

Göğüs Röntgeni Görüntülerinin Sınıflandırması için Fokal Modülasyona Dayalı Uçtan Uca Çok Etiketli Sınıflandırma

Focal Modulation Based End-to-End Multi-Label Classification for Chest X-ray Image Classification

Şaban Öztürk^{1,2,3} ve Tolga Çukur^{1,3}

¹Elektrik ve Elektronik Mühendisliği Bölümü, Bilkent Üniversitesi, Ankara, Türkiye

²Elektrik ve Elektronik Mühendisliği Bölümü, Amasya Üniversitesi, Amasya, Türkiye

³Ulusal Manyetik Rezonans Araştırma Merkezi (UMRAM), Bilkent Üniversitesi, Ankara, Türkiye
saban.ozturk@amasya.edu.tr, cukur@ee.bilkent.edu.tr

Özetçe —Çeşitli çevresel ve kalıtsal faktörlerin etkisiyle günümüzde artan göğüs hastalıklarının erken dönemde etkili biçimde teşhis edilebilmesi amacıyla göğüs röntgeni görüntüleri kritik öneme sahiptir. Göğüs röntgeni patolojik anormalliklerin tespitinde en sık kullanılan cihaz olmasına rağmen patolojik anormalliklerin değişken lokasyonları ve boyutları, görsel benzerlikler, ve karmaşık arkaplan gibi yanıltıcılar sebebiyle uzmanlar için oldukça zorlayıcı olabilmektedir. Geleneksel derin öğrenme (DÖ) mimarileri patolojik anormalliklerin nispeten küçük alanlarda olması ve hastalıklı-sağlıklı alanların benzerliklerinden ötürü yetersiz kalmaktadır. Ek olarak, standart sınıflandırma yaklaşımlarına sahip DÖ yapıları çoklu hastalık içeren problemler ile baş etmek için ideal değildirler. Bahsedilen problemlerin üstesinden gelebilmek için bu çalışmada ilk olarak bilindik bir konvolüsyonel sinir ağı (KSA) kullanılarak arkaplandan bağımsız özellik haritaları oluşturulmuştur. Daha sonra, özellik haritalarındaki nesnelere arasındaki ilişkiler öz-dikkat yaklaşımına göre daha efektif bir yenilikçi dikkat modülü olan fokal modülasyon ağı (FMA) kullanılarak çok etiketli sınıflandırma görevleri için uygun hale getirilmektedir. Toplamda 14 farklı hastalığa dair hem tekli ve hem de çoklu etiketler içeren bir göğüs röntgeni veri kümesi kullanılarak gerçekleştirilen deneyler önerilen yaklaşımın çok etiketli veri kümeleri için üstün performans sağlayabileceğini göstermektedir.

Anahtar Kelimeler—göğüs röntgeni, derin öğrenme, fokal modülasyon, çok etiketli sınıflandırma.

Abstract—Chest X-ray imaging is of critical importance in order to effectively diagnose chest diseases, which are increasing today due to various environmental and hereditary factors. Although chest X-ray is the most commonly used device for detecting pathological abnormalities, it can be quite challenging for specialists due to misleading locations and sizes of pathological abnormalities, visual similarities, and complex backgrounds. Traditional deep learning (DL) architectures fall short due to relatively small areas of pathological abnormalities and similarities between diseased and healthy areas. In addition, DL structures with standard classification approaches are not ideal for dealing with problems involving multiple diseases. In order to overcome the aforementioned problems, firstly, background-independent feature maps were created using a conventional convolutional neural network (CNN). Then, the relationships

between objects in the feature maps are made suitable for multi-label classification tasks using the focal modulation network (FMA), an innovative attention module that is more effective than the self-attention approach. Experiments using a Chest x-ray dataset containing both single and multiple labels for a total of 14 different diseases show that the proposed approach can provide superior performance for multi-label datasets.

Keywords—chest x-ray, deep learning, focal modulation networks, multi-label classification.

