



## Endüstriyel Kontrol Sistemlerinde Yenilikçi Anomali Tespit Sistemlerinin İncelenmesi

### Investigation of Innovative Anomaly Detection Systems in Industrial Control Systems

<sup>1,2</sup>Kerem ÇINAR , <sup>2</sup>Murat İSKEFİYELİ 

<sup>1</sup>*İstanbul Gedik Üniversitesi, Bilgisayar Programcılığı Programı, 34722, İstanbul, Türkiye*

<sup>2</sup>*Sakarya Üniversitesi, Bilgisayar ve Bilişim Bilimleri Fakültesi, 54050, Sakarya, Türkiye*

<sup>1</sup>kerem.cinar@gedik.edu.tr, <sup>2</sup>miskef@sakarya.edu.tr

Araştırma Makalesi/Research Article

#### ARTICLE INFO

##### Article history

Received: 5 January 2023

Accepted: 12 April 2023

##### Keywords:

Deep Learning, GAN, Industrial Control Systems, Machine Learning

#### ABSTRACT

Industrial Control Systems (ICS) or SCADA networks are becoming targets of cyber-attacks as their architectures move from proprietary hardware, software, and protocols to standard and open sources. Large-scale sensor data makes anomalies and cyber-attack events continuously monitored. Current unsupervised machine learning approaches have not fully exploited the spatiotemporal correlation and other dependencies between sensors in the system to detect anomalies. This article reviews the approaches used to detect anomalies in SCADA networks of various architectures such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Stacked Autoencoder (SAE), and Long Short-Term Memory. In addition to reviews of these methods in the article, an unsupervised multivariate anomaly detection method based on Generative Contradictory Networks (GANs) using Long-Short-Term-Memory Recurrent Neural Networks (LSTM-RNN) as basic models (i.e. generator and discriminator) is presented.

© 2023 Bandırma Onyeddi Eylül University, Faculty of Engineering and Natural Science. Published by Dergi Park. All rights reserved.

#### MAKALE BİLGİSİ

##### Makale Tarihleri

Gönderim : 5 Ocak 2023

Kabul : 12 Nisan 2023

##### Anahtar Kelimeler:

Derin Öğrenme, GAN, Endüstriyel Kontrol Sistemleri, Makine Öğrenimi

#### ÖZET

Endüstriyel Kontrol Sistemleri (ICS) veya SCADA ağları, mimarileri tescilli donanım, yazılım ve protokollerden standart ve açık kaynaklara geçtikçe siber saldırıların hedefi haline gelmektedir. Büyük ölçekli sensör verileri, olağan dışı durumları ve siber saldırı olaylarını sürekli olarak izlenebilir kılmaktadır. Mevcut denetimsiz makine öğrenimi yaklaşımları, anomalileri tespit etmek için sistemdeki sensörler arasındaki uzamsal-zamansal korelasyonu ve diğer bağımlılıkları tam olarak kullanmamıştır. Bu makale, Konvolüsyonel Sinir Ağı (CNN), Tekrarlayan Sinir Ağı (RNN), Stacked Autoencoder (SAE), Uzun Kısa Süreli Bellek gibi çeşitli mimarilerin SCADA ağlarındaki anomalilerin tespit edilmesinde kullanılan yaklaşımların incelenmesidir. Ayrıca makalede bu yöntemlerin incelenmesine ek olarak Uzun-Kısa Süreli-Bellek Tekrarlayan Sinir Ağlarını (LSTM-RNN) temel modeller (yani, üretici ve ayırıcı) olarak kullanan, Üretken Çelişkili Ağlara (GAN'lar) dayalı denetimsiz çok değişkenli bir anomali tespit yöntemini detaylı olarak sunmaktadır.

© 2023 Bandırma Onyeddi Eylül Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi. Dergi Park tarafından yayınlanmaktadır. Tüm Hakları Saklıdır.

ORCID: <sup>1</sup>0000-0002-6098-5521

<sup>2</sup>0000-0002-8210-5070

## 1. GİRİŞ

Endüstriyel kontrol sistemleri (ICS) yaygın olarak ilaç endüstrisi, imalat, elektrik, su arıtma tesisleri ve petrol rafinerileri gibi kritik altyapılar dahil olmak üzere çeşitli sektörlerin çalışması için hayati önem taşımaktadır.

Bu sistemler fiziksel olarak güvenli konumlarda özel donanım ve yazılımlar üzerinde çalışıyordu, ancak daha yakın zamanlarda ortak bilgi (BT) teknolojilerini ve uzaktan bağlantıyı kullanıma sundular. Bu değişiklikler, siber güvenlik açıklarını ve anomali olasılığını artırmaktadır [1]. ICS'lere yönelik siber saldırıları tespit etme yeteneği kritik bir görev haline geldi. Bu sorunu çözmek için kullanılan yaklaşımlardan biri, kötü amaçlı etkinliği belirlemek için geleneksel BT ağ tabanlı saldırı tespit Sistemlerini (IDS'ler) kullanmayı içerir. Bu çalışmada, sistemin anomali davranışını fiziksel düzeyde tespit etmeye çalıştığımız bir anomali tespit yaklaşımına odaklanıyoruz. Bu yaklaşım, saldırganın nihai amacının sistemin fiziksel davranışını etkilemek olduğu varsayımına dayanır ve sistemi ağ düzeyindeki savunma hattının ötesinde korumayı amaçlar. Fiziksel seviye tabanlı anomali tespiti, büyük bir ekonomik değere sahip olabilecek hatalı ekipmanın erken tespitini ve düzeltilmesini de kolaylaştırabilir.

Anomali tespit yöntemleri, sistemin kurallarına veya modellerine dayalı olabilir [2]. ICS'lere yönelik siber saldırıları tespit etme yeteneği kritik bir görev haline geldi. Bu sorunu çözmek için kullanılan yaklaşımlardan biri, kötü amaçlı etkinliği belirlemek için geleneksel BT ağ tabanlı saldırı tespit Sistemlerini (IDS'ler) kullanmayı içerir. Bu çalışmada, sistemin anomali davranışını fiziksel düzeyde tespit etmeye çalıştığımız bir anomali tespit yaklaşımına odaklanıyoruz.

Bu yaklaşım, saldırganın nihai amacının sistemin fiziksel davranışını etkilemek olduğu varsayımına dayanır ve sistemi ağ düzeyindeki savunma hattının ötesinde korumayı amaçlar. Fiziksel seviye tabanlı anomali tespiti, büyük bir ekonomik değere sahip olabilecek hatalı ekipmanın erken tespitini ve düzeltilmesini de kolaylaştırabilir.

Anomali tespit yöntemleri, sistemin kurallarına veya modellerine dayalı olabilir [3]. Ne yazık ki, karmaşık fiziksel süreçlerin kesin bir modelini oluşturmak çok zor bir iştir. Zaman alıcı ve büyük ve karmaşık sistemlere ölçeklenemeyen sistemin ve uygulamasının derinlemesine anlaşılmasını gerektirir. Son zamanlarda ilgi odağı haline gelen başka bir yaklaşım, ICS'leri modellemek ve anomali davranışları tespit etmek için makine öğrenimini kullanır. Yakın zamanda, ICS'lerde anomali tespiti için denetimli makine öğrenimini kullanan bir dizi çalışma yayınlandı [4]. Bu yaklaşım, normal ve saldırı senaryoları için etiketlenmiş eğitim verileri gerektirir, ancak siber saldırılar için etiketlenmiş verilerin elde edilmesi zor olabilir ve bu veriler doğal olarak bilinmeyen saldırı sınıflarını içermeyecektir. Son zamanlarda, denetimsiz makine öğreniminin, güvenli siber-fiziksel sistemlerin tasarımıyla ilgili araştırmaları desteklemek için oluşturulmuş özel bir su tesisi test ortamından (WADI) elde edilen verileri kullanarak siber saldırıları tespit etmede etkili olduğu gösterildi.

Bir anomali genellikle, sistem davranışının önceki normal durumdan önemli ölçüde farklı olduğu belirli zaman adımlarındaki noktalar olarak tanımlanır [5]. Anomali tespitinin temel görevi, bir anomalinin meydana gelmiş olabileceği zaman adımlarını belirlemektir. Geleneksel olarak, Cusum, Ewma ve Shewhart çizelgeleri gibi İstatistiksel Proses Kontrolü (SPC) yöntemleri, aralık dışı olan çalışma durumlarını bulmak için endüstriyel proseslerin kalitesini izlemek için popüler çözümlerdi [6]. Bu geleneksel algılama teknikleri, modern CPS'lerin giderek artan dinamik ve karmaşık doğası tarafından üretilen çok değişkenli veri akışlarıyla başa çıkamamaktadır. Bu nedenle, araştırmacılar spesifikasyon veya imza tabanlı tekniklerin ötesine geçtiler ve sistemler tarafından üretilen büyük miktarda veriden yararlanmak için makine öğrenimi tekniklerinden yararlanmaya başladılar [7]. Etiketli verilerin doğası gereği eksikliği nedeniyle, anomali algılama genellikle denetimsiz bir makine öğrenimi görevi olarak değerlendirilir. Bununla birlikte, mevcut denetimsiz yöntemlerin çoğu, çok değişkenli zaman serilerinin içsel korelasyonlarında doğrusal olmayan idare edemeyen doğrusal izdüşüm ve dönüşüm yoluyla oluşturulur. Ayrıca, mevcut tekniklerin çoğu, sistemlerin son derece dinamik doğası göz önüne alındığında yetersiz olabilen anomalileri tespit etmek için mevcut durumlar ve öngörülen normal aralıklar arasında basit karşılaştırmalar kullanır.

