



Negatif olmayan matris faktörizasyonuna dayalı LncRNA-Hastalık ilişkisi tahmini LncRNA-Disease association prediction based on nonnegative matrix factorization

Ahmet Toprak^{1,*} 

¹ Selçuk Üniversitesi, Elektrik ve Enerji Bölümü, Konya, Türkiye

Öz

Yapılan birçok biyolojik deneylerde lncRNA'nın çeşitli karmaşık insan hastalıklarının gelişimi ile yakından ilişkili olduğunu kanıtlanmıştır. Bu nedenle lncRNA ve hastalık arasındaki ilişkiyi bilmek sadece hastalık mekanizmasını anlamaya yardımcı olmakla kalmaz, aynı zamanda hastalığın teşhisini, tedavisini ve prognozunu da kolaylaştırır. Fakat lncRNA ile hastalık arasındaki ilişkiyi biyolojik deneyler yoluyla belirlemek hem maliyetli hem de çok zaman gerektiren bir süreçtir. Bu sebeple birçok araştırmacı potansiyel lncRNA-hastalık ilişkilerini tahmin etmek için hesaplamalı yöntemler geliştirmişlerdir. Bu çalışmada, fonksiyonel olarak benzer lncRNA'ların fenotipik olarak benzer hastalıklarla ilişki kurma eğiliminde olduğu varsayımına dayanarak, potansiyel lncRNA'ları tahmin etmek için NMF adı verilen bir hesaplama yöntemi öneriyoruz. Bu yöntem lncRNA ekspresyon benzerlik ağını, lncRNA kosinüs benzerlik ağını, hastalık semantik benzerlik ağını, hastalık kosinüs benzerlik ağını ve bilinen lncRNA-hastalık ilişkilendirme ağını entegre etmektedir. Yöntemimizin tahmin doğruluğunu göstermek için 5-katlı çapraz doğrulama ve birini dışarıda bırak çapraz doğrulama tekniklerini uyguladık ve ROC grafiklerini elde ettik. 5-katlı çapraz doğrulama için 0.7837 AUC değeri, birini dışarıda bırak çapraz doğrulama için 0.8551 AUC değeri hesaplandı. Sonuçlar NMF yönteminin güvenilir tahmin performansına sahip olduğunu göstermektedir.

Anahtar kelimeler: lncRNA, hastalık, lncRNA-hastalık ilişkisi, NMF

1 Giriş

Uzun kodlama yapmayan RNA'lar (lncRNA) 200 nükleotidden daha uzun bir uzunluğa sahip olup hücre döngüsü kontrolü, transkripsiyonel düzenleme, post-transkripsiyonel düzenleme ve kromatin modifikasyonu gibi çeşitli hücresel süreçlerde önemli düzenleyiciler olarak görev yapan bir RNA türüdür [1]. Anormal lncRNA ekspresyonu çok sayıda insan hastalığının başlamasına ve ilerlemesine neden olabilir [2]. Yapılan deneysel çalışmalarda lncRNA'ların mutasyonlarının ve düzensizliklerinin rahim ağzı kanseri [3], kolorektal kanser [4], yumurtalık kanseri [5], prostat kanseri [6] ve diyabet [7] dahil olmak üzere çeşitli hastalıklara neden olduğu açığa çıkmıştır. Örneğin lncRNA H19'un aşırı ekspresyonu meme kanseri hücrelerinin karsinogenezitesini arttırdığı

Abstract

Many biological experiments have proven that lncRNA is related to various complex human diseases. Therefore, knowing the lncRNA-disease relationships not only facilitates the diagnosis, treatment and prognosis of the disease helps to understand the disease mechanism. However, determining the lncRNA-disease relationships through biological experiments is both costly and time-consuming. For this reason, many researchers have suggested calculational methods to forecast potential relationships between lncRNAs and diseases. In this study, we suggest a computational method named NMF to forecast possible lncRNAs, based on the assumption that functionally similar lncRNAs tend to associate with phenotypically similar diseases. This method integrates the lncRNA expression similarity network, the lncRNA cosine similarity network, the disease semantic similarity network, the disease cosine similarity network, and the known lncRNA-disease relationship network. To demonstrate the prediction accuracy of our method, we applied 5-fold cross-validation and leave-out cross-validation techniques and obtained ROC plots. AUC of 0.7837 for 5-fold cross-validation and 0.8551 AUC for leave-out cross-validation were calculated. The results show that the NMF method has reliable prediction performance.