I. GİRİŞ

Akciğer kanseri, bronşit ve zatürre gibi göğüs hastalıkları günümüz toplumlarının sağlığını ciddi ölçüde tehdit eden hastalıkların başında gelmektedir. 10 yıl ara ile yapılan araştırmalar incelendiğinde 2013 yılında sadece Amerika Birleşik Devletlerinde 159,480 tanesi ölüm ile sonuçlanan 228,190 akciğer ve bronşit kaynaklı kanser vakasını [1], 2023 yılında sadece Amerika Birleşik Devletlerinde 127,070 tanesi ölüm ile sonuçlanan 238,340 akciğer ve bronşit kaynaklı kanser vakasını ortaya koymaktadır [2]. Geçmişten günümüze hasta sayısında artış olmasına rağmen şanslıyız ki erken teşhisi mümkün kılabilen teknolojik medikal cihazların yaygınlaşması sayesinde ölüm oranları azalmaktadır. Medikal cihazlardan, gelişmiş ülkelere ait sağlık kuruluşlarında oldukça etkin biçimde yararlanılmasına rağmen az gelişmiş ülkelerde ekonomik sebepler ve yeterince uzman doktorun olmaması gibi sebeplerden ötürü yeterince yararlanılamamaktadır. Fakat, göğüs röntgen görüntülerinin içerdiği karmaşık patoloji, değişken lezyon boyutları ve lokasyon çeşitliliği, dokuda çeşitlilikler gibi problemlerden ötürü gelişmiş ülkelerdeki eğitimli radyologlar bile teşhis sürecinde ciddi zorluklar yaşamaktadır [3]. Bu zorlukların aşılması için yapay zeka (YZ) yaklaşımları kullanılarak bilgisayar destekli teşhis (BDT) sistemlerinden sıklıkla faydalanılmaktadır [4]–[6]. BDT sürecinde geçmişte özellik mühendisliğine dayalı YZ yaklaşımları kullanılırken günümüzde gelişmiş uçtan-uca derin öğrenme (DÖ) yapıları tercih edilmektedir [7]–[9].

Konvolüsyonel sinir ağları (KSA) tabanlı yaklaşımlar görüntüye ait görsel özellikleri uçtan uca öğrenen günümüzdeki en etkin özellik çıkarma yaklaşımlarıdır [10]. Genel amaçlı ve çeşitli medikal görüntü analizi çalışmaları için standart yöntem haline gelmelerine rağmen göğüs röntgeni görüntülerinin analizinde üç temel problemi aşmada sorun yaşarlar [11]–[14]. İlki, patolojik anormalliklerin aynı sınıflar için bile tamamen farklı görünebilmesi ve farklı sınıflar için görsel farklılıkların az olması sebebiyle düşük duyarlılık performans üretmeleridir. İkincisi, patolojik anomalilerin boyutlarının normal görüntüde oldukça küçük kalması sebebiyle hastalıklı bölgeleri yüksek doğrulukla yakalayamamalarıdır. Üçüncüsü, bir görüntüde bir veya birden fazla hastalık olabilmesi ve bu bölgelerin keşif edilebilmesi ihtimaline karşı standart bir çıkışa sahip olmamalarıdır [15]. Bu zorluklara ek olarak, medikal veri kümelerinin çoğunda karşılaşılan sınıflar arası veri sayısı dengesizliği [16] ile karşılaştıklarında gürbüz bir performans ortaya koyamamaları bilindik KSA mimarilerinin çoklu sınıf etiketine sahip göğüs röntgeni görüntülerinin sınıflandırılması görevlerinde etkisiz kalmalarına sebep olmaktadır.