Son yıllarda Derin Öğrenme [8] yaklaşımları, doğal dil işleme (NLP), görüntü-video sınıflandırma gibi çalışmalarda oldukça popüler hale gelmiştir. Ağlarda anomali tespiti için çeşitli derin öğrenme yaklaşımları kullanılmıştır [9,10,11]. Derin öğrenmenin en önemli özelliklerinden biri, derin öğrenme modellerinde hiyerarşik özellikleri otonom olarak öğrenmek için denetimsiz yöntemlerin kullanılmasıdır [12,13,14,15,16]. Aslında, verilerin en göze çarpan özellikleri, derin mimarilerin otomatik öğrenme yeteneği kullanılarak denetimsiz bir şekilde öğrenmesi ve bu öğrenilen özelliklerin, anomali verileri normal olanlardan ayırt etmek için bir sınıflandırıcıda kullanılmasıdır. Bu makalede, derin özellik öğrenme yaklaşımını kullanan SCADA ağları ve Generative Adversarial Networks (GAN) dayalı denetimsiz çok değişkenli anomali tespit sistemleri incelenmiştir.

## 2. YÖNTEM

Denetimli algoritmaları eğitmek için etiketlenmiş anomali verilerinin doğal eksikliği göz önüne alındığında, anomali algılama yöntemleri çoğunlukla denetimsiz yöntemlere dayanmaktadır. Denetimsiz algılama yöntemlerini dört kategoriye ayırabiliriz: (i) doğrusal model tabanlı yöntem, (ii) mesafe tabanlı yöntemler, (iii) olasılık ve yoğunluk tahmini tabanlı yöntemler ve (vi) son zamanlarda oldukça popüler olan derin öğrenme tabanlı yöntemler.

Doğrusal model tabanlı denetimsiz anomali tespit yöntemleri için popüler bir yaklaşım Temel Bileşen Analizi (PCA) [14]. PCA, temel olarak, süreç ölçümlerinden çıkarılan önemli değişkenlik bilgilerini koruyan ve büyük miktarda ilişkili veri için boyutu azaltan çok değişkenli bir veri analizi yöntemidir [15]. PLS, model oluşturma ve anomali tespiti için yaygın olarak kullanılan başka birçok değişkenli veri analizi yöntemidir [16]. Ancak, bunlar yalnızca yüksek korelasyonlu veriler için etkilidir ve verilerin çok değişkenli Gauss dağılımını takip etmesini gerektirir [17].

Uzaklığa dayalı yöntemler için popüler bir yaklaşım, en yakın  $k$  komşusuna olan ortalama mesafeyi hesaplayan ve bu uzaklığa dayalı anomali puanları elde eden  $K$  En Yakın Komşu (KNN) algoritmasıdır. Clustering - Based Local Outlier Factor (CBLOF) yöntemi, uzaklığa dayalı yöntemlere başka bir örnektir. Yerel Aykırı Değer Faktörü (LOF) yönteminin geliştirilmiş bir versiyonu olan kümelemeye dayalı anomalileri belirlemek için önceden tanımlanmış bir anomali puanı işlevi kullanır [18]. Bazı durumlarda etkili olmakla birlikte, bu mesafeye dayalı yöntemler, anomali süreleri ve anomalilerin sayısı hakkında önceden bilgi sahibi olduğunda daha iyi performans gösterir.

Olasılıksal model tabanlı ve yoğunluk tahmini tabanlı yöntemler, veri dağılımlarına daha fazla dikkat edilerek mesafe tabanlı yöntemlerin iyileştirilmesi olarak önerilmiştir. Örneğin, Açık Tabanlı Aykırı Değer Algılama (ABOD) yöntemi [19] ve Özellik Paketleme (FB) yöntemi [20] değişken korelasyonları dikkate alarak verilerle ilgilenir. Ancak, bu yöntemler zaman adımları boyunca zamansal korelasyonu dikkate alamamaktadır ve bu nedenle çok değişkenli zaman serisi verileri için iyi çalışmamaktadır.

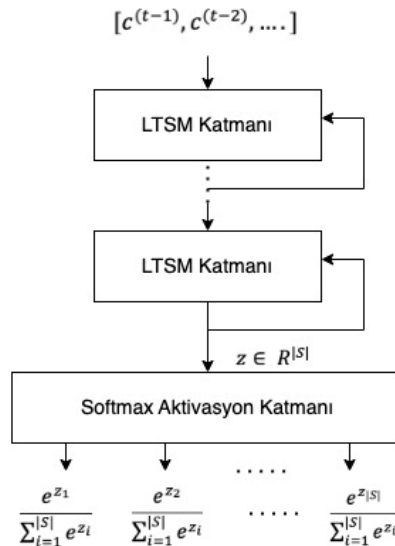
Derin öğrenme tabanlı denetimsiz anomali tespit yöntemleri, gelecek vaat eden performanslarıyla son zamanlarda çok popülerlik kazanmıştır. Örneğin, Otomatik Kodlayıcı (AE) [21], yeniden yapılandırma hatalarını inceleyerek anomali tespiti için popüler bir derin öğrenme modelidir. Derin Otomatik Kodlama Gauss Karışım Modeli (DAGMM) [22] ve LSTM Encoder-Decoder [23] gibi diğerleri de çok değişkenli anomali tespiti için iyi performans bildirmiştir. Bu çalışmada, derin öğrenme tabanlı denetimsiz anomali tespit yöntemlerinin umut verici başarısını takip ediyoruz ve Generative Adversarial Networks (GAN) temelinde oluşturulmuş yeni bir derin öğrenme tabanlı denetimsiz anomali tespit stratejilerini inceliyoruz.

Gerçekleştirilen bu çalışmada çok değişkenli zaman serileri için anomalileri tespit eden GAN tabanlı bir denetimsiz anomali tespit yöntemi olan MAD-GAN incelenmiştir. MAD-GAN mimarisi, GAN tarafından yakalanması için öğrenilen temel modeller olarak Uzun Kısa Vadeli-Tekrarlayan Sinir Ağlarını (LSTM-RNN) benimseyerek çok değişkenli zaman serisi verilerini analiz etmek için görüntü ile ilgili uygulamalar için önceden geliştirilmiş GAN çerçevesini uyarlar. Zamansal bağımlılık, her test numunesi için ayırım sonuçlarını ve yeniden yapılandırma kalıntılarını birleştiren yeni bir anomali skoru kullanarak anomalileri tespit etmek için hem GAN'ın ayırıcısını hem de oluşturucusu kullanmaktadır. İncelenen çalışmada, MAD-GAN'ın, iki CPS veri kümesi için siber saldırıların neden olduğu anomalileri tespit etmede mevcut yöntemlerden daha iyi performans ortaya koyduğu gösterilmiştir.

### 3. SCADA ANOMALİ TESPİT SİSTEMLERİNDE DENETİMSİZ ÖZELLİK ÖĞRENİMİNİN İNCELENMESİ

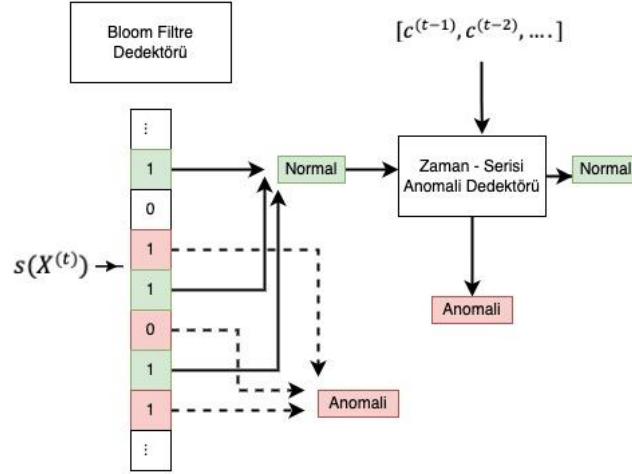
#### 3.1. LSTM/Bloom Filtresi Anomalilik Dedektörü

Bir gaz boru hattı SCADA sistemine veri/komut enjeksiyonu, keşif veya Hizmet Reddi (DoS) saldırılarından kaynaklanan anomalileri tespit etmek için [8], iki detektörden oluşan bir anomali tespit yaklaşımı önermektedir (Şekil 1).