Keywords: lncRNA, disease, lncRNA-disease association, NMF

bilinmektedir [8]. Tümör baskılayıcı lncRNA GAS5'nin akciğer kanseri dokularında önemli ölçüde aşağı regüle edildiği [9] ve HULC'nin anormal ekspresyonu Hepatosellüler karsinom (HCC) ile ilişkili birkaç gende önemli anormalliklere neden olduğu görüşmüştür [10]. Bu sebeple lncRNA'lar ilgili kanserlerin erken teşhisi ve prognozu için biyobelirteç olarak kullanılabilir, bu da lncRNA'lar ve hastalıklar arasındaki ilişkilerin tanımlanmasını bir araştırma odağı haline getirmektedir. Hastalıkla ilişkili lncRNA'ların bulunması, lncRNA düzeyinde karsinogeneze sebep olan karmaşık mekanizmaların anlaşılmasına, hastalık teşhisi ve tedavisi için yeni prognostik belirteçlerin ortaya çıkarılması açısından hayati önem taşımaktadır. Yapılan deneysel çalışmalara rağmen doğrulanmış lncRNA-hastalık ilişkileri

* Sorumlu yazar / Corresponding author, e-posta / e-mail: atoprak@selcuk.edu.tr (A. Toprak)

Geliş / Received: 07.04.2023 Kabul / Accepted: 10.08.2023 Yayınlanma / Published: 15.10.2023

doi: 10.28948/ngumuh.1279335

hala nispeten sınırlıdır. Bunun başlıca sebebi deneysel çalışmaların uzun süreç gerektirmesi ve artan maliyetleridir. Bu yüzden hastalıkla ilişkili lncRNA'ları bulmak için benzerlik tabanlı ve makine öğrenmesi tabanlı birçok hesaplamalı yöntemler geliştirilmiştir.

lncRNA ve hastalıklarla ilgili etkileşim bilgilerini barındıran Lnc2Cancer [11], MNDR [12] ve NRED [13] gibi çok sayıda açık kaynaklı veri tabanı bulunmaktadır. Lnc2Cancer veri tabanı lncRNA ile ilgili çok sayıda literatür ve biyomedikal deneysel verilerden 1614 insan lncRNA'sı ve 165 insan hastalığı arasında 4989 ilişki barındırmaktadır. Bu veri tabanları lncRNA ve hastalık arasındaki ilişkiyi incelemeye yardımcı olur.

Bu çalışmada, potansiyel insan lncRNA-hastalık ilişkilerini daha etkili bir şekilde tahmin etmek için negatif olmayan matris tamamlama (NMF) hesaplama yöntemini uyguladık. Bu yöntemde bilinen lncRNA-hastalık ilişkilerini, lncRNA ekspresyon benzerlik, hastalık benzerlik ve kosinüs benzerlik ağırları ile entegre ederek başarılı bir tahmin performansı elde ettik. NMF yönteminin tahmin performansını göstermek için beş katlı çapraz doğrulama (5-katlı CV) ve birini dışarıda bırak çapraz doğrulama (LOOCV) tekniklerini uyguladık. Önerilen NMF yöntemi ile ROC grafikleri hem 5-katlı CV hem de LOOCV için çizdirilmiş ve AUC değerleri hesaplanmıştır. 5-katlı CV tekniğinde 0.7837 AUC değeri, LOOCV tekniğinde ise 0.8551 AUC değeri elde edilmiştir.

2 Materyal ve metot

2.1 İnsan lncRNA-Hastalık ilişkisi

Deneysel olarak doğrulanmış 369 lncRNA ve 246 hastalık arasındaki 687 deneysel olarak doğrulanmış lncRNA-hastalık ilişki verisi LncRNADisease [14] veri tabanından elde edilmiştir. Hastalık ontolojisi bilgisi olmayan hastalıkları [15] ve ArrayExpress'te ekspresyon profili olmayan lncRNA'ları filtreledik [16]. Sonuçta 115 lncRNA ve 178 hastalık içeren, deneysel olarak doğrulanmış 540 lncRNA-hastalık ilişki veri seti elde ettik. Komşuluk matrisi $A \in R^{nl \times nd}$ 'yi lncRNA ile hastalık arasındaki ilişkiyi temsil etmek için tanımladık; burada $A(i, j)$ lncRNA $l(i)$ ve hastalık $d(j)$ arasında bir ilişki olup olmadığını göstermektedir. lncRNA $l(i)$ ve hastalık $d(j)$ arasında bilinen bir ilişki var ise $A(i, j) = 1$, ilişki bilinmiyorsa $A(i, j) = 0$ yapılmıştır.

$$A(i, j) = \begin{cases} 0 & \text{lncRNA } l(i) \text{'nin hastalık } d(j) \text{ ile ilişkisi yoksa} \\ 1 & \text{lncRNA } l(i) \text{'nin hastalık } d(j) \text{ ile ilişkisi varsa} \end{cases} \quad (1)$$

