

ACNE SEVERITY CLASSIFICATION WITH VISION TRANSFORMER (ViT)

S E N A Y I L D I Z

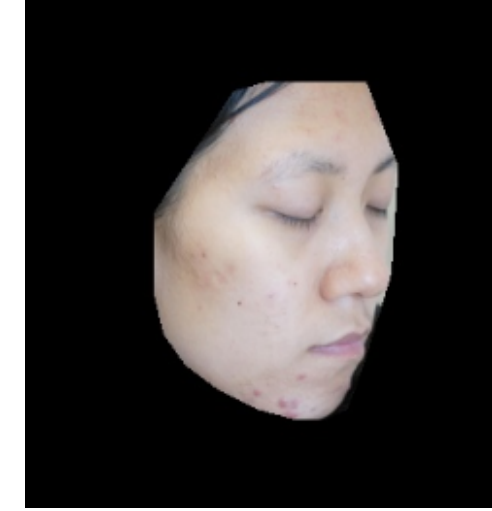
DATASET TANIMI

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.



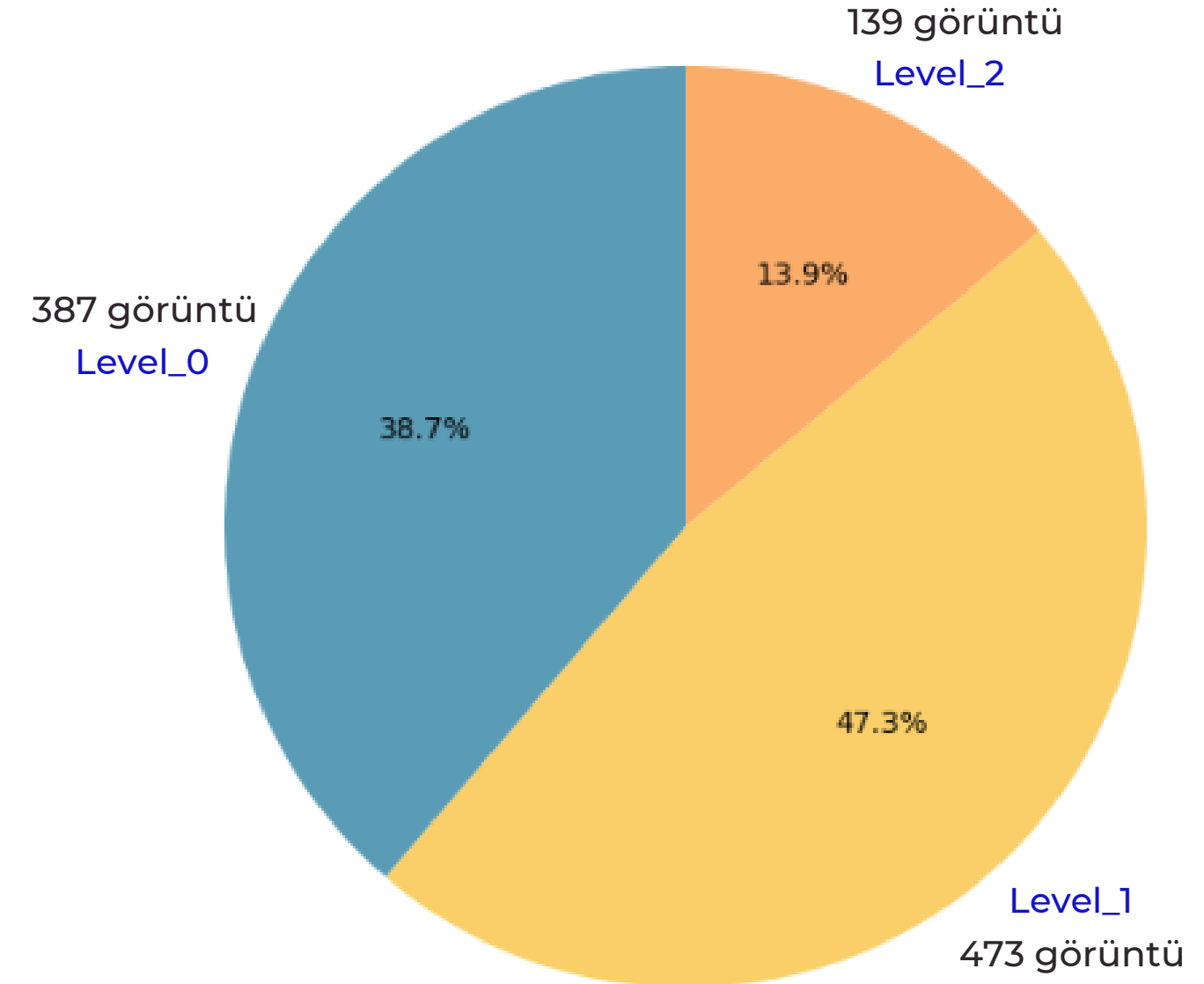
Level_0



Level_1

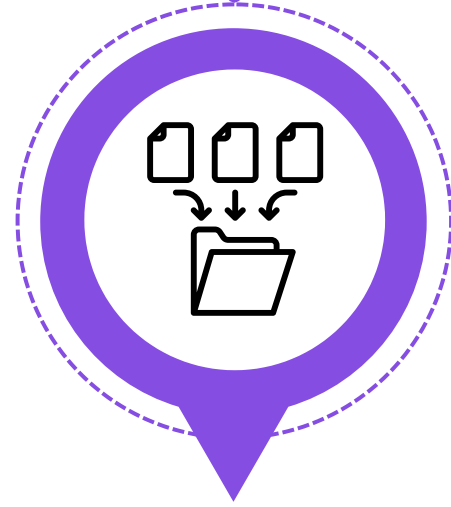


Level_2



UYGULANAN YÖNTEMLER

Girdi Görüntüleri ve Ön İşleme



Görüntüler ViT mimarisine uygun şekilde 224x224 boyutuna ölçeklendirilmiş ve normalize edilmiştir. Ek olarak, SSL (SimCLR) aşaması için görüntülerin farklı "view"larını oluşturan rastgele kırpma ve renk distorsiyonu içeren özel bir ön işleme hattı kurulmuştur.

Veri Dengeleme ve Augmentation



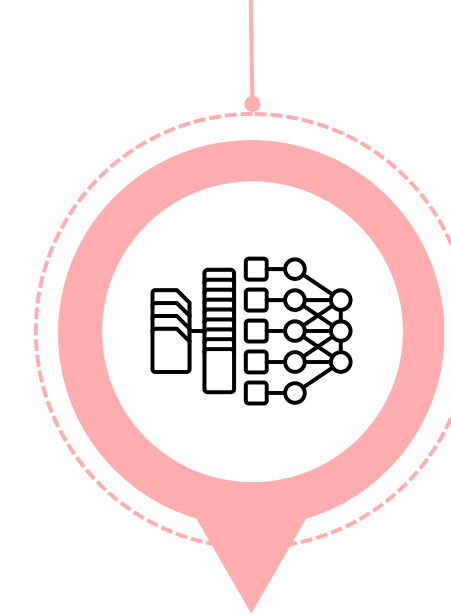
Sınıf dengesizliğini gidermek için Balanced Focal Loss kullanılarak Level_2 (Ağır Akne) sınıfına 3.60 kat daha fazla ağırlık verilmiştir. Eğitim sırasında modelin genelleme yeteneğini artırmak için döndürme, yatay çevirme ve parlaklık değişimlerini içeren dinamik veri artırımı uygulanmıştır.

Öz-Denetimli Öğrenme (SSL) ve SimCLR



Modelin kısıtlı veriden maksimum özellik öğrenmesi için SimCLR algoritması kullanılmıştır. Bu aşamada model, etiketlere ihtiyaç duymadan kontrastif kayıp (contrastive loss) kullanarak akne dokularını tanımayı öğrenmiş, ardından sınıflandırma eğitimine geçilmiştir.

Bayesian Optimizasyon ve Keras Tuner



Modelin en yüksek performansı vermesi için manuel deneme yerine Keras Tuner ve Bayesian Optimization kullanılmıştır. Bu sayede learning_rate (2e-5) ve dropout_rate (0.5) gibi kritik parametreler en optimum sonuçları verecek şekilde otomatik olarak seçilmiştir.