Şekil 1. Yığılmış LSTM tabanlı softmax modeli.

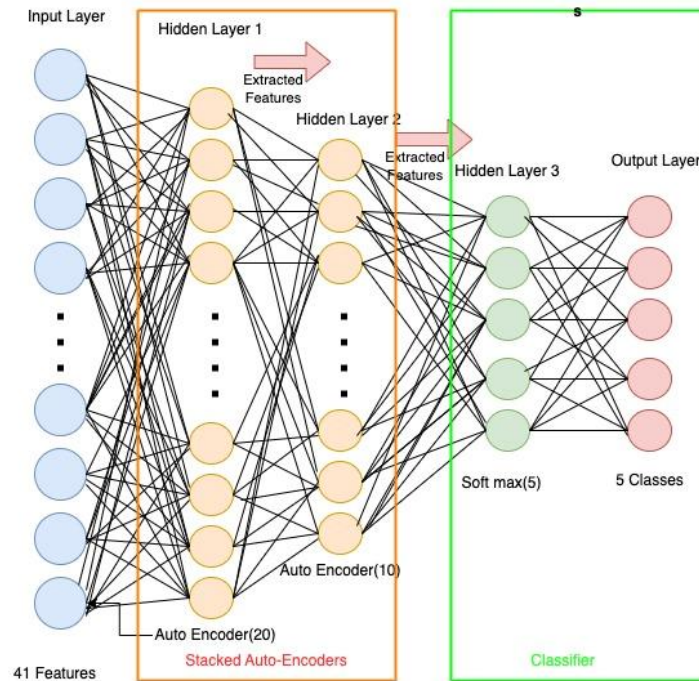
İlki, veri-tabanındaki bir paket imzasını kontrol eden paket düzeyinde bir anomali dedektörüdür. Veri tabanı, bir SCADA sisteminde kararlı oldukları için ağ modellerini ve iletişim modeli imzasını saklar. Bloom filtresi, analiz edilen paketin imzasını içermiyorsa paket anomali kabul edilir. Bir sonraki dedektör, bir sonraki adımın davranışını tahmin etmek için zaman adımlarının sayısı için bilgi ezberleme gücünü kullanan başka bir algılama seviyesi için Bloom filtresini geçen normal paketi alır. Bazı SCADA bileşenlerinin sınırlı bellek ve bilgi işlem kaynakları nedeniyle, Bloom filtresi olarak hızlı ve hafif ağırlıklı bir anomali dedektörü kullanmak büyük önem taşımaktadır. Zaman serilerinin girdisini alan LSTM Anomali Dedektörü (Şekil 2), çok sınıflı sınıflandırmaya uygun bir softmax fonksiyonunu en aza indirmek için eğitilerek bir sonraki veri noktasını tahmin etmek için önemli özelliklerini öğrenir [7,21]. Bir gaz boru hattı SCADA veri seti [22] üzerinde birleşik anomali tespit çerçevesinin değerlendirilmesi, diğer yaklaşımlara kıyasla daha yüksek olan %92'lik bir doğruluk oranı verir. Ancak, 35 dakikalık LSTM modelini 50 dönem boyunca eğitmek için gereken süre oldukça yüksektir.



Şekil 2. Paket ve zaman serisi düzeyinde anomali tespiti için birleşik çerçeve.

### 3.2. Yığılmış Otomatik Kodlayıcı (Stacked Auto-Encoder) Tabanlı Anomali Tespiti

Ağ bant genişliği ve veri artışı nedeniyle [23], DoS, Probe, R2L ve U2R saldırılarının tespitine izin verecek gerekli özelliği öğrenmek için derin bir paket incelemesi önermektedir. Yazarlar, mimarinin, sınıflandırma için bir softmax katmanının eklendiği, özellik öğrenme için yığılmış bir otomatik kodlayıcı olduğu bir Derin Sinir Ağları (DNN) yaklaşımı kullanmışlardır (Şekil 3). Şekil 3 'te yer alan DNN mimarisinden görüleceği üzere, sistem Giriş Katmanı (İnput Layer), Gizli Katmanlar (Hidden Layer 1, Hidden Layer 2, Hidden Layer 3, Auto Encoder) ve Çıkış Katmanı (Output Layer) bölümlerinden oluşmaktadır.



Şekil 3. DNN mimarisi.

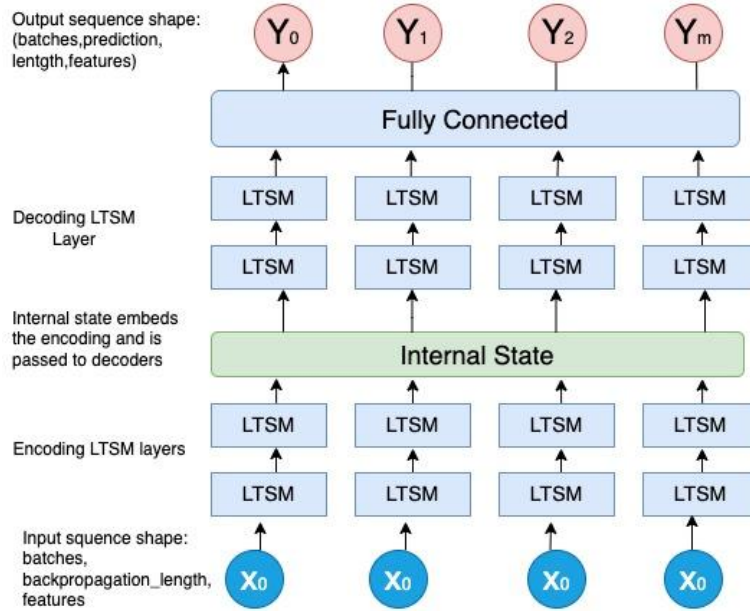
İstiflenmiş otomatik kodlayıcı, biri 20 düğümlü ve ikincisi 10 düğümlü olmak üzere iki gizli katmana sahiptir. Öğrenilen özelliklerin boyutu, NSL-KDD veri setinin 41 orijinal özelliğine kıyasla 10'dur. Genel süreç dört adımı kapsar, yani yığınlı otomatik kodlayıcı ile bir özellik öğrenme adımı, denetimli bir softmax eğitimi ile ilk ince ayar adımı. Bu ilk ince ayar adımının girdisi, verilerin sıkıştırılmış temsidir. Bir sonraki adım, ilk ince ayar adımından sonra tüm ağ katmanlarına uygulanan bir geri yayılım eğitimi ile ikinci bir ince ayardır. İkinci ince ayar adımı, kayıp fonksiyonunu en aza indirmek için ağ ağırlıklarını ayarlayarak saldırı tespit görevi için daha alakalı hale getirmek için ara katmanların özelliklerini iyileştirmeyi amaçlar.

Son olarak, sürecin son adımı, modelin verimliliğini değerlendirmek için ince ayarlı ağa bir test veri setinin sunulduğu bir sınıflandırma ve test adımıdır. Önerilen yaklaşımı k-means, DBN, SOM, AdaBoost gibi standart tekniklere karşı değerlendirmek için hatırlama, doğruluk, kesinlik ve f-mesure metrikleri kullanılır.

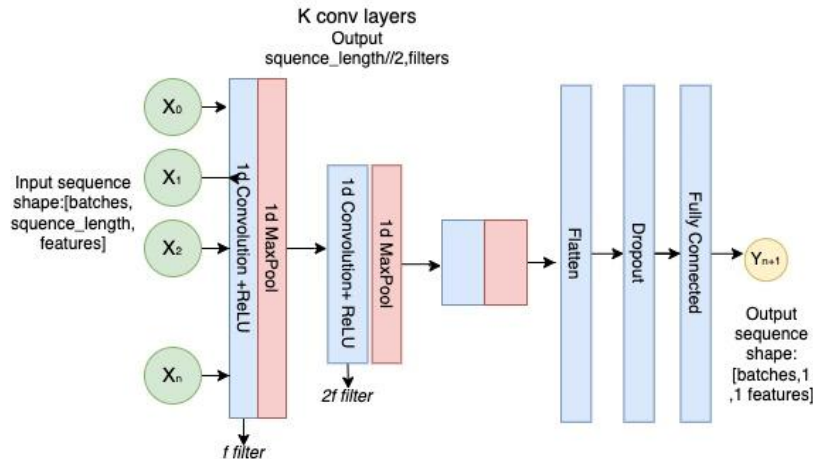
Deneysel sonuçlar, DoS ve Probe saldırıları için iyi algılama doğruluğuna (sırasıyla %97,6 ve %86,34) rağmen, R2L(Remote to user) ve U2R(User to Root) saldırılarının zayıf sonuçlar verdiğini (sırasıyla %12,98 ve %39,62) göstermektedir. Son iki saldırı kategorisinin düşük performansı, R2L ve U2R ile ilgili yeterli miktarda veri bulunmamasından kaynaklanmaktadır (sırasıyla %0,04 ve %0,79). Prob saldırılarında olduğu gibi R2L ve U2R saldırı kategorileri için %9 ila %10 eğitim verisi örnekleri daha iyi tespit sonuçları verebilirdi.

