

ACNE SEVERITY CLASSIFICATION WITH VISION TRANSFORMER (ViT)

S E N A Y I L D I Z

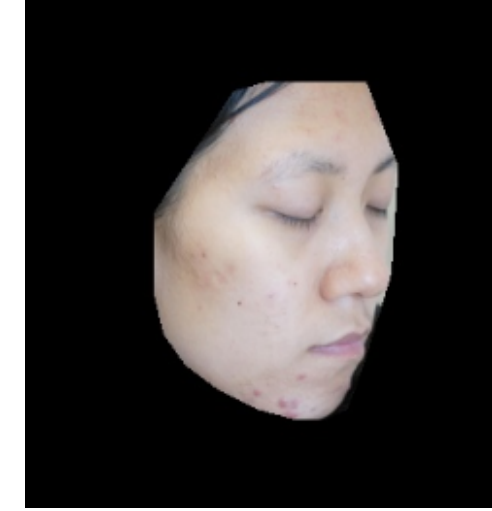
DATASET TANIMI

Kaggle Acne Grading Classification Dataset

Veri seti, RGB frontal yüz görüntülerinden oluşan ve akne şiddetini Level_0, Level_1, Level_2 olmak üzere üç sınıfta ele alan çok sınıflı bir görüntü sınıflandırma problemidir. Toplam 999 görüntü bulunmaktadır. Level_2 sınıfının verinin yalnızca %13,9'unu oluşturması, nadir ve ağır vakaların öğrenilmesini zorlaştıran belirgin bir sınıf dengesizliği problemine işaret etmektedir.