Fine-Tuning ve Dinamik Eğitim



Ön eğitimi tamamlanmış ViT-B16 modeli, özel sınıflandırma katmanları eklenerek akne teşhisi için fine-tune edilmiştir. Eğitim sürecinde ReduceLROnPlateau ile öğrenme hızı dinamik olarak yönetilmiş ve EarlyStopping ile aşırı öğrenme (overfitting) engellenmiştir.

MODEL

Vision Transformer (ViT-B16)

Model olarak Google tarafından geliştirilen vit_b16 (Base 16) mimarisi temel alınmıştır.

SSL ve SimCLR Entegrasyonu

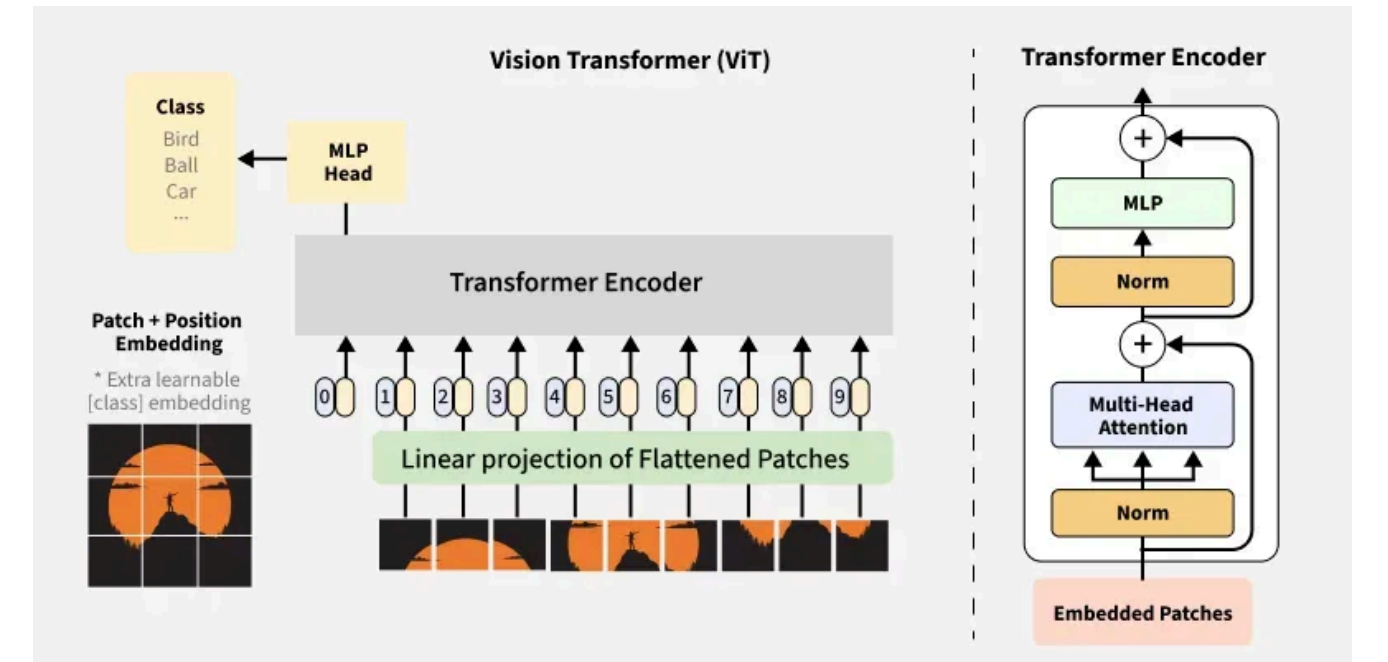
Self-Supervised Learning (SSL); sınırlı veriyi daha verimli kullanmak adına, etiketli eğitimden önce modelin dokuları kendi başına öğrenmesi sağlanmıştır. SimCLR Mimarisi, görüntülerin farklı varyasyonları (kırpılmış, döndürülmüş) arasındaki benzerliği ölçen bir "Projection Head" katmanı eklenerek kontrastif öğrenme (Contrastive Learning) uygulanmıştır.

