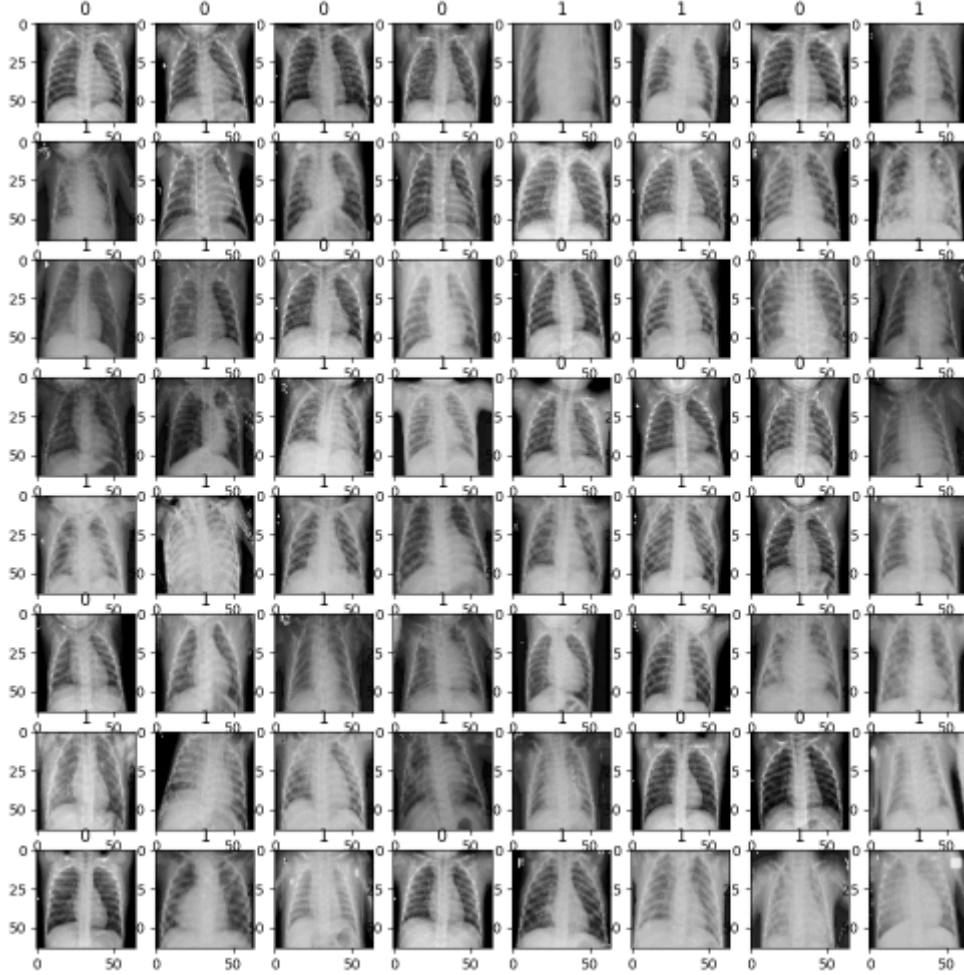


## SAĞLIKTA YAPAY ZEKA - SENTETİK VERİ ÜRETİMİ

### Veri Seti

Bu çalışmada Göğüs Röntgeni görüntüleri kullanılmıştır ([Veri Seti](#)). Veri setinde 3 alt başlık bulunmaktadır. Test, Train, Validation. Her alt başlık 2 kısımdan oluşmaktadır. NORMAL (sağlıklı röntgen görüntüleri), PNEUMONIA (zatürre röntgen görüntüleri). Modellerimizi geliştirirken eğitim için Train alt başlığı kullanılacaktır. Train alt başlığında 1341 normal , 3875 zatürre olmak üzere toplam 5216 göğüs röntgeni görüntüsü bulunmaktadır.



Şekil 1: Veri Seti

### Veri Ön İşleme

Veri ön işleme aşamalarında performans için:

- RGB kanalları tek boyutlu Siyah-Beyaz kanala indirgenmiştir.
- Resim boyutu 64x64 piksel olarak yeniden düzenlenmiştir.
- Resimlere 0 (normal) , 1 (zatürre) olacak şekilde atamalar yapıp bu atamaları temsil edecek dizin oluşturulmuştur.

## Generatif Modeller

### 1. Variational Auto Encoder (VAE)

1.1. **Temel Yapısı:** VAE, geleneksel autoencoder mimarisiyle benzer şekilde iki ana bileşenden oluşur: encoder (kodlayıcı) ve decoder (kod çözücü).

1.1.1. **Encoder (Kodlayıcı):** Encoder, girdiyi (örneğin bir resim veya vektör) daha düşük boyutlu bir latent uzaya (gizli uzay) projeler. Ancak VAE'de bu latent uzaya sabit bir nokta yerine bir olay dağılımı üretilir. Yani, encoder, ortalama ( $\mu$ ) ve standart sapma ( $\sigma$ ) gibi parametreleri tahmin ederek gizli değişkenlerin normal dağılımını modellemeye çalışır.

1.1.2. **Decoder (Kod Çözücü):** Decoder, bu latent uzayda üretilen rastgele bir örneği (sampling), orijinal veri uzayına geri çevirir. Bu süreç, yeni

### 2. Generative Adversarial Networks (GAN)

2.1. **Temel Yapısı:** GAN'lar, iki ayrı sinir ağı olan üretici (generator) ve ayırmacı (discriminator) ağlarının birbirlerine karşı rekabet ettikleri bir yapıdan oluşan bir derin öğrenme modelidir. Bu rekabet, modelin karmaşık ve gerçekçi veri örnekleri üretmesini sağlar.

2.1.1. **Generator (Üretici):** Amacı, gerçek verilere benzeyen sahte veri örnekleri üretmektir. Genellikle rastgele gürültüden (örneğin bir normal dağılımdan elde edilen rastgele vektör) başlar ve bu gürültüyü gerçek verilere benzer örneklere dönüştürmeye çalışır. Üretilen veriler, görüntü, ses, metin gibi çeşitli formatlarda olabilir.

2.1.2. **Discriminator (Ayırmacı):** Amacı, gerçek veriler ile generator tarafından üretilen sahte verileri ayırt etmektir. Bu ağ, girdiyi alır ve bunun gerçek mi yoksa sahte mi olduğunu sınıflandırır. Discriminator, gerçek veriyi ve generator tarafından üretilen sahte veriyi görür ve bu iki sınıf arasında ayırım yapmaya çalışır.