Görsel dikkat modülleri hastalıkla ilişkili olmayan bölgeleri elimine ederek küçük boyutlu patolojik anomalilerin tespit edilmesine ve birden fazla anomaliye odaklanılmasına katkı sağlayabildiğinden göğüs röntgeni görüntülerinin sınıflandırılması görevlerinde sıkça tercih edilmektedir [17]. Görsel özelliklerin uzun menzilli ilişkilerini gürbüz biçimde yakalayabilmek için literatürdeki bazı göğüs röntgeni sınıflandırma çalışmaları görüntüye direk görüntü dönüştürücü yapılarını (ViT) uygulamaktadır [13]. Bu kullanım biçimi büyük miktarda bellek gereksinimi doğurduğundan literatürde bir omurga KSA tarafından çıkarılmış özellik haritalarına görsel dikkat uygulayan çalışmalar daha yaygındır [12]. Hastalıklar ile ilişkisi olmayan bölgelerin etkisini daha da azaltmak için bazı çalışmalar bölütleme maskeleri ile akciğer bölgesini veya riskli alanları bölütleyerek sadece ilgili alanların sınıflandırılması üzerine çalışmaktadır [18]. Literatürdeki mevcut yaklaşımlar ve içerdikleri teknikler temsil gücü oldukça yüksek görüntü özellikleri üretebilmektedir. Fakat çoklu etiketler arasındaki ilişkileri etkili biçimde yakalayabilecek bir çıkış mimarisinden muzdariptirler ve veri sayısı dengesizlikleri ile adaptif biçimde başa çıkabilecek yaklaşımlardan yoksundurlar. Çoklu etiket ilişkilerinin yakalanması için ön tanımlı ilişki ağları kullanmak [19] ilgili problemi rahatlatır da yüksek performanslı çözüm sunmamaktadır.

ViT yapılarındaki öz-dikkat modülleri [20] özellikle yüksek çözünürlüklü görüntüler için görsel token sayısı üzerinde ikinci dereceden karmaşıklık problemine sahip olmaları sebebiyle göğüs röntgeni görüntülerinin sınıflandırılması için efektif değildir. Görsel özelliklerin uzun-mesafeli bağlantılarını daha efektif biçimde ortaya çıkarabilmek için biz fokal modülasyon ağları (FMA) [21] tabanlı bir çözüm ortaya koymaktayız. FMA, görsel içeriği kısıdan uzuna aralıklar ile kodlamak için derin konvolüsyon katmaları yığını kullanarak odaklar arası bağlar oluşturur, her bir sorgu tokeni için odak noktalarını seçer ve bir modülatörde toplar ve son olarak modülatörü sorguya enjekte etmek için eleman bazında afin dönüşüm uygular. Bu yolla öz-dikkat modülünün problemlerini çözerek efektif biçimde daha yüksek performans üretir. Bu çalışma hem FMA kullanarak çoklu etikete sahip veri kümelerinin sınıflandırılması problemini araştıran hem de FMA ile göğüs röntgeni görüntülerinin sınıflandırılmasını araştıran literatür-

deki ilk çalışmadır.

II. YÖNTEM

A. Önerilen Yaklaşım

N adet eğitim görüntüsünden oluşan bir veri kümesi tanımlayalım, $\{x_i, y_i\}_{i=1}^N$. Burada x_i , i 'nci giriş görüntüsü ifade ederken, y_i ise i 'nci hedef çıkış vektörünü temsil etmektedir. L adet farklı hastalık içeren göğüs röntgen veri kümesinde y_i , '0' ve '1' değerlerinden oluşan L -boyutlu bir vektördür, $\{y_i^k\}_{k=1}^L \in \{0, 1\}$ ('0' sağlıklı, '1' hastalık içerdiğini belirtir). Unutulmamalıdır ki çoklu etikete sahip veri kümelerinde her bir vektör birden fazla '1' değeri içerebilmektedir. Amacımız bir x_i giriş görüntüsünü modelleyip yüksek doğruluğa sahip bir çoklu etiketli y_i vektörü üretmektir, $f : x_i \rightarrow y_i$. Bu amaçla Şekil 1'de görüldüğü gibi önce görüntü özellik haritalarını çıkararak, daha sonra ise bu haritalarda kısa ve uzun menzilli özellik ilişkilerini inceleyen bir FMA ağı kullandık.

Hastalıkla ilişkisi olmayan alanların elimine edilmesi ve etkin hafıza kullanımının sağlanması için ImageNet veri kümesi ile ön eğitilmiş bir VGG16 yapısı (ξ_{VGG16}) ile göğüs röntgeni görüntülerinin C derinliğinde özellik haritaları çıkartmaktayız, $\xi_{VGG16} : x_i \rightarrow m_i^C$. VGG16 mimarisini için C değeri 512 ve özellik haritalarının boyutları ise 7×7 'dir. Bu işlem için VGG16 mimarisinin tam bağlı katmanlarını çıkarttık. Daha sonra bir FMA (ψ_{FMA}) kullanarak özellik haritalarından, temsil gücü yüksek çıkış vektörleri \tilde{y}_i ürettik, $\psi_{FMA} : m_i^C \rightarrow \tilde{y}_i$. FMA operasyonu Eşitlik 1'deki gibi ifade edilebilir.