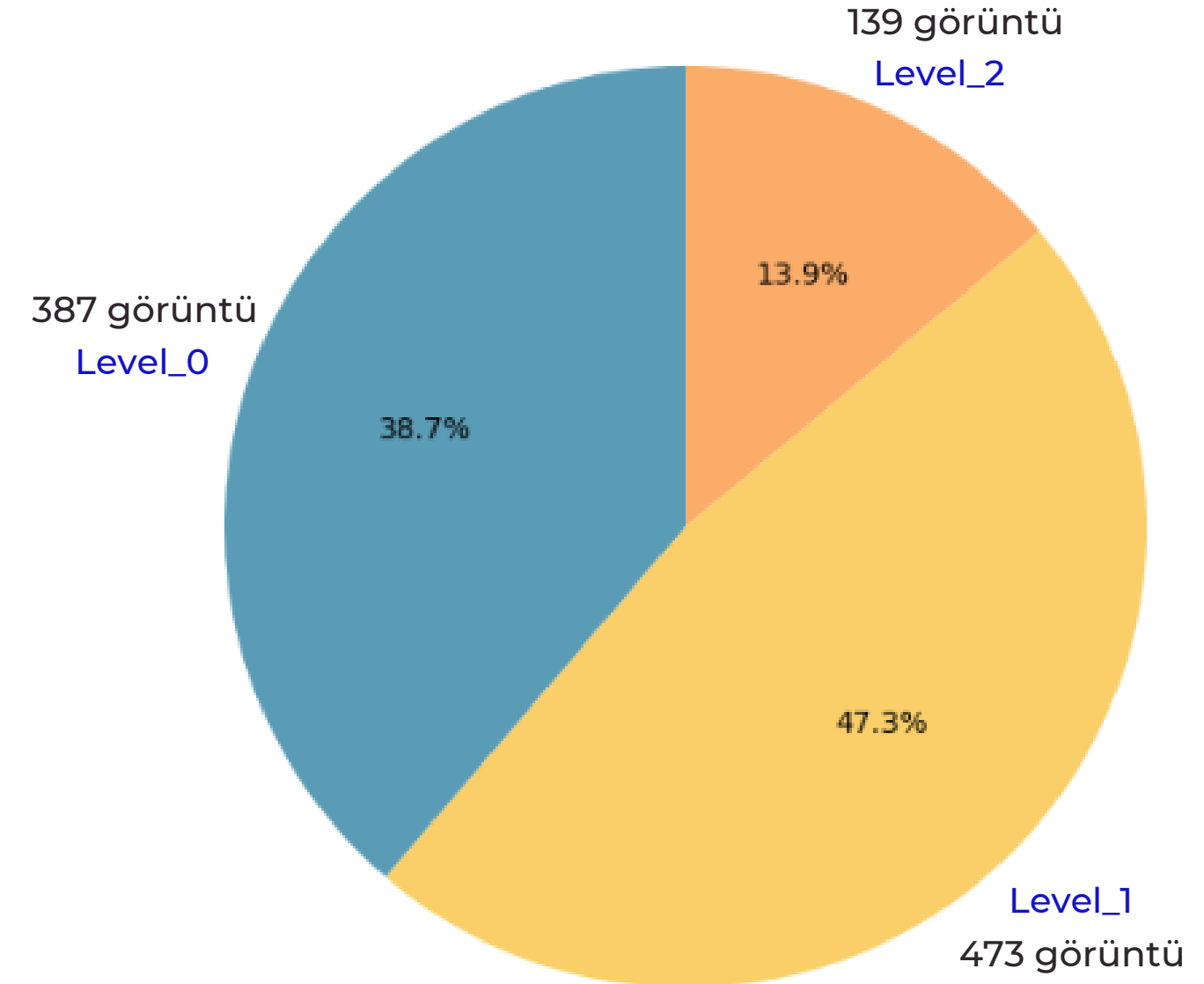
Level_0



Level_1

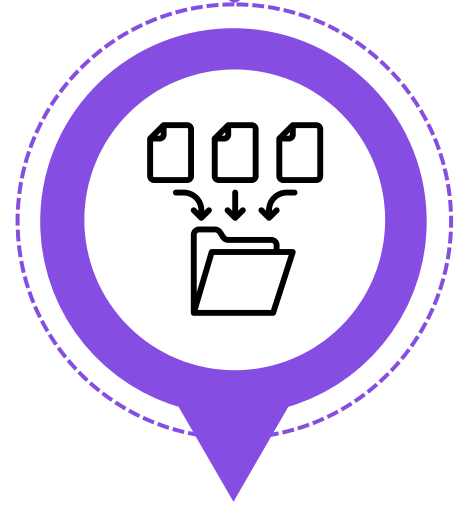


Level_2



UYGULANAN YÖNTEMLER

Girdi Görüntüleri ve Ön İşleme



Görüntüler, Vision Transformer (ViT) mimarisine uygun şekilde 224x224 boyutuna ölçeklendirilmiş ve 1/255 oranında normalize edilmiştir.

Veri Dengeleme ve Augmentation



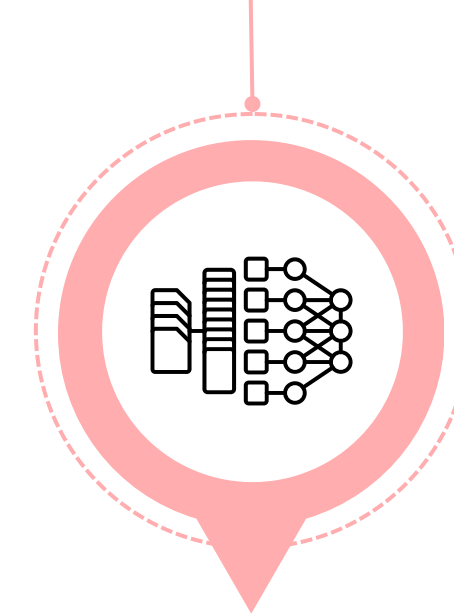
Sınıf dengesizliğini gidermek için Balanced Focal Loss kullanılmış; azınlık sınıfı olan Level_2'ye 3.60 kat daha fazla ağırlık verilmiştir. Ayrıca Mixup Augmentation ile veri çeşitliliği artırılmıştır.

Bayesian Optimizasyon ve Keras Tuner



Model için en uygun öğrenme hızı, dropout oranı ve katman derinliği, Keras Tuner kullanılarak Bayesian Optimization ile belirlenmiş ve en verimli çalışma noktası istatistiksel olarak tespit edilmiştir.

Fine-Tuning ve Dinamik Eğitim



Akneye özgü öznetelikleri yakalamak için ViT(Vision Transformer)'in üst katmanları açılarak Fine-Tuning uygulanmış; Early Stopping ile aşırı öğrenme önlenmiş, ReduceLROnPlateau ile öğrenme hızı dinamik olarak ayarlanmıştır.

Performans Değerlendirme



Geleneksel CNN yapıları yerine, görüntüyü yamalara (patches) ayırarak işleyen ve global ilişkileri Self-Attention mekanizmasıyla yakalayan ViT-Base modeli tercih edilmiştir.