Classification Head

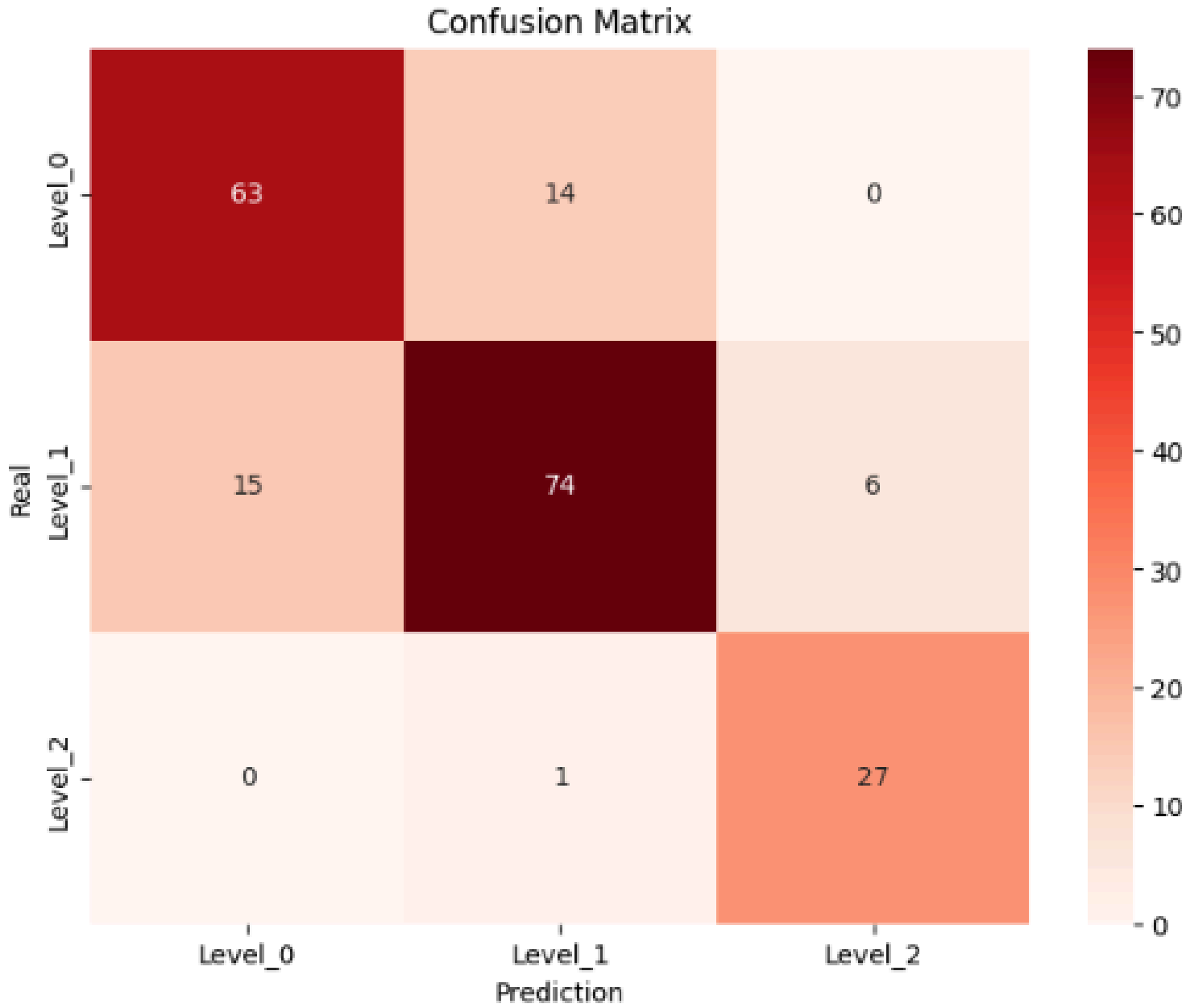
Kodda belirlenen en iyi parametrelerle şu katmanlar eklenmiştir: Global Average Pooling (1D), Dropout (0.5), Dense Layer (Level 0, 1, 2) ve Softmax katmanı.

Hiperparametre ve Optimizasyon

Optimizer olarak AdamW, Balanced Focal Loss, ve ReduceLROnPlateau kullanılmıştır.