2.2 lncRNA ekspresyon benzerliği

lncRNA ekspresyon profilleri Array-Express veri tabanından indirilmiştir [17]. Array-Express veri tabanı RNA-sekanslama teknolojisi ile üretilen 16 insan dokusundan 60.000'den fazla ifade profilini içermektedir. Her bir lncRNA çifti için Spearman korelasyon katsayısı lncRNA ekspresyon benzerliği olarak hesaplanır [18]. lncRNA $l(i)$ ve lncRNA $l(j)$ 'nin 0 ile 1 arasındaki

ekspresyon benzerlik skorunu temsil etmek için $LS(l_i, l_j)$ matrisi kullanılmıştır.

2.3 Hastalık Semantik benzerliği

Yönlendirilmiş döngüsüz grafik (DAG) hastalık semantik benzerliğini hesaplamak ve potansiyel lncRNA-hastalık ilişkilerini tahmin etmek için yaygın olarak kullanılmaktadır. İlk olarak ABD Ulusal Tıp Kütüphanesinden (<https://meshb.nlm.nih.gov/>) hastalıkların MeSH tanımları elde edilir ve her bir hastalık için DAG $G_{d_i} = (d_i, T_{d_i}, E_{d_i})$ yapısı oluşturulur. DAG yapısında düğümler hastalık MeSH tanımlayıcılarını temsil eder ve tüm MeSH tanımlayıcıları ana düğümlerden alt düğümlere olan ilişkiyi temsil etmek için doğrudan bir kenar kullanır. T_{d_i} kendisi de dahil olmak üzere d_i hastalığının ata düğümünü, E_{d_i} ise d_i ile olan bağlantıları gösterir. Bu çalışmada hastalık semantik benzerliği [19, 20]'de anlatıldığı gibi hesaplanmıştır. Hastalık d_i ile ata hastalığı p arasındaki hastalık semantik benzerliği Denklem (2) ve Denklem (3) ile hesaplanır:

$$d_i(p) = \begin{cases} 1 & \text{eğer } p = d_i \\ \max(\Delta \times D_{d_i}(p') | p' \in p \text{'nin alt hastalığı}) & \text{eğer } p \neq d_i \end{cases} \quad (2)$$

$$DV(d_i) = \sum_{p \in T_{d_i}} D_{d_i}(p) \quad (3)$$

Burada $D_{d_i}(p)$ ve $D_{d_j}(p)$ sırasıyla hastalık d_i ve hastalık d_j 'nin semantik değerini göstermektedir. Hastalık d_i ve hastalık d_j arasındaki semantik benzerlik Denklem (4) ile hesaplanır:

$$DS(d_i, d_j) = \frac{\sum_{p \in T(d_i) \cap T(d_j)} (D_{d_i}(p) + D_{d_j}(p))}{DV(d_i) + DV(d_j)} \quad (4)$$

2.4 lncRNA'lar ve hastalıkları için kosinüs benzerliği

lncRNA kosinüs benzerliği ilkesi eğer lncRNA l_i ve lncRNA l_j birbirine benzerse o zaman lncRNA-hastalık ilişki matrisinde ikili vektör $IP(l_i)$ ve ikili vektör $IP(l_j)$ 'nin birbirine benzer olması gerektiği varsayımına dayanır. Her biri aynı varsayım hastalıklar için de geçerli olmalıdır. Bilinen lncRNA-hastalık ilişkilendirme verilerine göre, lncRNA'ların kosinüs benzerlik matrisi Denklem (5) ve Denklem (6) ile hesaplanır:

$$CL = (CL(l_i, l_j))_{N_l \times N_l} \quad (5)$$

$$CL(l_i, l_j) = \frac{IP(l_i) \cdot IP(l_j)}{\|IP(l_i)\| \|IP(l_j)\|} \quad (6)$$

Burada $CL(l_i, l_j)$ lncRNA l_i ve lncRNA l_j arasındaki kosinüs benzerliğidir. İkili vektör $IP(l_i)$, lncRNA $l(i)$ ile her

bir hastalık (komşuluk matris A 'nın i 'inci sütunu) arasındaki ilişkinin varlığını veya yokluğunu gösterir.

Benzer şekilde, hastalık d_i ve hastalık d_j arasındaki kosinüs benzerliği **Denklem (7)** ve **Denklem (8)** ile hesaplanabilir:

$$CD = (CD(d_i, d_j))_{N_d \times N_d} \quad (7)$$

$$CD(d_i, d_j) = \frac{IP(d_i) \cdot IP(d_j)}{\|IP(d_i)\| \|IP(d_j)\|} \quad (8)$$

Burada $CD(d_i, d_j)$ hastalık d_i ve hastalık d_j arasındaki kosinüs benzerliğidir. İkili vektör $IP(d_i)$, hastalık $d(i)$ ile her bir lncRNA (komşuluk matris A 'nın i 'inci satırı) arasındaki ilişkinin varlığını veya yokluğunu gösterir.