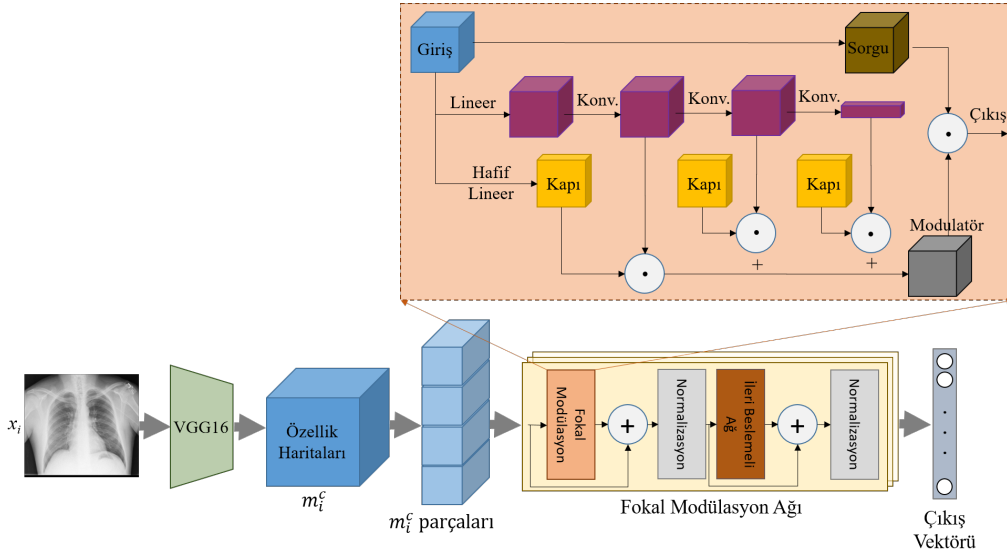
$$\tilde{y}_i = q(x_i) \odot h \left(\sum_{c=1}^{C+1} g_i^c \cdot z_i^c \right) \quad (1)$$

Eşitlik 1'de $q(\cdot)$ sorgu projeksiyon fonksiyonunu, \odot eleman bazında çarpma işlemini, $h(\cdot)$ lineer katmanını ifade eder. g_i^c ve z_i^c ise sırasıyla G^c ve Z^c için i lokasyonunda kapılma değeri ve görsel özellikleri ifade eder. G^c ve Z^c aşağıdaki gibi hesaplanır:

$$G^c = f_g(G^{c-1}) = \text{GeLU}(Z^{c-1}) \quad (2)$$

$$Z^c = f_a(Z^{c-1}) = \text{GeLU}(\text{DWConv}(Z^{c-1})) \quad (3)$$

Eşitlik 2'de f^g lineer bir aktivasyon fonksiyonunu temsil eder, bu amaçla biz GeLU aktivasyon fonksiyonunu kullandık. Eşitlik 3'te f_a^c derinlik güdümlü konvolüsyon işlemi (DWConv) uygulayarak c 'nci seviyede bağ oluşturma fonksiyonudur. FMA tarafından üretilen özellik vektörü daha sonra bir sigmoid aktivasyon fonksiyonundan (ς) geçirilerek $[0,1]$ aralığındaki değerlere çekilir, $\varsigma(\tilde{y}_i) = 1 / (1 + e^{-\tilde{y}_i})$. Ek olarak, medikal veri kümelerinde yaygın olan veri sayısı dengesizliği probleminin [22] çözümü için etiket var olma istatistiğine dayalı bir ağırlıklandırma kullanılmaktadır. Bu amaçla örnek sayısı fazla olan sınıflara toplam örnek sayısına oranlı biçimde azınlık sınıflara oranla görece düşük ağırlıklar atamaktayız. Çoklu etiketli çıkış vektörlerinin kayıplarını hesaplayabilmek ve ağ parametrelerini güncelleyebilmek için Eşitlik 4'teki gibi bir çarpaz-entropi fonksiyonu (E) kullanılmaktadır.