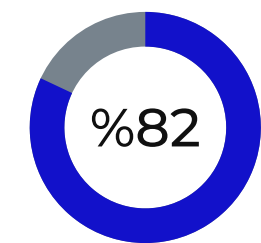
MODEL PERFORMANSI



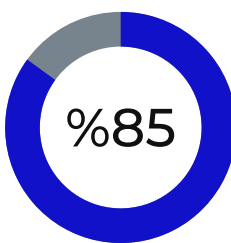
Sınıf Bazlı Performans

Sınıf	Precision	Recall	F1-score
Level_0	0.81	0.82	0.81
Level_1	0.83	0.78	0.80
Level_2	0.82	0.96	0.89

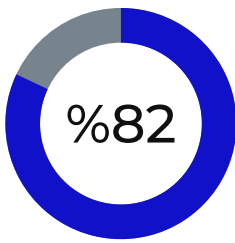
Genel Bazlı Performans



Accuracy

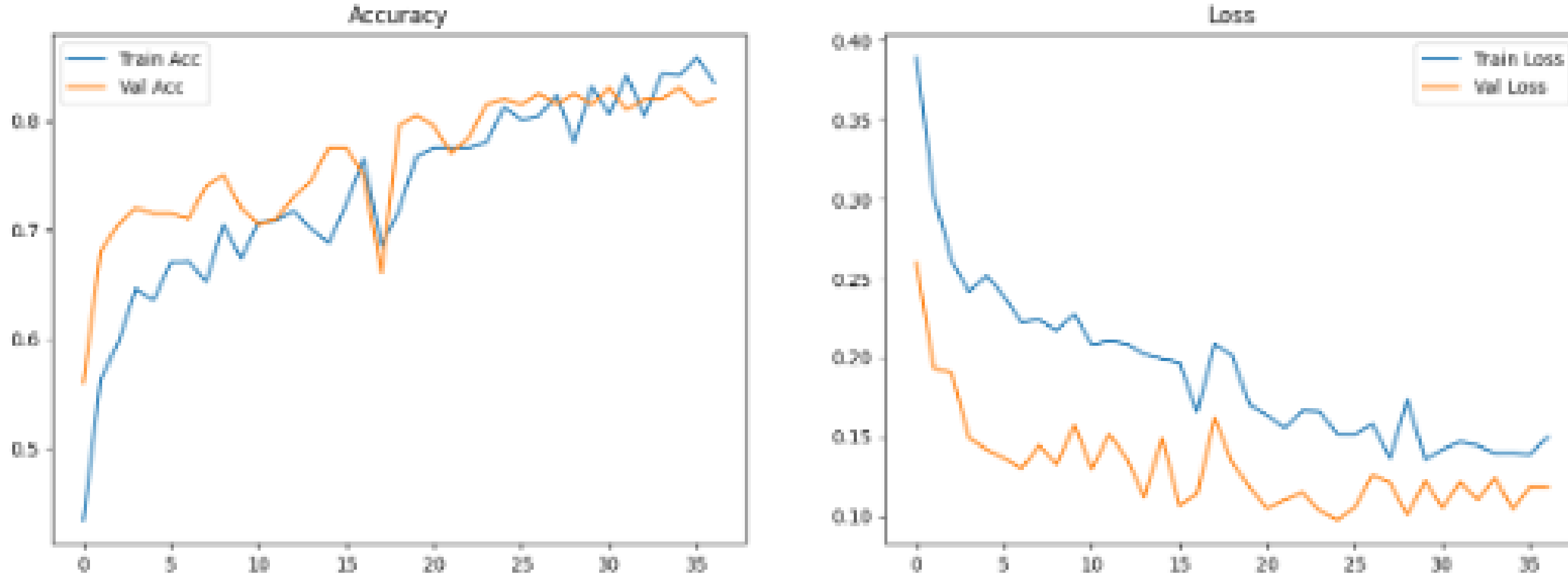