### 3.3. SCADA Üzerinde CNN/LSTM Anomali Tespiti

Güvenli Su Arıtma test yatağı (SWaT) veri seti, 36'ya kadar farklı siber saldırı içerir. Bu tür bir sistemde izinsiz giriş tespiti için denimsiz özellik öğrenmenin kullanımını değerlendirmek için [28], özellik öğrenici olarak LSTM (Şekil 4) veya 1D CNN kullanan iki model önermektedir (Şekil 5).

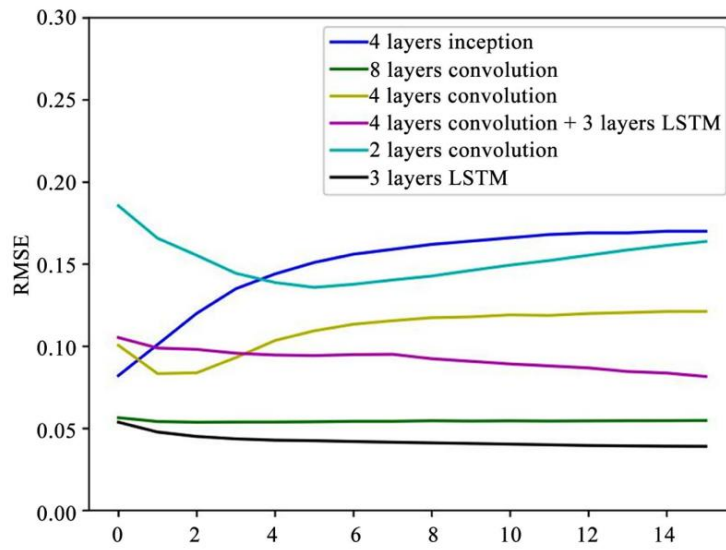


Şekil 4. LSTM autoencoder modeli.



Şekil 5. 1D geleneksel sinir ağları.

Bir hata işlevi olarak ortalama MSE'yi ve ağırlık azalmasıyla AdamOptimizer'ı kullanırlar. Bir düzenleme tekniği olarak ağırlık azalması, modelin fazla takılmasını önler ve AdamOptimizer [29] hesaplama açısından verimlidir ve çok az bellek gerektirir. İlk Derin Sinir Ağı (DNN) mimarisi, sınıflandırma amacıyla en üstte tamamen bağlı bir katmana sahip yığılmış bir LSTM (Long short-term memory)'dir. LSTM modeliyle, bir öğrenme oranı (0,001 ile 0,00001 arasında) ve bir azalma değeri (0,9 ile 0,99 arasında) ayarlayarak, LSTM katmanlarının çeşitli derinliklerini (64'ten 2048'e) ve dizi uzunluklarını (50 ile 1000 arasında) test edebildiler. 1D CNN mimarisi, ReLU-MaxPooling şemasını benimsemiştir. Deneyler için farklı çekirdek boyutları kullanılmıştır. Konvolüsyonlar katmanının üstüne, tahmin için tamamen bağlı bir katman eklenir ve fazla uydurmayı önlemek için bırakma kullanılır. Yazarlar, bir toplu normalleştirme katmanı ekleyerek veya temel CONV-RELU-POOL bloğunu (CONV-RELU)  $\times$  N-MAXPOOL mimarisi ile değiştirerek bu CNN mimarisinin çeşitli varyasyonlarını test ettiler. Ayrıca, evrişim katmanlarını, daha iyi performans ve daha düşük hesaplama maliyeti sağladığı bilinen Başlangıç katmanları [30] ile değiştirdiler. Başlangıç katmanları, evrişim katmanları tarafından kullanılan tam bağlantılar yerine seyrek seyrek bağlantıları kullanır, bu nedenle hesaplama yükü azalır. Deneyler, 36 farklı siber saldırı içeren SWaT veri seti üzerinde gerçekleştirilmiştir. Önerilen 1D CNN modeli %89 algılama oranına sahiptir, bu oldukça iyidir, ancak iyileştirilmesi gerekmektedir.



Şekil 6. Test hataları.

Farklı mimarilerin karşılaştırılması (Şekil 6), LSTM'lerin ve başlangıç tabanlı evrişimin yalnızca daha hızlı yakınsamadığını, aynı zamanda daha düşük eğitim hatası oranı verdiğini de göstermektedir. Anomali algılama yöntemi, sekiz katmanlı evrişimli ağ için yüksek Eğri Altında Alan (AUC), yani 0,967 verir. CNN(Convolutional neural network)'nin eğitim ve test süreleri LSTM ağına göre daha düşüktür. CNN ağları, LSTM ağlarına kıyasla anomali tespiti için iyi performans gösterdi. Önerilen CNN, %100 hassasiyetle %85'e ulaşan algılama oranlarına sahiptir.

### 3.4. Üretken Düşman Eğitimi (Generative Adversarial Training - GAN) ile Anomali Tespiti

Zaman serileri için anomali tespitinin temel görevi, test verilerinin normal veri dağılımlarına uyup uymadığını belirlemektir; uyumsuz noktalara çeşitli uygulama alanlarında anomaliler, aykırı değerler, izinsiz girişler, arızalar veya kirleticiler denir (7). Şekil 7, önerilen GAN'ın genel mimarisini göstermektedir.

#### 3.4.1. GAN Mimarisi

Tipik olarak, bir GAN iki ağdan oluşur: üretici ve ayırıcı (aka eleştirici). Oluşturucu, gizli bir koddan bir örnek, örneğin bir görüntü üretir ve bu görüntülerin dağılımı ideal olarak eğitim dağılımından ayırt edilemez olmalıdır. Durumun böyle olup olmadığını söyleyen bir fonksiyon tasarlamak genellikle mümkün olmadığından, değerlendirmeyi yapmak için bir ayırıcı ağ eğitilir ve ağlar türevlenebilir olduğundan, her iki ağı da doğru yöne yönlendirmek için kullanabileceğimiz bir gradyan elde ederiz. Tipik olarak, üretici ana ilgi konusudur- ayırıcı, üretici eğitildikten sonra atılan uyarlanabilir bir kayıp işlevidir (24).

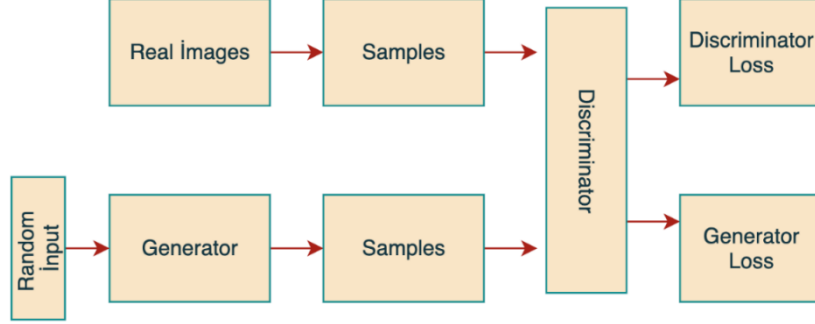
#### 3.4.2. DR-Skoru: Hem Ayırıcılık hem de Yeniden Yapılanma Kullanılarak Anomali Tespiti

GAN kullanmanın bir avantajı, aynı anda eğitilmiş bir ayırıcı ve bir jeneratöre sahip olmamızdır. Anomalileri tanımlamak için normal anatomik değişkenliği temsil etmek üzere ortaklaşa eğitilmiş hem ayırıcı hem de

oluşturucudan yararlanmayı öneriyoruz. (12)'deki formülasyonu takiben, GAN tabanlı anomali tespiti aşağıdaki iki bölümden oluşur.

### Ayrımcılığa Dayalı Anomali Tespiti

Eğitilmiş ayrımcı D'nin sahte verileri (yani anomalileri) gerçek verilerden yüksek hassasiyetle ayırt edebildiği göz önüne alındığında, anomali tespiti için doğrudan bir araç olarak hizmet eder.



**Şekil 7.** Denetimsiz GAN tabanlı anomali algılama. Solda, üretici ve ayrımcının yinelemeli çekişmeli eğitimle elde edildiği bir GAN çerçevesi var. Sağda, hem GAN tarafından eğitilmiş ayrımcının hem de oluşturucunun, ayırım ve yeniden yapılandırmaya dayalı birleşik bir anomali skoru hesaplamak için uygulandığı anomali algılama süreci yer almaktadır.