2.5 LncRNA ve hastalık benzerliklerinin entegre edilmesi

Kapsamlı ve güvenilir bir lncRNA benzerlik matrisi elde etmek için lncRNA'nın kosinüs benzerlik matrisini ve lncRNA ekspresyon benzerlik matrisini entegre ettik. LncRNA l_i ve lncRNA l_j 'nin komşuluk matrisi A 'da ortak ilişkili hastalıkları yoksa, kosinüs benzerlik matrisi aralarındaki değer 0'dır. Bu nedenle, lncRNA l_i ve lncRNA l_j 'nin kosinüs benzerlik puanı olmadığında, entegre benzerlik puanı olarak doğrudan aralarındaki lncRNA ekspresyon benzerlik puanını kullanırız. LncRNA l_i ve lncRNA l_j 'nin kosinüs benzerlik matrisinde bir benzerlik puanı varsa, kosinüs benzerlik puanları ve ekspresyon benzerlik puanları nihai benzerlik puanı olarak entegre edilir. LncRNA entegre benzerlik puanları **Denklem (9)** ile hesaplanır.

$$= \begin{cases} \frac{IL(l_i, l_j) + CL(l_i, l_j)}{2} & \text{eğer } l_i \text{ ve } l_j \text{ ilişkili ise} \\ LS(l_i, l_j) & \text{değilse} \end{cases} \quad (9)$$

Benzer şekilde, hastalık semantik benzerlik matrisi ve hastalık kosinüs benzerlik matrisi entegre edilerek entegre hastalık benzerlik matrisi elde edilir. Hastalık d_i ve hastalık d_j 'nin komşuluk matrisi A 'da ortak ilişkili lncRNA'ları yoksa, kosinüs benzerlik matrisinde aralarındaki değer 0 olduğu için doğrudan hastalık semantik benzerliğini kullanırız. Hastalık kosinüs benzerlik matrisinde benzerlik puanı varsa, hastalık kosinüs benzerlik matrisi ve hastalık semantik benzerlik matrisindeki puanları entegre ederiz. Hastalık entegre benzerlik matrisi **Denklem (10)** ile hesaplanır:

$$= \begin{cases} \frac{ID(d_i, d_j) + CD(d_i, d_j)}{2} & \text{eğer } d_i \text{ ve } d_j \text{ ilişkili ise} \\ DS(d_i, d_j) & \text{değilse} \end{cases} \quad (10)$$

2.6 Negatif olmayan matris faktörizasyonu

Negatif olmayan matris faktörizasyonu (nonnegative matrix factorization-NMF) tahmin sistemlerinde uygulanabilen önemli bir algoritmadır [21]. Son yıllarda NMF ilaç-ilaç [22], lncRNA-protein [23], miRNA-hastalık [24], mikrop-hastalık [25], ilaç-hastalık [26] ve circRNA-hastalık [27] potansiyel etkileşimlerini tahmin etmek için başarıyla kullanılmıştır. NMF komşuluk matrisini doğrudan iki düşük boyutlu negatif olmayan matrise ayrıştırarak lncRNA-hastalık ilişkisi skor matrisini elde etmeyi amaçlamaktadır [28]. Bu çalışmada, lncRNA-hastalık ilişki matrisi $Y \in R^{468 \times 262}$, $U \in R^{k \times 468}$ ve $V \in R^{k \times 262}$ olmak üzere iki alt matrise ayrıştırılır, burada k ($k < rn/(r+n)$) alt uzay boyutunu temsil etmektedir. LncRNA-hastalık ilişki tahmini matematiksel ifadesi **Denklem (11)** ve **Denklem (12)** ile formüle edilmiştir:

$$Y \cong U^T V \quad (U \geq 0, V \geq 0) \quad (11)$$

Amaç fonksiyonu şu şekilde yazılabilir:

$$\min_{U, V} \|Y - U^T V\|_F^2 \quad (U \geq 0, V \geq 0) \quad (12)$$

Burada $\|\cdot\|_F$ Frobenius normu temsil etmektedir. $U, V \geq 0$ ise U ve V 'nin tüm öğelerinin negatif olmadığını gösterir. $\|U\|_F^2 = Tr(U^T U)$, $Tr(U^T) = Tr(U)$ ve $Tr(UV) = Tr(VU)$ matris özelliklerine göre **Denklem (13)**'ü elde edebiliriz.