Şekil 1: Önerilen göğüs röntgeni sınıflandırma yaklaşımı öncelikle bir VGG16 yapısı ile özellik haritaları oluşturulur. Daha sonra bu özellik haritaları eşit büyüklükte parçalara ayrılarak FMA tarafından işlenerek çoklu etiketlere sahip sınıflandırma işlemine uygun çıkış vektörü üretir.

$$E(y_i, \tilde{y}_i) = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\tilde{y}_i) + (1 - y_i) \log(1 - \tilde{y}_i)] \quad (4)$$

B. Ağ Mimarisi Detayları ve Eğitim

Tüm deneyler TensorFlow kütüphanesi kullanılarak iki adet NVidia RTX 3090 ekran kartı ile gerçekleştirilmiştir. Uçtan uca eğitim 'Adam' optimizasyon yaklaşımı ile 120 epok eğitilmiştir. Öğrenme oranı olarak 10^{-4} ve mini küme olarak 30 göğüs röntgeni görüntüsü kullanılmıştır. ImageNet veri seri ile ön eğitilmiş VGG16 tarafından üretilen özellik haritalarının boyutları 7×7 ve derinliği ise 512'dir. Özellik haritasını 4 eşit parçaya bölebilmek için sıfır-ekleme yöntemiyle harita etrafına '0' değeri ekleyerek 8×8 boyutlarına çıkarttık. Böylece FMA girişine 4 adet $4 \times 4 \times 512$ boyutlarında parçalı özellik haritaları iletilir. FMA mimarisinde orijinal parametre değerleri korundu [21].

III. DENEYSEL SONUÇLAR

A. Veri Kümesi

Deneylerde 30,805 farklı hastadan elde edilen 112,120 adet görüntüden oluşan bir göğüs röntgeni veri kümesi kullanılmaktadır [23]. Görüntüler yaşları 1-95 arasında değişen hastalardan alınmıştır ve verilerin %56.49'u erkek %43.51'i kadın hastalara aittir. Her görüntüde "atelectasis, cardiomegaly, effusion, infiltration, mass, nodule, pneumonia, pneumothorax, consolidation, edema, emphysema, fibrosis, p-thickening, hernia" olmak üzere 14 sınıftan bir veya birkaçı bulunabilir veya 'No Finding' ifadesi tek başına bulunabilir. Veri kümesinin sahip olduğu çoklu etiket durumu, sınıflar arası veri sayı dengesizliği (çoğunluk sınıf 25366 örnek, azınlık sınıf 284 örnek) ve lezyonların genellikle tüm görüntü alanının sadece %0.075'ini kaplaması sınıflandırma algoritmaları için zorlayıcı etmenlerdir.

B. Sonuçlar

Bu çalışmada çoklu etiketli verilerin sınıflandırılması amacıyla önerilen FMA tabanlı yaklaşımın performansı ön-eğitilmiş bir VGG16 yapısı, VGG16 çıkışına eklenmiş bir ViT yapısı, 'transition' blokları bir lineer derin ağ (Transition) [5], medikal görüntüler için transfer öğrenmesi yaklaşımı (MorphHR) [10], paralel bağlı çoklu KSA'lar ile etiket ilişkilerine dayalı ağ (LCL) [15], ve paralel öz dikkat modülleri (DuaLAnet) [8] yaklaşımları kullanılarak karşılaştırılmıştır. Tablo 1'de her bir hastalık sınıfının AUC performansları ayrı ayrı hesaplanmış ve en sonunda ise ortalama AUC performansı hesaplanmıştır. Buna göre FMA'nın 'Edema' ve 'Pneumonia' sınıfları için en yüksek performansı sergilediği, diğer sınıflar için ortalama üstü performanslar sergilediği gözlenmektedir. Ortalama olarak, önerilen yaklaşım rakiplerine göre daha başarılıdır ve ikinci en iyi yaklaşıma göre %0.2 daha üstün performans sergilemektedir.