Macro Average



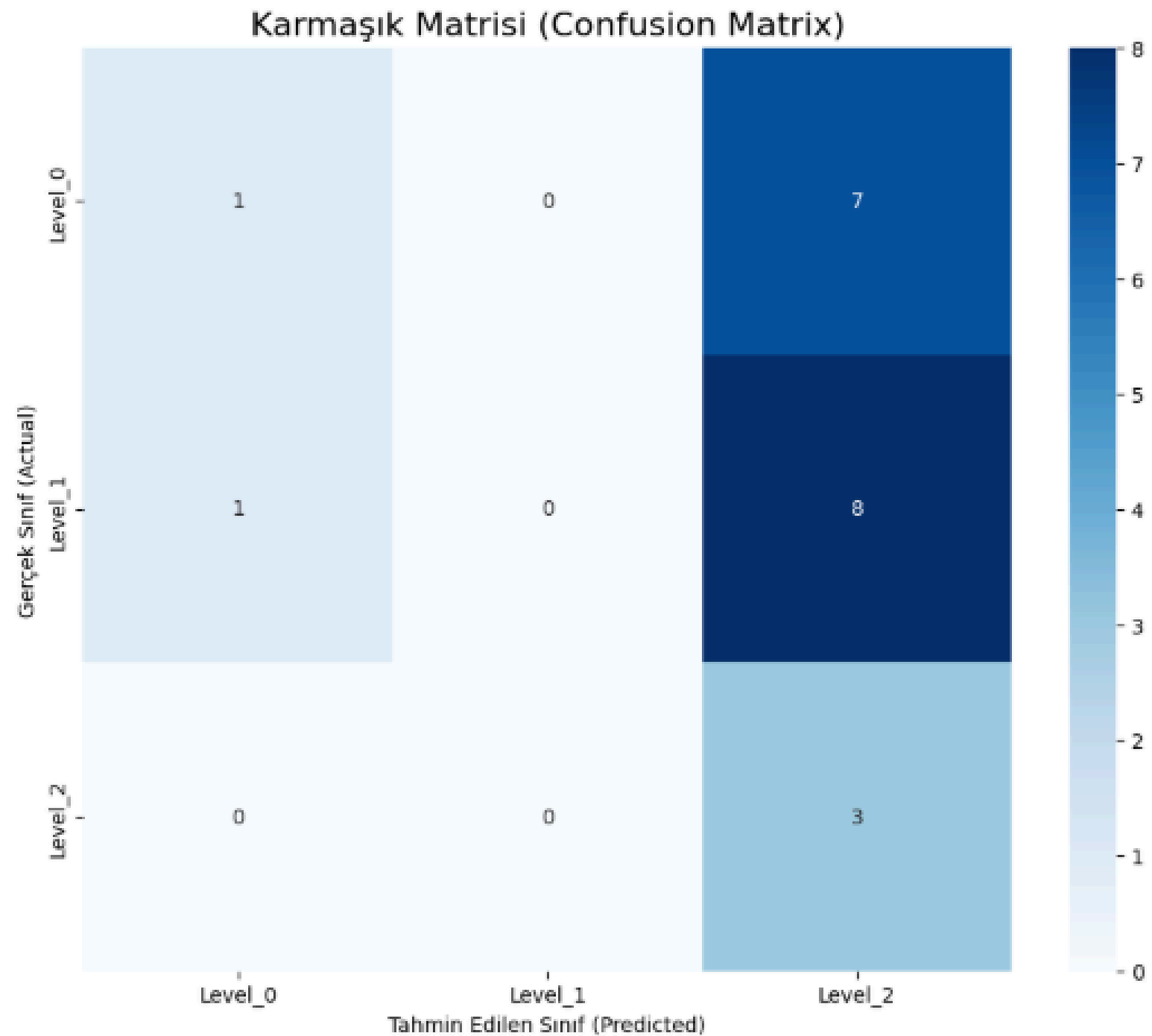
Weighted Average

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.

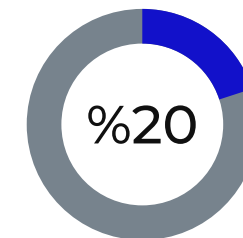
SSL MODEL PERFORMANSI



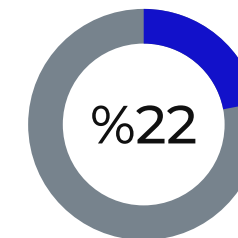
Sınıf Bazlı Performans

Sınıf	Precision	Recall	F1-score
Level_0	0.50	0.12	0.20
Level_1	0.80	0.00	0.00
Level_2	0.17	1.00	0.29