### Yeniden Yapılandırmaya Dayalı Anomali Tespiti

Gerçekçi örnekler üretebilen eğitilmiş jeneratör  $G$  aslında gizli alandan gerçek veri alanına bir eşlemedir:  $G(Z): Z \rightarrow X$  ve normal verilerin dağılımını yansıtan açık olmayan bir sistem modeli olarak görülebilir. [24]'te bahsedilen gizli uzayın yumuşak geçişlerinden dolayı, gizli uzaydaki girişler yakınsa, jeneratör benzer örnekleri çıkarır. Bu nedenle, test verileri için gizli uzayda karşılık gelen  $Z_k$ 'yi bulmak mümkünse  $X_{tes}$  ve  $G(Z^k)$  (yeniden yapılandırılmış test örnekleri olan) arasındaki benzerlik,  $X_{tes}$ 'in dağılımı ne ölçüde takip ettiğini açıklayabilir.  $G$  tarafından yansıtılır. Başka bir deyişle, test verilerindeki anomalileri belirlemek için  $X_{tes}$  ve  $G(Z^k)$  arasındaki artıkları da kullanabiliriz.

Test verilerine karşılık gelen optimal  $Z^1$ 'yi bulmak için, önce gizli uzaydan rastgele bir  $Z^1$  kümesini numuneleriz ve jeneratöre besleyerek yeniden yapılandırılmış ham numuneler  $G(Z^1)$  elde ederiz. Daha sonra  $X_{tes}$  ve  $G(Z)$  ile tanımlanan hata fonksiyonundan elde edilen gradyanlar ile latent uzaydan örnekleri güncelleriz.  $\min_{Z^k} Er(X_{tes}, G_{rnn}(Z^k)) = 1 - Simi(X_{tes}, G_{rnn}(Z^k))$  burada diziler arasındaki benzerlik basitlik için kovaryans olarak tanımlanabilir.

Hata yeterince küçük olacak şekilde yeterli yineleme turlarından sonra,  $Z$  k örnekleri, test örnekleri için gizli uzayda karşılık gelen eşleme olarak kaydedilir. Test numuneleri için  $t$  zamanındaki artık  $Res(X_t^{tes}) = \sum_{i=1}^n |x_t^{tes,i} - G_{rnn}(Z_t^{k,i})|$  olarak hesaplanır.

## 3.5. CPS ve Siber saldırılar

### 3.5.1. Su Arıtma ve Dağıtım Sistemi

#### SWaT

Güvenli Su Arıtma (SWaT) sistemi, büyük şehirlerde bulunan büyük bir modern su arıtma tesisinin küçük ölçekli bir versiyonunu temsil eden su arıtma için operasyonel bir test yatağıdır [25]. Genel test yatağı tasarımı, genel fiziksel süreç ve kontrol sisteminin sahadaki gerçek sistemlere çok benzemesini sağlamak için ülke çapında su hizmetleri şirketi olan Singapur Kamu Hizmet Kurulu ile koordine edildi. SWaT veri seti toplama süreci, sistem günde 24 saat çalıştırılarak 11 gün sürmüştür. 2016 SWaT veri toplama sürecinin son 4 gününde toplam 36 saldırı başlatıldı [25]. Genellikle saldırıya uğrayan noktalar arasında sensörler (ör. su seviyesi sensörleri, akış hızı ölçer vb.) ve aktüatörler (ör. vana, pompa vb.) bulunur. Bu saldırılar, son dört gün içinde farklı amaçlarla ve çeşitli kalıcı sürelerle (birkaç dakikadan bir saate kadar) test alanında başlatıldı. Ya başka bir saldırı başlatılmadan önce sistemin normal çalışma durumuna gelmesine izin verildi ya da saldırılar art arda başlatıldı.

SWaT 'deki su arıtma işlemi, P1'den P6'ya [26] olarak adlandırılan altı alt süreçten oluşur. İlk süreç ham su temini ve depolaması içindir ve P2, su kalitesinin değerlendirildiği ön arıtma içindir. İstenmeyen materyaller, P3'te ultra filtrasyon (UF) geri yıkaması ile uzaklaştırılır. Kalan korin, Deklorinasyon işleminde (P4) yok edilir. Ardından, inorganik safsızlıkları azaltmak için P4'ten gelen su Ters Ozmos (RO) sistemine (P5) pompalanır. Son olarak, P6 suyu dağıtımına hazır olarak depolar.

## WADI

Tipik olarak güvenli bir yerde bulunan bir su arıtma sistemi tesisinin aksine, bir dağıtım sistemi geniş bir alana yayılan çok sayıda boru hattından oluşur. Bu, bir dağıtım ağına fiziksel saldırı riskini oldukça artırır. Su Dağıtım (WADI) test ortamı, eksiksiz ve gerçekçi bir su arıtma, depolama ve dağıtım ağı oluşturmak için SWaT'lerin ters ozmos permeatını ve ham suyu alarak SWaT sisteminin bir uzantısıdır. Su dağıtım sisteminde üç kontrol süreci vardır. İlk işlem, ham suyun SWaT, Public Utility Board (PUB) girişinden veya WADI'deki dönüş suyundan alınması ve ham suyun iki tankta depolanmasıdır. P2, suyu iki yükseltilmiş rezervuar tankından ve altı tüketici tankından önceden belirlenmiş bir talep modeline göre dağıtır. Su geri dönüştürülür ve üçüncü işlemde P1'e geri gönderilir.

WADI test ortamı benzer şekilde kimyasal dozlama sistemleri, takviye pompaları ve valfleri, enstrümantasyon ve analizörlerle donatılmıştır (27). WADI, ağlar aracılığıyla PLC'lerde gerçekleştirilen saldırıları ve savunmaları simüle etmenin yanı sıra, su sızıntısı ve kötü niyetli kimyasal enjeksiyonlar gibi fiziksel saldırıların etkilerini simüle etme yeteneklerine sahiptir. WADI veri toplama süreci, 14 günü normal operasyonda ve 2 günü saldırı senaryolarında olmak üzere 16 günlük sürekli operasyonlardan oluşmaktadır. Veri toplama sırasında tüm ağ trafiği, sensör ve aktüatör verileri toplanmıştır. WADI veri seti hakkında daha fazla ayrıntı için lütfen WADI web sitesine bakın.

### 3.5.2. Siber Saldırıları

Bir saldırganın amacı, tesisin normal operasyonlarını manipüle etmektir. Saldırganın SWaT ve WADI'nin SCADA sistemine uzaktan erişimi olduğu ve sistemlerin nasıl çalıştığı hakkında genel bilgiye sahip olduğu varsayılmaktadır.

İlgili sistem yanıtlarını araştırmak için SWaT ve WADI sistemlerinde çeşitli deneyler yapılmıştır. Toplamda SWaT ve WADI'ye sırasıyla 36 saldırı ve 15 saldırı eklendi [25]. Örnek olarak, her sistem için bir örnek saldırı açıklayalım.

**SWaT** Bir saldırı hedefi, SWaT performansını nominal seviyeden düşürmektir (örneğin, P4'teki Reverse Osmosis (RO) besleme tankının su seviyesi 5 galon. Saldırgan, LIT401'e saldırarak, RO besleme tankının seviyesini düşürdü. 800mm'den 200mm'ye kadar, bu PLC-4'ün pompayı durdurmasına yol açacaktır P401 ve daha az su P5'e pompalandı. Son olarak, LIT401'e saldıran sensörün olumsuz etkisi, RO ünitesinin çıkış su akış hızına yansıdı (değerler FIT501 ile ölçülen P5). Sistem özelliklerine göre bu debi yaklaşık 1,2cm kalmalıdır.

**WADI** Bir saldırı hedefi, P1'deki su seviyesi sensörünün okumalarını manipüle etmektedir. Saldırgan, sensör değerini tank kapasitesinin %76'sından %10'una değiştirerek "düşük durum" gösterir. Sonuç olarak, PLC-1 (P1 kontrolörü) WADI dönüşünden, SWaT çıkışından veya PUB girişinden daha fazla su almak için giriş suyu pompasını açmak için bir komut gönderir. Aynı zamanda, ham su deposundaki hatalı düşük su seviyesi durumu nedeniyle, P1'den P2'ye su beslemesi kesilirken, P2 tüketici tanklarına su sağlamaya devam eder. Böylece P2'deki tankların su seviyeleri azaldı. Yükseltilmiş tanklardaki (P2) su seviyesi düşük bir seviyeye ulaştığında, tüketici tanklarına (P2) verilen besleme kesildi. Sonuç olarak, P1'deki su seviye sensörünün okumalarını düşük bir seviyeye değiştirerek, P1'in tanklarında bir taşma olacak ve P2'de su akışı olmayacaktır.

**Tablo 1.** Veri kümeleri hakkında genel bilgiler.

	Swat	WADI	KDDCUP99
<b>Değişken</b>	51	103	34
<b>Saldırı</b>	36	15	2
<b>Saldırı Süresi</b>	2 – 25	1.5 -30	NA
<b>Training Boyutu</b>	496800	1048571	562387
<b>Test Boyutu</b>	449919	172801	494021
<b>N_Rate(%)</b>	88.02	94.01	19.69

(1) N oranı, normal veri noktalarının test veri kümelerindeki tümüne oranıdır.