$$\begin{aligned} \|Y - U^T V\|_F^2 &= Tr((Y - U^T V)(Y - U^T V)^T) \\ &= Tr(Y Y^T) - 2Tr(Y V^T U) \\ &\quad + Tr(U^T V V^T U) \end{aligned} \quad (13)$$

Lee ve arkadaşları [28] U ve V 'nin çarpımsal güncelleme kurallarına dayanan, negatif olmayan matris çarpanlara ayırma algoritmasını önermişlerdir. **Denklem (13)**'ün güncelleme kuralları **Denklem (14)** ve **Denklem (15)**'deki gibidir:

$$u_{ki} \leftarrow u_{ki} \frac{(V Y^T)_{ki}}{(V V^T U)_{ki}} \quad (14)$$

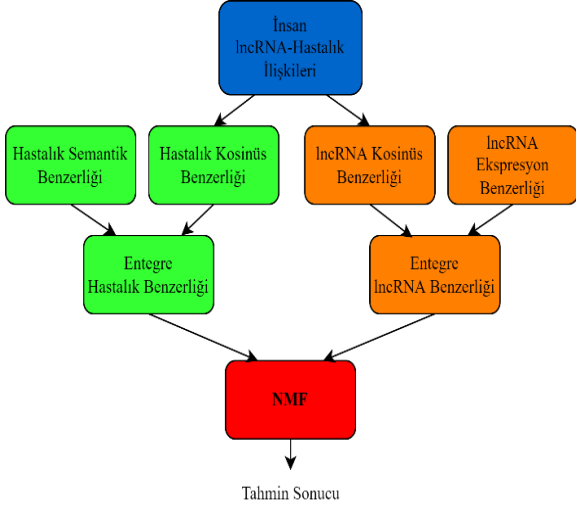
$$v_{kj} \leftarrow v_{kj} \frac{(U Y^T)_{kj}}{(U U^T V)_{kj}} \quad (15)$$

Denklem (14) ve **Denklem (15)** ile U ve V matrisleri yakınsayana kadar güncellenir. LncRNA-hastalık çiftleri için tahmin edilen ilişkilendirme skoru matrisi $Y = U^T V$ ile elde edilir. Sonuç olarak lncRNA-hastalık çifti puanı ne kadar yüksekse ilişkili olma olasılıkları o kadar yüksektir. Uygulamış olduğumuz NMF yönteminin akış diyagramı **Şekil 1**'de görülmektedir.

3 Performans hesaplaması

Bu çalışmada, NMF'nin lncRNA-hastalık etkileşimlerini tahmin etme performansını değerlendirmek için beş katlı çapraz doğrulama ve LOOCV teknikleri uyguladık. Beş katlı

çapraz doğrulamada bilinen 540 lncRNA-hastalık ilişki veri setini rastgele beş eşit alt kümeye ayırdık. Dördü eğitim seti olarak kullanılır ve geri kalan bir alt küme test örneği olarak kullanılmıştır. LOOCV tekniğinde ise bilinen 540 lncRNA-hastalık ilişki veri setinin her biri test örneği olarak kullanılırken diğerleri eğitim seti olarak kullanılmıştır.



Şekil 1. NMF yönteminin akış diyagramı

ROC (alıcı işlem karakteristiği) eğrisi ve AUC (alıcı işlem karakteristiği eğrisinin altındaki alan) değeri makine öğrenmesinde performans tahmini için yaygın olarak kullanılır [29]. ROC eğrisi gerçek pozitif oranı TPR (*Sensitivity*) karşılık yanlış pozitif oranı FPR ($1 - Specificity$) çizdirilerek elde edilir. ROC eğrisini çizdirebilmek için oluşturulan karışıklık matrisi ile duyarlılık (*Sensitivity*) ve özgüllük (*Specificity*) Denklem (16) ve Denklem (17) ile hesaplanır:

$$TPR(Sensitivity) = \frac{TP}{TP + FN} \quad (16)$$

$$FPR(1 - Specificity) = \frac{TN}{TN + FP} \quad (17)$$

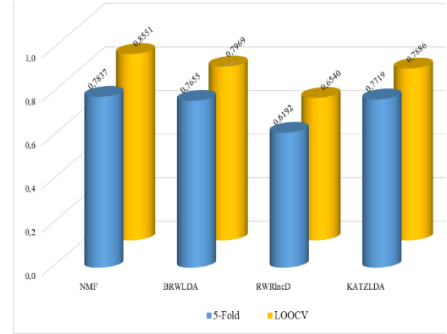
Burada, TP gerçek pozitif, FN yanlış negatif, TN gerçek negatif ve FP yanlış pozitif tahmin sayısını göstermektedir.

