), 42-56, 2017.
[20] Aktürk SM. Grabcut Etkileşimli Bölütleme Yöntemi Üzerinde İyileştirme Çalışmaları, Yüksek Lisans Tezi, Karadeniz Teknik Üniversitesi, Trabzon, Türkiye, 2018.
[21] Akalın F, Yumuşak N. "DNA genom dizilimi üzerinde dijital sinyal işleme teknikleri kullanılarak elde edilen ekson ve intron bölgelerinin EfficientNetB7 mimarisi ile sınıflandırılması". *Journal of the Faculty of Engineering and Architecture of Gazi University*, 37(3), 1355-1372, 2022.
[22] Cancer Gene Expression Data Sets and Their Visualizations. "Data Set Name: Leukemia" https://file.biolab.si/biolab/supp/bi-cancer/projections/ (2022).
[23] Dias R, Torkamani A. "Artificial intelligence in clinical and genomic diagnostics". *Genome Medicine*, 11(1), 1-12, 2019.
[24] El Mrabet MA, El Makkaoui K, Faize A, "Supervised machine learning: a survey". *Proceedings 4th International Conference on Advanced Communication Technologies and Networking, CommNet 2021*, Rabat, Morocco, 03-05 December, 2021.
[25] Atila Ü, Uçar M, Akyol K, Uçar E. "Plant leaf disease classification using EfficientNet deep learning model". *Ecological Informatics,* 61, 1-13, 2021.
[26] Karahan T, Nabiyev V. "Plant identification with convolutional neural networks and transfer learning". *Pamukkale University Journal of Engineering Sciences*, 27(5), 638-645, 2021.
[27] Elmas B. "Identifying species of trees through bark images by convolutional neural networks with transfer learning method". *Journal of the Faculty of Engineering and Architecture of Gazi University*, 36(3), 1253-1269, 2021.
[28] Sreng S, Maneerat N, Hamamoto K, Win KY. "Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images". *Applied Sciences*, 10(14), 1-19, 2020.
[29] Zhang M, Cui Z, Neumann M, Chen Y. "An end-to-end deep learning architecture for graph classification". *The Thirty-Second AAAI Conference on Artificial Intelligence,* 32(1), 4438-4445, 2018.
[30] Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS. "A Comprehensive Survey on Graph Neural Networks". *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4-24, 2021.
[31] Xu R, Anagnostopoulos GC, Wunsch DC. "Multi-class cancer classification by semi-supervised ellipsoid ARTMAP with gene expression data". *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* San Francisco, CA, USA, 01-05 September, 2004.
[32] Wang X, Gotoh O. "Cancer classification using single genes". *Genome Informatics,* 23(1), 179-188, 2009.
[33] Ghorai S, Mukherjee A, Dutta PK. "Gene expression data classification by VVRKFA". *Procedia Technology*, 4, 330-335, 2014.
[34] Maulik U, Chakraborty D. "Fuzzy preference based feature selection and semisupervised SVM for cancer classification". *IEEE Transactions on Nanobioscience*, 13(2), 152-160, 2014.