Önerilen tekniğin, 'Edema' ve 'Pneumonia' gibi ortalama veri sayısına sahip sınıflar için yüksek performans üretmesi çeşitli geliştirmeler ile çoğunluk ve azınlık sınıfları içinde en üst performansları üretebileceğinin göstergesidir. Yapılan deneysel çalışmalar FMA tabanlı yaklaşımların medikal görüntüler için ve özellikle çoklu etiketlere sahip veri kümeleri için üstün sergileyebileceğini ortaya koymaktadır. Performansın daha da artması için veri sayısı dengesizliği ile başa çıkabilecek daha adaptif bir çözüm, VGG16'dan daha gürbüz bir özellik haritası oluşturma yaklaşımı, bölütleme veya ek dikkat yardımcı bir mimari gibi çeşitli çözümler üzerinde durulabilir.

IV. TARTIŞMA

Bu çalışmada öz-dikkat yaklaşımlarının sahip olduğu dezavantajların üstesinden gelmek için ortaya atılan fokal modülasyon ağlarının çoklu etiketli göğüs röntgeni görüntülerinin sınıflandırma performansı analiz edilmiştir. Bu amaçla, ilk olarak hastalık içermeyen bölgelerin elenmesi ve bellek dostu özellik haritalarının oluşturulması için bir VGG16 amaca uygun

TABLO I: ÖNERİLEN YAKLAŞIM VE DİĞER TEKNİKLERİN AUC PERFORMANSLARININ KARŞILAŞTIRILMASI.

	VGG16	VGG16 ve ViT	Transition	MorphHR	LCL	DualAnet	Önerilen
Atelectasis	73.8	79.3	81.4	77.6	77.3	78.3	78.5
Cardiomegaly	82.1	84.2	89.9	86.5	88.0	88.4	89.3
Consolidation	70.6	80.4	78.9	77.6	73.6	74.6	79.6
Edema	78.8	84.7	87.4	85.4	84.5	84.1	87.4
Effusion	82.2	87.5	87.3	83.0	82.3	83.2	86.7
Emphysema	82.8	90.2	92.4	89.8	93.8	93.7	91.1
Fibrosis	70.4	76.4	80.9	81.4	82.3	82.0	81.3
Hernia	78.0	87.2	92.3	88.2	89.9	91.2	85.9
Infiltration	65.7	70.2	70.1	72.4	69.7	70.8	70.1
Mass	75.5	86.6	84.0	83.8	82.6	83.7	85.2
Nodule	67.5	75.2	77.5	77.1	78.6	80.0	75.2
P-Thickening	70.3	78.5	77.2	80.0	78.2	79.6	78.3
Pneumonia	62.5	69.6	66.2	72.1	72.0	73.5	73.5
Pneumothorax	83.5	88.4	86.5	87.0	87.1	86.6	88.0
ORTALAMA	74.6	81.3	82.0	81.6	81.4	82.0	82.2

düzenlenmiştir. Ardından özellik haritaları fokal modülasyon ile işlenerek özellik vektörleri üretilmiştir. Önerilen yaklaşım tarafından üretilen sonuçlar literatürdeki diğer yöntemlere göre daha yüksek performans sergilemektedir. Bu durum, fokal modülasyon içerikli yaklaşımların çoklu etiketli göğüs röntgen görüntülerinin sınıflandırılması problemleri için elverişli olduğunu ortaya koymaktadır. Gelecek çalışmalar, veri sayısı dengesizliğini dikkate alan kayıp fonksiyonları ve probleme özel fokal modülasyon yaklaşımlarının araştırılmasını kapsayacaktır.