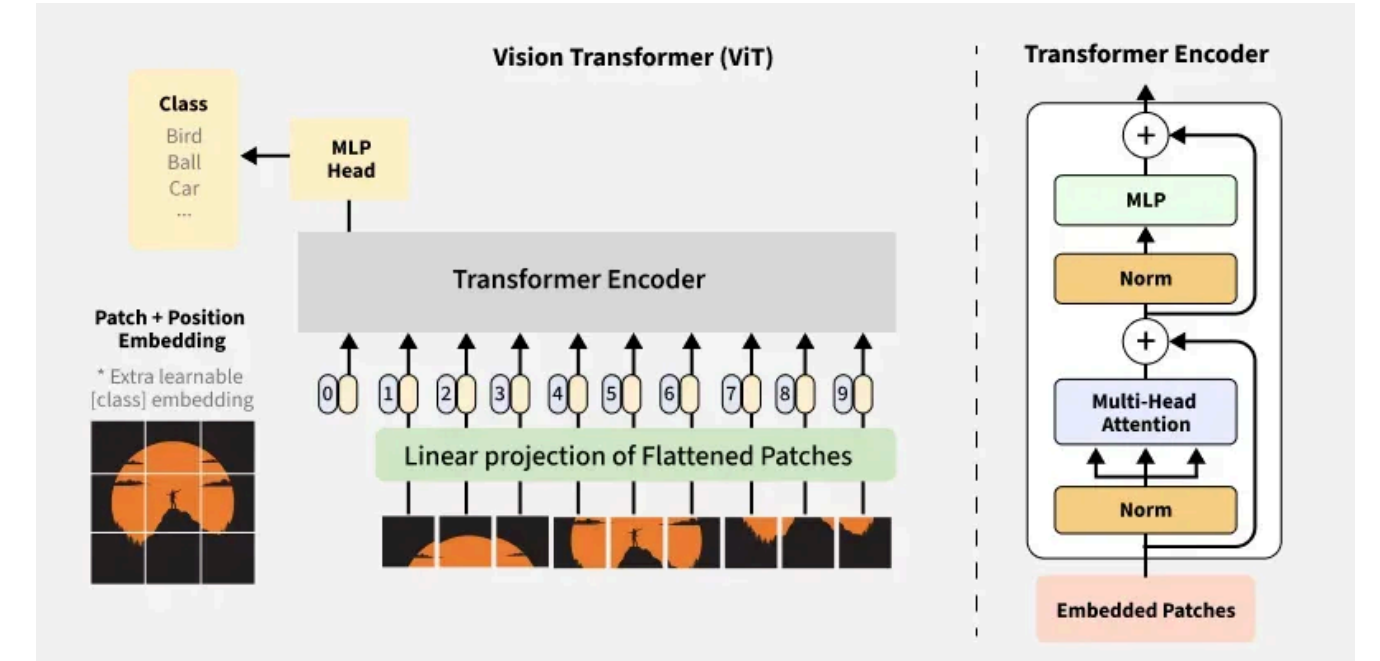
MODEL

ViT

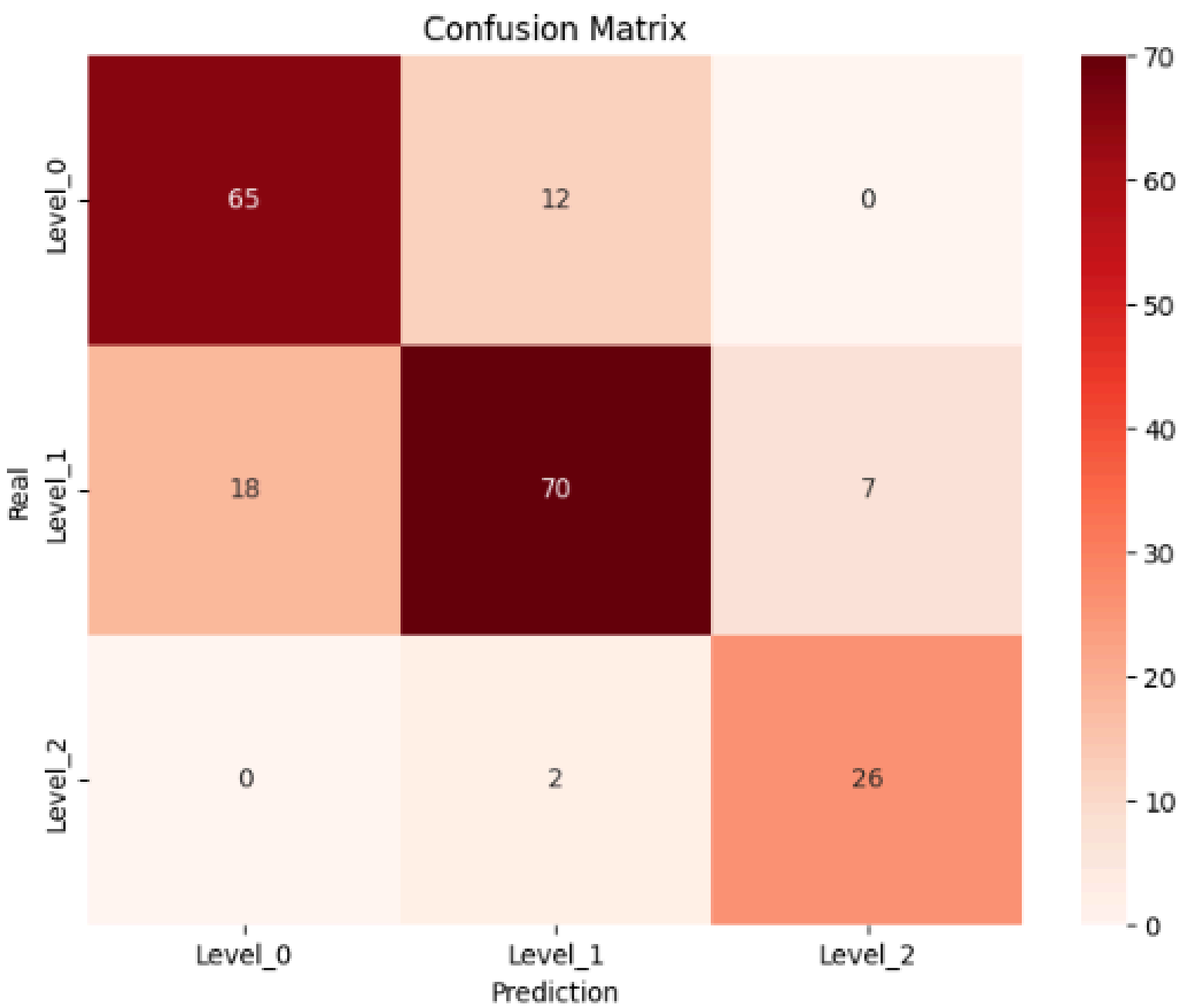
ImageNet-21k ön eğitimli Vision Transformer (ViT), görüntüyü yamalara bölerek Self-Attention mekanizmasıyla işler. Bu yapı, akne lezyonlarının farklı ölçeklerdeki dokusal özelliklerini hem yerel hem de küresel bağlamda eşzamanlı öğrenerek, uzun menzilli ilişkileri analiz eder ve geleneksel yöntemlere kıyasla daha yüksek hassasiyet sağlar.

EĞİTİM AŞAMASI

- Model Yapısı: ImageNet-21k ön eğitimli ViT kullanılmış, Class Token çıktısı üzerine GELU'lu 512 ve 256 birimlik yoğun katmanlar eklenmiştir.
- Eğitim Stratejisi: Model, aşırı öğrenmeyi engellemek için Dropout (0.4) ve Early Stopping (patience: 15) ile eğitilmiştir.
- Optimizasyon: Öğrenme hızı, ReduceLROnPlateau callback mekanizmasıyla val_loss durakladığında otomatik olarak düşürülmüştür.



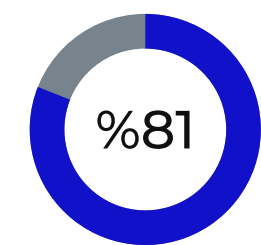
MODEL PERFORMANSI



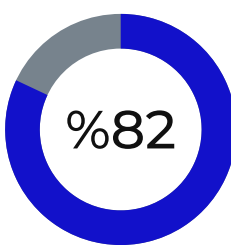
Sınıf Bazlı Performans

Sınıf	Precision	Recall	F1-score
Level_0	0.78	0.84	0.81
Level_1	0.83	0.74	0.78
Level_2	0.79	0.93	0.85