AUC değeri ise ROC eğrisinin altında kalan hesaplanarak elde edilir. AUC değeri 0 ile 1 arasında değişmektedir. Büyük bir AUC değeri modelin daha iyi tahmin performansını temsil etmektedir.

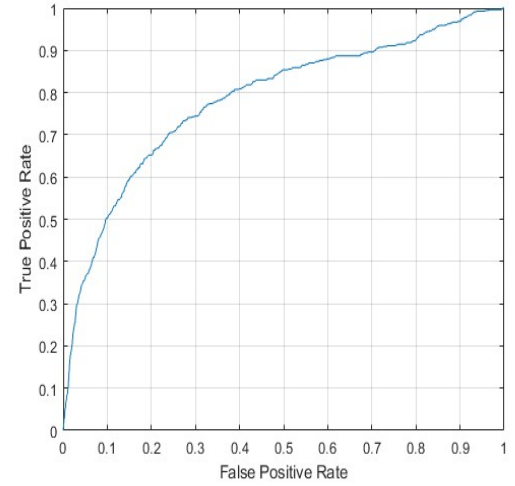
NMF yönteminin ROC eğrileri Şekil 3 ve Şekil 4'te görülmektedir. Yöntemin 5-katlı CV tekniğinde hesaplanan AUC değeri 0.7837, LOOCV tekniğinde hesaplanan AUC değeri ise 0.8551'dir. Hesapladığımız AUC değerlerini bizden önce yapılmış ve aynı veri tabanını kullanan çalışmalar ile Tablo 1'de ve Şekil 2'de görüldüğü gibi karşılaştırdık. Sonuçlar incelendiğinde NMF yönteminin muhtemel lncRNA'lar ve hastalıklar arasındaki ilişkileri belirlemede kullanılabileceği görülmüştür.

Tablo 1. NMF yöntemi ve diğer yöntemlerin karşılaştırmalı AUC değerleri

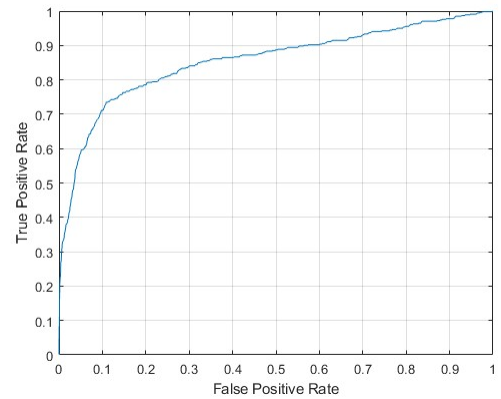
	5-Fold	LOOCV
NMF	0.7837	0.8551
BRWLDA [30]	0.7655	0.7969
RWRlncD [31]	0.6192	0.6540
KATZLDA [32]	0.7719	0.7886



Şekil 2. AUC değerlerinin karşılaştırmalı grafiği



Şekil 3. 5-fold CV ROC



Şekil 4. LOOCV ROC

4 Sonuç ve tartışma

Çok sayıda biyolojik deney ve klinik çalışma lncRNA'nın birçok karmaşık insan hastalığının gelişiminde hayati bir rol oynadığını göstermiştir. Bu nedenle potansiyel lncRNA-hastalık ilişkilerini tahmin etmek önemlidir. Bu tahmin yalnızca hastalığın arkasındaki temelleri anlamaya yardımcı olmakla kalmaz, aynı zamanda hastalığın prognozuna ve önlenmesine de katkıda bulunur. Bununla birlikte biyolojik deneyler ve klinik çalışmalar yoluyla ilişkilendirmelerin doğrulanması zaman alıcı ve pahalıdır. Bu çalışmada potansiyel lncRNA-hastalık ilişkilerini tahmin etmek için NMF hesaplama yöntemi öneriyoruz. Önermiş olduğumuz yöntem ile 5-katlı CV tekniğinde 0.7837 AUC değeri, LOOCV tekniğinde ise 0.8551 AUC değeri hesaplandı.

Diğer üç gelişmiş yöntemle karşılaştırıldığında NMF 5-katlı CV tekniğinde ve LOOCV tekniğinde daha iyi tahmin performansına sahiptir. NMF yönteminin başarılı tahmin performansı elde etmesinin sebeplerinden ilki NMF lncRNA kosinüs benzerliği ile lncRNA ekspresyon benzerliğini ve hastalık kosinüs benzerliği ile hastalık semantik benzerliğini bütünleştirir. İkincisi, NMF negatif örnekleri olmayan bir modeldir. Üçüncüsü, NMF bilinen herhangi bir ilişkili hastalık olmadan lncRNA'ları tahmin edebilir.