KAYNAKLAR

- [1] R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics, 2013," *CA: A Cancer Journal for Clinicians*, vol. 63, no. 1, pp. 11–30, 2013.
- [2] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," *CA: A Cancer Journal for Clinicians*, vol. 73, no. 1, pp. 17–48, 2023.
- [3] M. E. Kruk, A. D. Gage, N. T. Joseph, G. Danaei, S. García-Saisó, and J. A. Salomon, "Mortality due to low-quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries," *The Lancet*, vol. 392, no. 10160, pp. 2203–2212, 2018.
- [4] Y. Korkmaz, S. U. H. Dar, M. Yurt, M. Özbey, and T. Çukur, "Unsupervised mri reconstruction via zero-shot learned adversarial transformers," *IEEE Transactions on Medical Imaging*, vol. 41, no. 7, pp. 1747–1763, 2022.
- [5] H. Wang, Y.-Y. Yang, Y. Pan, P. Han, Z.-X. Li, H.-G. Huang, and S.-Z. Zhu, "Detecting thoracic diseases via representation learning with adaptive sampling," *Neurocomputing*, vol. 406, pp. 354–360, 2020.
- [6] G. Elmas, S. U. Dar, Y. Korkmaz, E. Ceyani, B. Susam, M. Özbey, S. Avestimehr, and T. Çukur, "Federated learning of generative image priors for mri reconstruction," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2022.
- [7] A. Güngör, S. U. Dar, Ş. Öztürk, Y. Korkmaz, G. Elmas, M. Özbey, and T. Çukur, "Adaptive diffusion priors for accelerated mri reconstruction," *arXiv:2207.05876*, 2022.
- [8] V. Teixeira, L. Braz, H. Pedrini, and Z. Dias, "Dualanet: Dual lesion attention network for thoracic disease classification in chest x-rays," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020, pp. 69–74.
- [9] M. Özbey, O. Dalmaz, S. U. Dar, H. A. Bedel, Ş. Öztürk, A. Güngör, and T. Çukur, "Unsupervised medical image translation with adversarial diffusion models," *arXiv:2207.08208*, 2022.
- [10] T. Wei, A. I. Aviles-Rivero, S. Wang, Y. Huang, F. J. Gilbert, C.-B. Schönlieb, and C. W. Chen, "Beyond fine-tuning: Classifying high resolution mammograms using function-preserving transformations," *Medical Image Analysis*, vol. 82, p. 102618, 2022.
- [11] I. Sivgin, H. A. Bedel, Ş. Öztürk, and T. Çukur, "A plug-in graph neural network to boost temporal sensitivity in fmri analysis," *arXiv preprint arXiv:2301.00439*, 2023.
- [12] K. Chen, X. Wang, and S. Zhang, "Thorax disease classification based on pyramidal convolution shuffle attention neural network," *IEEE Access*, vol. 10, pp. 85 571–85 581, 2022.
- [13] M. S. Lee and S. W. Han, "Duetnet: Dual encoder based transfer network for thoracic disease classification," *Pattern Recognition Letters*, vol. 161, pp. 143–153, 2022.
- [14] S. U. Dar, Ş. Öztürk, M. Özbey, and T. Çukur, "Learning deep mri reconstruction models from scratch in low-data regimes," *arXiv preprint arXiv:2301.02613*, 2023.
- [15] B. Chen, Y. Lu, and G. Lu, "Multi-label chest x-ray image classification via label co-occurrence learning," in *Pattern Recognition and Computer Vision: Second Chinese Conference, PRCV 2019, Xi'an, China, November 8–11, 2019, Proceedings, Part II 2*. Springer, 2019, pp. 682–693.
- [16] Ş. Öztürk and T. Çukur, "Deep clustering via center-oriented margin free-triplet loss for skin lesion detection in highly imbalanced datasets," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 9, pp. 4679–4690, 2022.
- [17] H.-G. Jung, W.-J. Nam, H.-W. Kim, and S.-W. Lee, "Weakly supervised thoracic disease localization via disease masks," *Neurocomputing*, vol. 517, pp. 34–43, 2023.
- [18] B. Chen, J. Li, G. Lu, and D. Zhang, "Lesion location attention guided network for multi-label thoracic disease classification in chest x-rays," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 7, pp. 2016–2027, 2020.
- [19] H. H. Pham, T. T. Le, D. Q. Tran, D. T. Ngo, and H. Q. Nguyen, "Interpreting chest x-rays via cnns that exploit hierarchical disease dependencies and uncertainty labels," *Neurocomputing*, vol. 437, pp. 186–194, 2021.
- [20] O. Dalmaz, M. Yurt, and T. Çukur, "Resvit: Residual vision transformers for multimodal medical image synthesis," *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2598–2614, 2022.
- [21] J. Yang, C. Li, X. Dai, and J. Gao, "Focal modulation networks," *Advances in Neural Information Processing Systems*, vol. 35, pp. 4203–4217, 2022.
- [22] Ş. Öztürk, E. Çelik, and T. Çukur, "Content-based medical image retrieval with opponent class adaptive margin loss," *Information Sciences*, vol. 637, p. 118938, 2023.
- [23] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3462–3471.