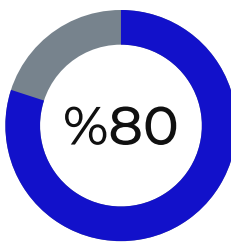
Genel Bazlı Performans



Accuracy

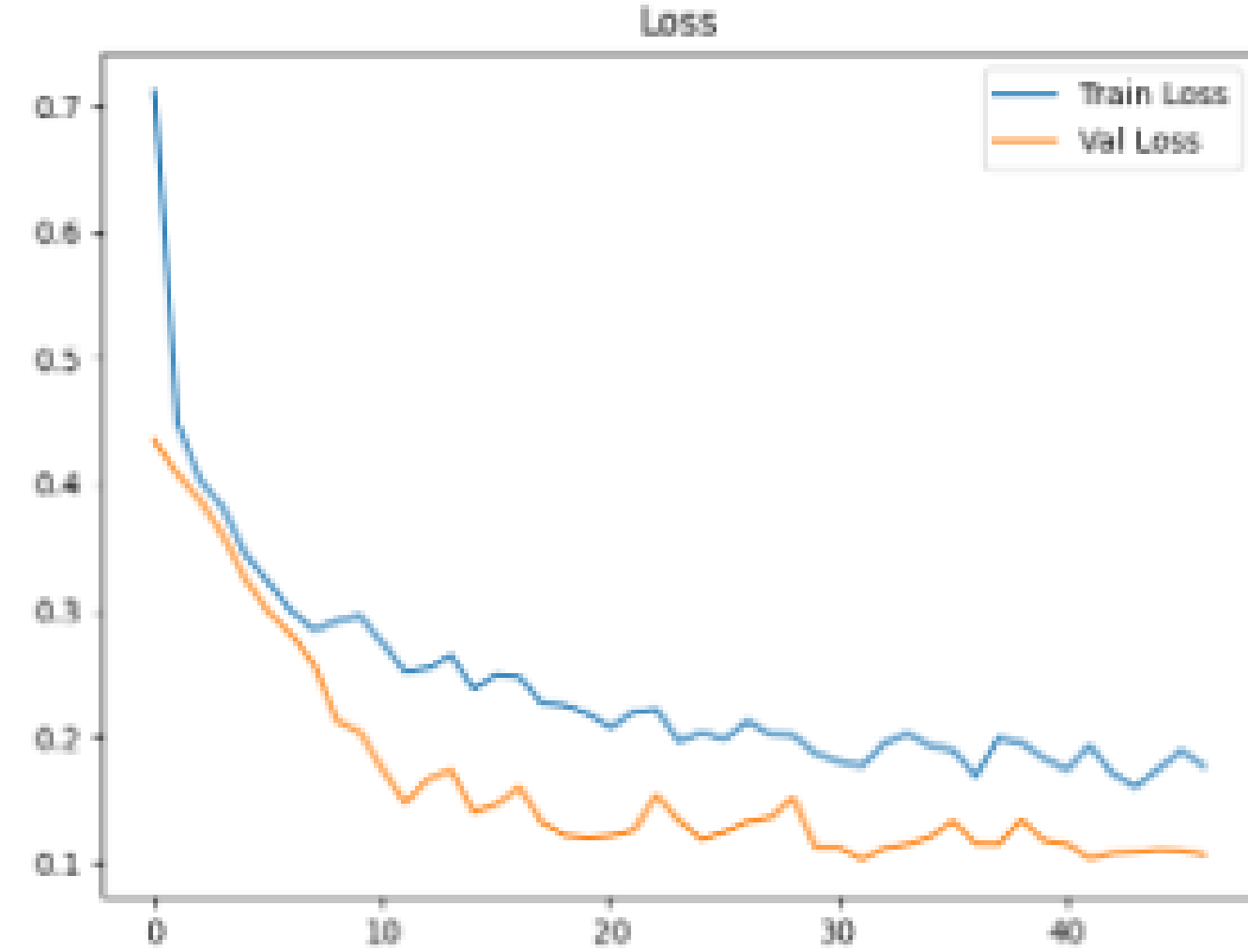


Macro Average



Weighted Average

GENEL DEĞERLENDİRME



Vision Transformer (ViT) mimarisi, akne lezyonları arasındaki dokusal ilişkileri CNN tabanlı modellere kıyasla daha hassas yakalamıştır. Balanced Focal Loss ve Mixup kullanımı sayesinde en zor sınıf olan ağır vakalar (Level_2) %93 recall ile başarıyla tespit edilmiştir. Model, doğrulama setinde %81 genel doğruluk elde ederek klinik karar destek sistemleri için güçlü bir temel sunmuştur.

TEŞEKKÜRLER