NMF yönteminin başarılı tahmin performansı olmasına rağmen, bazı sınırlamaları da vardır. Bilinen lncRNA-hastalık ilişki yoğunluğu nispeten azdır ve bu da tahmin sonuçlarının doğruluğu üzerinde belirli bir etkiye sahip olacaktır. Ayrıca, lncRNA kosinüs benzerlik ağı ve hastalık kosinüs benzerlik ağı büyük ölçüde bilinen lncRNA-hastalık ilişki ağına dayanır. Bu durum daha fazla bilinen ilişkili lncRNA'lara sahip hastalıklar ve daha fazla bilinen ilişkili hastalıklara sahip lncRNA'lar için tahmin yanlılığına yol açar. Tahmin yanlılığını azaltmak ve tahmin performansını iyileştirmek için gelecekte NMF yöntemine daha büyük bir lncRNA-hastalık ilişkilendirme ağı, hastalık fenotip benzerlik ağı ve lncRNA fonksiyonel benzerlik ağı gibi ek deneysel olarak doğrulanmış biyolojik veriler uygulanmalıdır. Ayrıntılı bir füzyon yöntemiyle daha fazla biyolojik bilgi uygulandığında tahmin doğruluğu geliştirilebilir.

Çıkar çatışması

Yazarlar çıkar çatışması olmadığını beyan etmektedir.

Benzerlik oranı (iThenticate): %6

Kaynaklar

- [1] Washietl, S., M. Kellis, and M. Garber, Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome research*, 2014. 24(4): p. 616-628. DOI: <https://doi.org/10.1101/gr.165035.113>.
- [2] Schmitt, A.M. and H.Y. Chang, Long noncoding RNAs in cancer pathways. *Cancer cell*, 2016. 29(4): p. 452-463. DOI: <https://doi.org/10.1016/j.ccell.2016.03.010>.
- [3] Huang, H., et al., Upregulation of lncRNA PANDAR predicts poor prognosis and promotes cell proliferation in cervical cancer. *Eur Rev Med Pharmacol Sci*, 2017. 21(20): p. 4529-4535.
- [4] Zhu, Y., et al., Oncogenic activity of Wrap53 in human colorectal cancer in vitro and in nude mouse xenografts. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, 2018. 24: p. 6129. DOI: <https://doi.org/10.12659/MSM.910214>.
- [5] Tripathi, M.K., et al., Role of lncRNAs in ovarian cancer: defining new biomarkers for therapeutic purposes. *Drug Discovery Today*, 2018. 23(9): p. 1635-1643. DOI: <https://doi.org/10.1016/j.drudis.2018.04.010>.
- [6] Sun, S.-C., et al., Expression of long non-coding RNA H19 in prostate cancer and its effect on the proliferation and glycometabolism of human prostate cancer cells. *Zhonghua nan ke xue= National Journal of Andrology*, 2017. 23(2): p. 120-124.
- [7] Thomas, A.A., et al., lncRNA H19 prevents endothelial-mesenchymal transition in diabetic retinopathy. *Diabetologia*, 2019. 62: p. 517-530. DOI: <https://doi.org/10.1007/s00125-018-4797-6>.
- [8] Collette, J., X. Le Bourhis, and E. Adriaenssens, Regulation of human breast cancer by the long non-coding RNA H19. *International journal of molecular sciences*, 2017. 18(11): p. 2319. DOI: <https://doi.org/10.3390/ijms18112319>.
- [9] Iacoangeli, A., et al., BC200 RNA in invasive and preinvasive breast cancer. *Carcinogenesis*, 2004. 25(11): p. 2125-2133. DOI: <https://doi.org/10.1093/carcin/bgh228>.
- [10] Pasmant, E., et al., Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer research*, 2007. 67(8): p. 3963-3969. DOI: <https://doi.org/10.1158/0008-5472.CAN-06-2004>.
- [11] Gao, Y., et al., Lnc2Cancer 3.0: an updated resource for experimentally supported lncRNA/circRNA cancer associations and web tools based on RNA-seq and scRNA-seq data. *Nucleic acids research*, 2021. 49(D1): p. D1251-D1258. DOI: <https://doi.org/10.1093/nar/gkaa1006>.
- [12] Cui, T., et al., MNDR v2. 0: an updated resource of ncRNA-disease associations in mammals. *Nucleic acids research*, 2018. 46(D1): p. D371-D374. DOI: <https://doi.org/10.1093/nar/gkx1025>.
- [13] Dinger, M.E., et al., NRED: a database of long noncoding RNA expression. *Nucleic acids research*, 2009. 37(suppl_1): p. D122-D126. DOI: <https://doi.org/10.1093/nar/gkn617>.
- [14] Chen, G., et al., LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic acids research*, 2012. 41(D1): p. D983-D986. DOI: <https://doi.org/10.1093/nar/gks1099>.
- [15] Zhang, X., et al., Long non-coding RNA expression profiles predict clinical phenotypes in glioma.