### 3.5.3. SwaT ve WADI Veri Kümeleri

SWaT/WADI veri toplama süreci, sistemlerin günde 24 saat çalıştırılmasıyla 11/16 gün sürmüştür. Son 4/2 günde, test yataklarına farklı amaçlarla ve farklı süreli sürelerle (birkaç dakikadan bir saate kadar) çeşitli siber saldırılar gerçekleştirildi. Sistemlerin ya başka bir saldırı başlatılmadan önce normal çalışma durumuna gelmesine izin verildi ya da saldırılar art arda başlatıldı. Bu iki veri kümesi hakkında bazı genel bilgiler Tablo 1'de özetlenmiştir. Tespit görevlerinin karmaşıklığını daha iyi anlamak için, iki veri kümesinin ve bunlarla ilişkili normal koşulların ve saldırıların aşağıdaki özelliklere sahip olduğunu belirtmekte fayda var.



Farklı senaryolar nedeniyle farklı saldırılar farklı sürelerde sürebilir. Bazı saldırılar hemen etkili olmadı. Sistem istikrar süreleri de saldırılara göre değişir. Değişen akış hızlarını hedefleyen saldırılar gibi daha basit saldırılar, sistemin dengelenmesi için daha az zaman gerektirirken, sistem dinamikleri üzerinde daha güçlü etkilere neden olan saldırılar, istikrar için daha fazla zamana ihtiyaç duyacaktır.

Bir sensöre/aktüatöre yapılan saldırılar, diğer sensörler/aktüatörler veya tüm sistem üzerindeki performansı, genellikle belirli bir zaman gecikmesinden sonra. Ayrıca, benzer türdeki sensörler/eyleyiciler, saldırılara benzer şekilde yanıt verme eğilimindedir. Örneğin, LIT101 sensörüne (SWaT 'nin P1'indeki bir su seviyesi sensörü) saldırılar, hem LIT101'de hem de LIT301'de (SWaT 'nin P3'ünde başka bir su seviyesi sensörü) bariz anomali artışlara neden oldu, ancak diğer sensörlerin ve aktüatörlerin okumaları üzerindeki etkiler nispeten daha küçüktü. Yukarıda bahsedilen gözlemler, farklı alt süreçlerin performanstaki genel değişimi toplu olarak saldırıları daha iyi tanımayaya yardımcı olabileceğinden, anomali tespiti için sistemlerin modellenmesinde çok değişkenli bir yaklaşımın benimsenmesinin önemli olduğunu göstermektedir. Başka bir deyişle, sensörler ve aktüatörler arasındaki temel korelasyonlar, saldırının neden olduğu sistem davranışlarındaki anomalileri tespit etmek için faydalı olabilir.

## 4. KONFIGÜRASYON VE PERFORMANS METRİKLERİ

### 4.1. Veri Hazırlama ve Sistem Mimarisi

SWaT veri setinde, 11 gün boyunca 51 değişken (sensör okumaları ve aktüatör durumları) ölçülmüştür. Ham veriler içerisinde normal çalışma koşullarında (ilk 7 günde toplanan veriler) 496.800 örnek, sonradan sisteme çeşitli siber saldırılar yapıldığında ise 449.919 örnek toplanmıştır. Benzer şekilde WADI veri seti için normal çalışma koşullarında ilk 14 günde 103 değişken için 789371 örnek, son 2 gün içinde sisteme çeşitli siber saldırılar yapıldığında 172801 örnek toplanmıştır. Bu veri setlerinin her ikisi için, sistem [26]'ya göre ilk açıldığında stabilizasyona ulaşması 5-6 saat sürdüğü için eğitim verilerinden (normal veriler) ilk 21.600 örneği çıkarılmıştır. Anomali algılama sürecinde, ham akışlar boyunca kayan bir pencere olarak orijinal uzun çoklu dizileri daha küçük zaman serilerine bölünür. Alt dizi gösterimi için optimal pencere uzunluğuna karar vermek zaman serisi çalışmasında önemli bir konu olduğundan, sistem durumunu farklı çözünürlüklerde yakalamak için bir dizi farklı pencere boyutu denedik, yani  $sw = 30 \times i$ ,  $i = 1, 2, \dots, 10$ . SWaT verilerinin ilgili dinamiklerini yakalamak için, pencere, kayma uzunluğu  $ss=10$  olan normal ve test veri kümelerine uygulanır.

Bu çalışma için, jeneratör için derinlik 3 ve 100 gizli (dahili) üniteye sahip bir LSTM ağı kullandık. Diskriminatör için LSTM ağı, 100 gizli birim ve derinlik 1 ile nispeten daha basittir. [9]'deki gizli uzay boyutu hakkındaki tartışmadan esinlenerek, farklı boyutlar da denedik ve özellikle çok değişkenli üretirken daha yüksek gizli uzay boyutunun genellikle daha iyi örnekler ürettiğini bulduk. Bu çalışmada latent uzayın boyutunu 15 olarak belirledik.

### 4.2. Değerlendirme Metrikleri

GAN'ın anomali tespit performansını değerlendirmek için Precision (Pre), Recall (Rec) ve F1 puanları gibi standart metrikleri kullanırız:

$$Rec=(TP)/(TP+FN) \quad (1)$$

$$Pre=(TP)/(TP+FP) \quad (2)$$

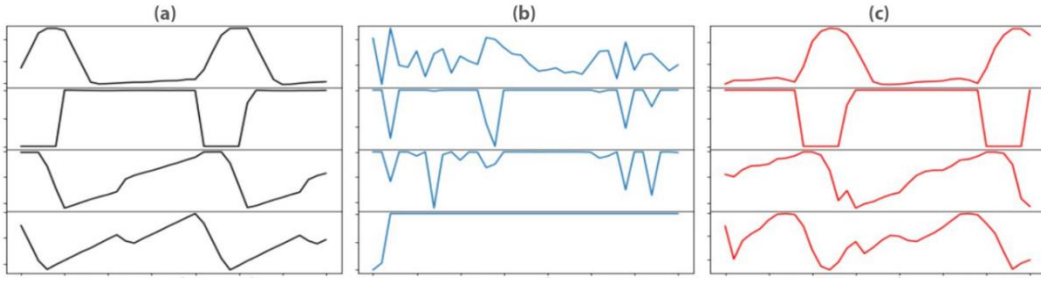
$$F1=2 \times (Pre \times Rec)/(Pre+Rec) \quad (3)$$

Bu çalışmadaki uygulamamız izinsiz girişleri (siber saldırı) tespit etmek olduğundan, sistemin birkaç yanlış alarmı tolere etmeyi gerektirse bile tüm saldırıları tespit etmesi önemlidir. Bu nedenle, gerçek pozitifleri doğru bir şekilde tespit etmeye çalışırken, yanlış pozitifler aşırı olmadığı sürece önemli değildir. Bu nedenle, bu çalışma için anomali tespiti performansını ölçmek için ana metrik olarak hatırlama kullanılır.

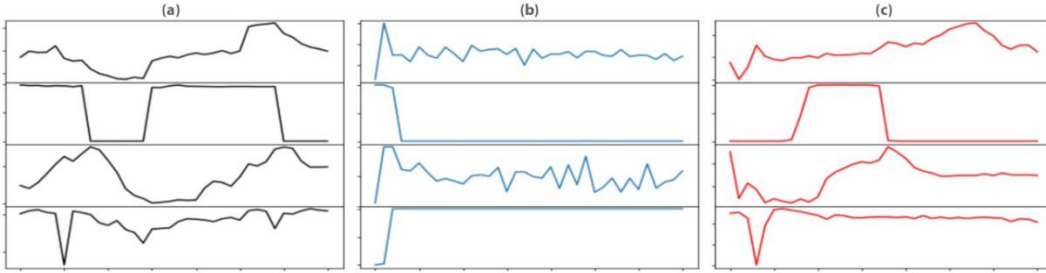
Bu çalışma gerçekleştirilen uygulama izinsiz girişleri (siber saldırı) tespit etmek olduğundan, sistemin birkaç yanlış alarmı tolere etmeyi gerektirse bile tüm saldırıları tespit etmesi önemlidir. Bu nedenle, gerçek pozitifleri doğru bir şekilde tespit etmeye çalışırken, yanlış pozitifler aşırı olmadığı sürece önemli değildir. Bu nedenle, bu çalışma için anomali tespiti performansını ölçmek için ana metrik olarak hatırlamayı kullanıyoruz.

## 5. SONUÇ, TARTIŞMA VE ÖNERİLER

GAN'ın anomali tespit performansını yukarıda bahsedilen iki veri seti SWaT ve WADI üzerinde değerlendiriyoruz. Daha önce açıklandığı gibi, alt diziler MAD-GAN modeline beslenir. Hesaplama yükünü azaltmak için, PC varyans oranına göre PC boyutunu seçerek orijinal boyutu PCA ile boyut indirgeme yapılmıştır. Anomali algılama performansı üzerinde karşılaştırma yapmak için, veri setlerinde popüler denetimsiz anomali algılama yöntemleri olan PCA, K-En Yakın Komşu (KNN), Özellik Paketleme (FB) ve Otomatik Kodlayıcı (AE) da uyguladık. GAN tabanlı bir yöntemle karşılaştırmak için, GAN'ı, ayırıcısı ve üretici tam bağlı sinir ağları olarak uygulanan [12]'nin Efficient GAN tabanlı (EGAN) yöntemiyle karşılaştırılmıştır.

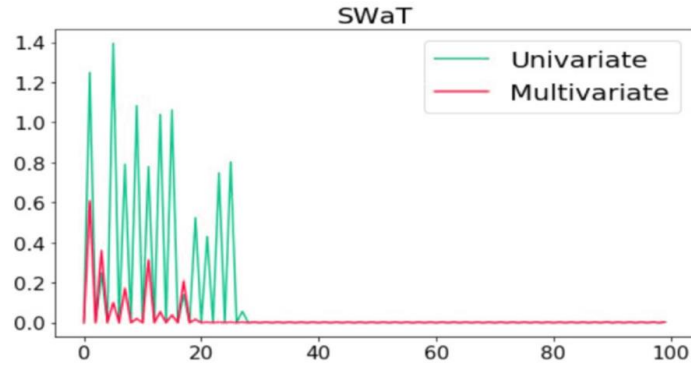


Şekil 8. Farklı eğitim aşamalarında oluşturulan örnekler arasında karşılaştırma: Swat.

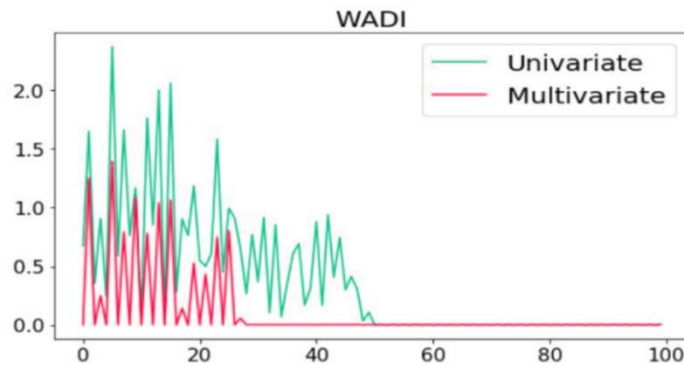


Şekil 9. Farklı eğitim aşamalarında oluşturulan örnekler arasında karşılaştırma: Wadi.

Şekil 4-1'te her iki veri kümesi için çok değişkenli örnekler oluşturmak için GAN eğitim yinelenmelerinde MMD değerleri çizilmektedir. Her iki veri kümesi için 30-50 yinedemeden sonra MMD değerlerinin küçük değerlere yaklaşma eğiliminde olduğunu gözlemleyebiliriz. Ayrıca tek değişkenli numune üretimi için MMD değerleri karşılaştırılmıştır. Çok değişkenli örneklerin erken MMD değerleri tek değişkenli örneklerinkinden daha düşük olması ve çok değişkenli örnekler için MMD de tek değişkenli durumdan daha hızlı yakınsamıştır. Bu, birden fazla veri akışının kullanılmasının GAN modelinin eğitimine yardımcı olabileceğini düşündürmektedir.



Şekil 10. MMD: Çoklu zaman serisi.



Şekil 11. MMD: Tek zaman serisi.

### 5.1. Anomali Tespit Performansı

Tablo 2'de, popüler denetimsiz yöntemlerle (PCA, KNN, FB ve AE) en iyi performansı altı çizili olarak ve genel olarak en iyi performansı koyu renkle gösteriyoruz.

F1 denge hassasiyeti ve geri çağırma bu yana en iyi F1 tarafından seçilen sonuçlara odaklanan SWaT veri seti için ADGAN, AE tarafından kesinlik ve geri çağırma için sırasıyla %26.34 ve %11.11 ile verilen dört popüler yöntemle en iyi performansı geride bıraktı. Aslında, GAN, SWaT için tüm anonim noktaları yanlış alarmlar olmadan doğru bir şekilde algılayarak, burada neredeyse %100 hassasiyet ve geri çağırma elde etti.

Aynı zamanda en iyi F\_1 tarafından seçilen sonuçlara odaklanan WADI veri seti için, GAN tarafından yapılan geri çağırma, AE'ninkinden biraz daha zayıftır (%3.02 daha düşük). Ancak, en iyi hatırlama durumu için, GAN, geri çağırma değerlerine göre %65,64 %94,36 ile diğerlerinden daha iyi performans gösterdi. GAN'ın hassasiyeti zayıf görünse de, %100'e yakın bir geri çağırma değerine ulaşabilir. Yanlış pozitifin maliyeti tüm izinsiz girişleri tespit etmek için tolere edilebilir olduğundan (Bölüm 5.2'de bahsedildiği gibi) bu siber saldırı ortamında kabul edilebilir. Karşılaştırıldığında, popüler tespit yöntemlerinin hiçbiri tatmin edici bir geri çağırma sağlayamaz.

İki veri kümesi arasında GAN, SWaT için belirgin şekilde daha iyi performans gösterdi. Tablo 1'deki "N oranı" ile gösterildiği gibi, WADI veri seti SWaT'den daha dengesizdir (yani daha fazla gerçek negatif), bu da daha fazla yanlış bildirilen pozitiflere yol açar. Ayrıca, Tablo 1'de gösterildiği gibi, WADI veri kümesinin SWaT'den daha büyük bir özellik boyutuna sahip olduğunu da not ediyoruz (WADI 103 değişkene sahipken SWaT yalnızca 51 değişkene sahiptir.).

**Tablo 2.** Farklı veri kümeleri için anomali tespit metrikleri.

Datasets	Metod	Precesision	Recall	F1
Swat	PCA	24.92	21.65	0.23
	KNN	7.83	7.85	0.08
	FB	10.17	10.17	0.1
	AE	72.63	52.63	0.61
	EGAN	40.57	67.73	0.51
	GAN*	99.99	54.8	0.7
	GAN**	12.2	99.98	0.22
	GAN***	98.97	63.74	0.77
WADI	PCA	39.53	5.63	0.1
	KNN	7.76	7.75	0.08
	FB	8.6	8.6	0.09
	AE	34.35	34.35	0.34
	EGAN	11.33	37.84	0.17
	GAN*	46.98	24.58	0.32
	GAN**	6.46	99.99	0.12
	GAN***	41.44	33.92	0.37
GAN*	En iyi Precision tarafından seçilen sonuçlar listelenir.			
GAN**	En iyi Recall tarafından seçilen sonuçlar listelenir.			
GAN***	En iyi F1 tarafından seçilen sonuçlar listelenir.			

Burada ayrıca GAN'ı daha dengeli bir veri kümesine, KDDCUP99 veri kümesine uyguladık. Bu veri setinde GAN, %85'in üzerinde hassasiyet ve %94'ün üzerinde geri çağırma ile 0.90 F1 puanına ulaşabilir. KDDCUP99 veri setindeki ([12] tarafından rapor edilen) EGAN sonuçları MAD-GAN'ınkinden daha iyi olsa da, GAN hem SWaT hem de WADI veri setleri (dengesiz veri setleri) için EGAN'dan daha iyi performans gösterdi. Bunun nedeni, GAN'da kullanılan LSTM-RNN'nin, EGAN'da kullanılan CNN'lerden daha iyi karmaşık zaman serilerini öğrenebilmesidir. Aslında, EGAN'ın diğer GAN olmayan yöntemlerle görece performansına baktığımızda, zamansal korelasyonu uygun şekilde modellemesek GAN tabanlı anomali tespitinin diğer geleneksel yöntemlerle rekabet edemediğini görebiliriz.

Genel olarak, GAN, popüler denetimsiz algılama yöntemlerinden tutarlı bir şekilde daha iyi performans gösterdi. Tek dezavantajı, LSTM-RNN'nin daha uzun alt dizilerle uğraşmasının daha fazla zaman almasıdır (spesifik olmak gerekirse, alt dizi uzunluğu sw 200'den büyük olduğunda model yavaşlar). Zamansal korelasyonu dahil etmek için diğer Sinir Ağlarını kullanmayı keşfetmek ve gelecekteki çalışmalar için alt dizi uzunluğu seçimini düşünmek faydalı olacaktır.

Ağa bağlı sensörler ve aktüatörlerle donatılmış günümüzün siber-fiziksel sistemleri, siber saldırıların neden olduğu anomalileri tespit etmek için sistem davranışlarını izlemek için kullanılacak büyük miktarda veri akışı üretir. Bu yazıda, EKS'ler tarafından oluşturulan zaman serisi verilerinde çok değişkenli anomali tespiti için GAN'ın kullanımını araştırdık. LSTM-RNN'leri çok değişkenli zaman serisi verileri üzerinde eğitmek için yeni bir GAN (GAN ile Çok Değişkenli Anomali Tespiti) çerçevesi önerdik ve ardından yeni bir Ayrıcılık ve Yeniden Yapılandırma Anomali Puanı (DR-) kullanarak anomalileri tespit etmek için hem ayırıcılığı hem de oluşturucuyu

kullandık. Puan). GAN'ı Secure Water Treatment Testbed (SWaT) ve Water'dan alınan iki karmaşık siber saldırı CPS veri kümesi üzerinde test edilmiştir.