- Neurobiology of disease, 2012. 48(1): p. 1-8. DOI: <https://doi.org/10.1016/j.nbd.2012.06.004>.
- [16] Schriml, L.M., et al., Disease Ontology: a backbone for disease semantic integration. *Nucleic acids research*, 2012. 40(D1): p. D940-D946. DOI: <https://doi.org/10.1093/nar/gkr972>.
- [17] Parkinson, H., et al., ArrayExpress—a public database of microarray experiments and gene expression profiles. *Nucleic Acids Research*, 2006. 35(suppl_1): p. D747-D750. DOI: <https://doi.org/10.1093/nar/gkl995>.
- [18] Chen, X. and G.-Y. Yan, Novel human lncRNA–disease association inference based on lncRNA expression profiles. *Bioinformatics*, 2013. 29(20): p. 2617-2624. DOI: <https://doi.org/10.1093/bioinformatics/btt426>.
- [19] Toprak, A. and E. Eryilmaz, Prediction of miRNA-disease associations based on Weighted K-Nearest known neighbors and network consistency projection. *Journal of Bioinformatics and Computational Biology*, 2021. 19(1): p. 2050041. DOI: <https://doi.org/10.1142/S0219720020500419>.
- [20] Toprak, A. and E. Eryilmaz Dogan, Prediction of Potential MicroRNA-Disease Association Using Kernelized Bayesian Matrix Factorization. *Interdisciplinary Sciences: Computational Life Sciences*, 2021. 13(4): p. 595-602. DOI: <https://doi.org/10.1007/s12539-021-00469-w>.
- [21] Luo, X., et al., An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. *IEEE Transactions on Industrial Informatics*, 2014. 10(2): p. 1273-1284. DOI: <https://doi.org/10.1109/TII.2014.2308433>.
- [22] Zhang, W., et al., Manifold regularized matrix factorization for drug-drug interaction prediction. *Journal of biomedical informatics*, 2018. 88: p. 90-97. DOI: <https://doi.org/10.1016/j.jbi.2018.11.005>.
- [23] Zhang, T., et al., LPGNMF: predicting long non-coding RNA and protein interaction using graph regularized nonnegative matrix factorization. *IEEE/ACM transactions on computational biology and bioinformatics*, 2018. 17(1): p. 189-197. DOI: <https://doi.org/10.1109/TCBB.2018.2861009>.
- [24] Xiao, Q., et al., A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics*, 2018. 34(2): p. 239-248. DOI: <https://doi.org/10.1093/bioinformatics/btx545>.
- [25] Liu, Y., S.-L. Wang, and J.-F. Zhang, Prediction of microbe–disease associations by graph regularized non-negative matrix factorization. *Journal of Computational Biology*, 2018. 25(12): p. 1385-1394. DOI: <https://doi.org/10.1089/cmb.2018.0072>.
- [26] Zhang, W., et al., Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC bioinformatics*, 2018. 19: p. 1-12. DOI: <https://doi.org/10.1186/s12859-018-2220-4>.
- [27] Wei, H. and B. Liu, iCircDA-MF: identification of circRNA-disease associations based on matrix factorization. *Briefings in bioinformatics*, 2020. 21(4): p. 1356-1367. DOI: <https://doi.org/10.1093/bib/bbz057>.
- [28] Lee, D.D. and H.S. Seung, Learning the parts of objects by non-negative matrix factorization. *Nature*, 1999. 401(6755): p. 788-791. DOI: <https://doi.org/10.1038/44565>.
- [29] Fawcett, T., An introduction to ROC analysis. *Pattern Recognition Letters*, 2006. 27(8): p. 861-874. DOI: <https://doi.org/10.1016/j.patrec.2005.10.010>.
- [30] Yu, G., et al., BRWLDA: bi-random walks for predicting lncRNA-disease associations. *Oncotarget*, 2017. 8(36): p. 60429. DOI: <https://doi.org/10.18632/oncotarget.19588>.
- [31] Sun, J., et al., Inferring novel lncRNA–disease associations based on a random walk model of a lncRNA functional similarity network. *Molecular BioSystems*, 2014. 10(8): p. 2074-2081. DOI: <https://doi.org/10.1039/c3mb70608g>.
- [32] Chen, X., KATZLDA: KATZ measure for the lncRNA-disease association prediction. *Scientific Reports*, 2015. 5: p. 16840. DOI: <https://doi.org/10.1038/srep16840>.