Dağıtım Sistemi (WADI) ve GAN tabanlı bir yaklaşım da dahil olmak üzere mevcut denetimsiz algılama yöntemlerine göre üstün performans göstermiştir. Dağıtım Sistemi (WADI) ve mevcut denetimsiz algılama yöntemlerine göre üstün performans göstermiştir. Bunun GAN kullanılarak zaman serisi verilerinde çok değişkenli anomali tespiti için erken bir girişim olduğu göz önüne alındığında, daha fazla araştırmayı bekleyen ilginç sorunlar var. Gelecekteki çalışmalar için, çok değişkenli anomali tespiti için özellik seçimi konusunda daha fazla araştırma yapmayı ve teorik garantilerle gizli boyut ve PC boyutunu seçmek için ilkeli yöntemleri araştırmaları yapılabilir. Ayrıca algılama modelinin kararlılığı hakkında ayrıntılı bir çalışma yapmayı umuyoruz. Uygulamalar açısından, akıllı binalar ve makineler için kestirimci bakım ve arıza teşhisi gibi diğer anomali algılama uygulamaları için GAN kullanımını araştırmayı planlıyoruz.

Derin Öğrenme yaklaşımları, SCADA sistemlerinde anomali tespiti için giderek daha fazla kullanılmaktadır. Yüksek anomali algılama oranı sağlamak için mevcut SCADA ağı büyük verilerinden önemli özelliklerin öğrenilmesini mümkün kılan denetimsiz özellik öğrenme yeteneği, derin öğrenme yaklaşımlarına artan ilgiye katkıda bulunur. SCADA veri özelliklerini öğrenmek için CNN, LSTM, DBN, SAE, SDAE veya bunların bir kombinasyonu gibi çoklu mimariler ve Softmax katmanı, Tam bağlı sinir ağı gibi sınıflandırıcılar; Sınıflandırma için ELM, DAE veya MLP kullanılır. Çoğu durumda, derin öğrenme yaklaşımları standart yaklaşımlardan daha iyi performans gösterir, ancak Aşıl topuğu, eğitim(training) için gereken yüksek eğitim süresi bir dezavantaj olarak devam eder. Bilimsel topluluk tarafından yüksek eğitim süresi eksikliğinin üstesinden gelmek için alınan ilginç araştırma yönü, Endüstriyel Kontrol Sistemlerinde anomali tespiti için dağıtılmış derin öğrenme yaklaşımlarının kullanılmasıdır. Gelecekteki bir çalışmada, Endüstriyel Kontrol Sistemlerinde anomali tespiti için dağıtılmış bir derin öğrenme yaklaşımı önerilebilir. Çoğu durumda, derin öğrenme yaklaşımları standart yaklaşımlardan daha iyi performans gösterir.

## Yazar Katkıları

Yazarlar makaleye eşit derecede katkı sağladılar

## Çıkar Çatışması

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler

## KAYNAKÇA

- [1] K.A. Stouffer, J.A. Falco, and K.A. Scarfone "Guide to Industrial Control Systems (ICS) Security: Supervisory Control and Data Acquisition (SCADA) Systems, Distributed Control Systems (DCS), and Other Control System Configurations Such As Programmable Logic Controllers (PLC)", Gaithersburg, MD, United States: NIST Special Publication vol.82 (800), 2014.
- [2] Y. Zhang, L.Wang, W. Sun, R.C. Green II and M. Alam "Distributed Intrusion Detection System in a Multi-Layer Network Architecture of Smart Grids", IEEE Transactions on Smart Grid. vol. 2, pp. 796-808, 2011
- [3] F. Pasqualetti, F. Dörfler, F. Bullo "Cyber-physical attacks in power networks: Models, fundamental limitations and monitor design", IEEE Conference on Decision and Control and European Control Conference, 2011.
- [4] J. M. Beaver, R. Borges, M. Buckner "An Evaluation of Machine Learning Methods to Detect Malicious SCADA Communications", 12th International Conference on Machine Learning and Applications, 2013.
- [5] V. Chandolai, V. Mithal, V. Kumar "Comparative evaluation of anomaly detection techniques for sequence data", In Eighth IEEE International Conference on Data Mining, pp. 743-748. 2008.
- [6] B. Sun, P.B. Luh, Q.-S. Jia, Z.O. Neill, F. Song. "Building energy doctors: An spc and kalman filter-based method for system-level fault detection in hvac systems", IEEE Transactions on Automation Science and Engineering, vol. 11, pp. 215-229, 2014.
- [7] K. Donghwoon, H. Kim, J. Kim, S.C. Suh, I. Kim, K. J. Kim "A survey of deep learning-based network anomaly", Cluster Comp., vol. 22, pp. 1-139, 2017.
- [8] O. Mogren "C-rnn-gan: Continuous recurrent neural networks with adversarial training", arxiv:1611.09904, 2016.
- [9] E. Cristbal, S.L. Hyland, and G. Rtsch "Real-valued (medical) time series generation with recurrent conditional gans", arXiv:1706.02633, 2017.
- [10] X. Yuan, T. Xu, H. Zhang, R. Long, and X. Huang "Segan: Adversarial network with multi-scale l1 loss for medical image segmentation", Neuroinform, vol. 16, pp. 383-392, 2018.
- [11] S. Tim, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen "Improved techniques for training gans", In Advances in Neural Information Processing Systems, arXiv:1606.03498, 2016.
- [12] S. Thomas, P. Seebck, S.M. Waldstein, U. Schmidt-Erfurth, G. Langs "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery", Lecture Notes in Computer Science, vol. 10265, pp. 146-157, 2017.
- [13] Z. Houssam, C.S. Foo, B. Lecouat, G. Manek, V.R. Chandrasekhar "Efficient gan-based anomaly detection", arXiv:1802.06222, 2018.

- [14] S. Li and J. Wen “A model-based fault detection and diagnostic methodology based on pca method and wavelet transform”, *Energy and Buildings*, vol. 68, pp. 63–71, 2014.
- [15] S. Wol, E. Kim, P. Geladi “Principal component analysis”, *Chemometrics and intelligent laboratory systems*, vol. 2, pp. 37–52, 1987.
- [16] S. Kotz and N.L. Johnson “Partial least squares”, In *Encyclopedia of Statistical Sciences*, vol. 6, pp. 581–591, 1985.
- [17] D. Xuewu and Z. Gao “From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis”, *IEEE Transactions on Industrial Informatics*, vol. 9, pp. 2226–2238, 2013.
- [18] M.R. Breuni, P. Kröger, R.T. Ng, J. Sander “Lof: identifying density-based local outlier”, *ACM SIGMOD Record*, vol. 29, no. 2, pp. 93–104, 2000.
- [19] M. Schuber, H.P. Kriegel and A. Zimek “Angle-based outlier detection in high-dimensional data”, *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 444–452, 2008.
- [20] L. Aleksandar and V. Kumar. “Feature bagging for outlier detection”, *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pp. 157–166, 2005.
- [21] B. Zong, Q. Song, M.R. Min, W. Cheng, C. Lumezanu, D. Cho, H. Chen “Deep autoencoding gaussian mixture model for unsupervised anomaly detection”, *ICLR 2018 Conference Blind Submission*, 2018.
- [22] H. Edan and A. Shabtai. “Using lstm encoder-decoder algorithm for detecting anomalous ads-b messages”, *Computers and Security*, vol. 78, 2018.
- [23] T. Karras, T. Aila, S. Laine, J. Lehtinen “Progressive Growing Of Gans For Improved Quality, Stability, and Variation”, *ICLR*, pp. 1, 2018.
- [24] A. Mathur N.O. Tippenhauer “Swat: A water treatment testbed for research and training on ics security”, *2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater)*, pp. 31–36, 2016.
- [25] G. Jonathan, S. Adepun, K.N. Junejo, A. Mathur. “A dataset to support research in the design of secure water treatment systems”, *International Conference on Critical Information Infrastructures Security*, vol. 10242, 2017.
- [26] C.M. Ahmed, V.R. Palletti and A.P. Mathur. “Wadi: A water distribution testbed for research in the design of secure cyber physical systems”, In *Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks*, pp. 25–28, 2017.
- [27] Y. Raymond, C. Chen, T.Y. Lim, M. Hasegawa-Johnson, and M. N. Do. “Semantic image inpainting with perceptual and contextual losses”, *arXiv:1607.07539*, vol. 1607, 2016.
- [28] S. Tim, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen “Improved techniques for training gans in *In Advances in Neural Information Processing Systems*”, Part of *Advances in Neural Information Processing Systems* vol. 29, pp. 2234–2242. 2016.