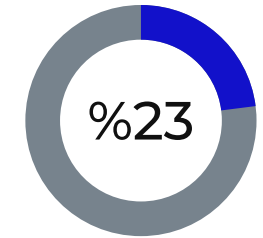
Genel Bazlı Performans



Accuracy

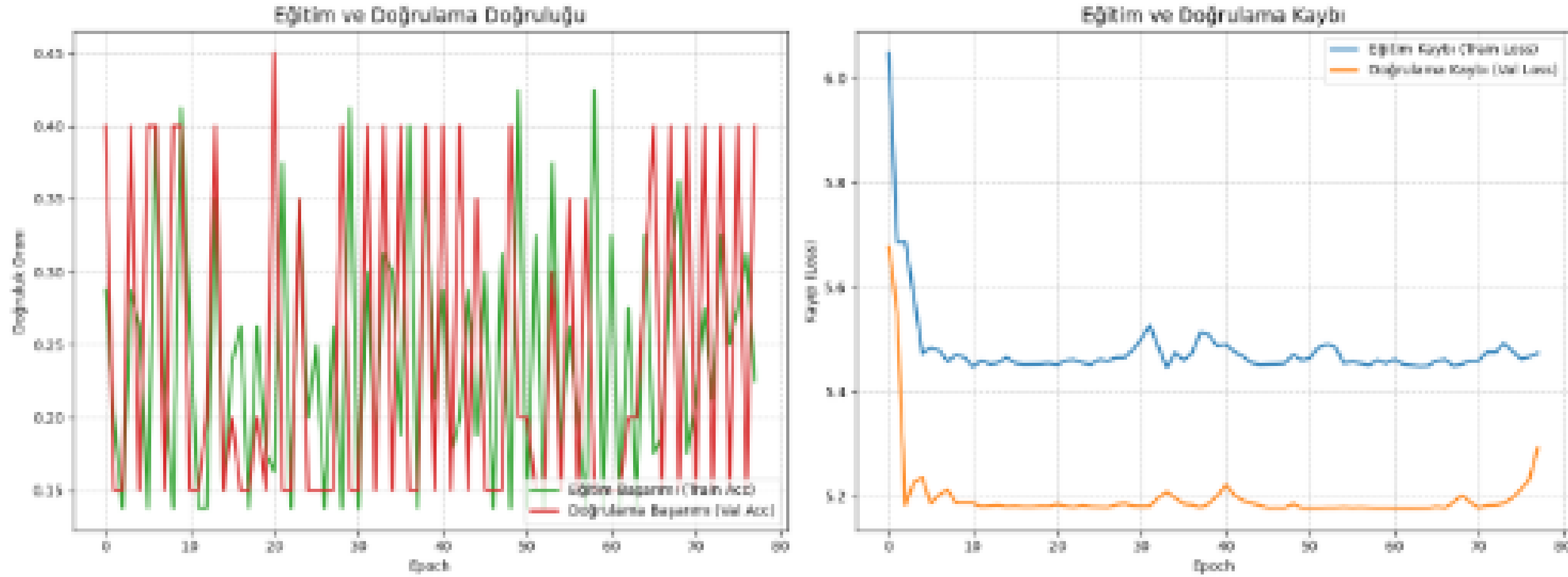


Macro Average



Weighted Average

SSL GENEL DEĞERLENDİRME



SSL aşamasında SimCLR ile akne dokularına ait ayırt edici temsillerin öğrenilmesi amaçlanmıştır; ancak doğrulama başarısı %20 seviyesinde kalmıştır. Bunun başlıca nedenleri, SSL için gerekli geniş ölçekli etiketlenmemiş verinin sınırlı olması, kontrastif öğrenmenin yüksek epoch ve hesaplama maliyeti gerektirmesi ve şiddetli akne (Level_2) sınıfının dengesiz dağılımıdır.

SONUÇ

Bu çalışmada, akne şiddeti sınıflandırmasında Vision Transformer (ViT) mimarisi standart eğitim ve Öz-denetimli Öğrenme (SSL) yaklaşımlarıyla karşılaştırılmıştır. Bayesian optimizasyonu ve Balanced Focal Loss ile eğitilen ViT-B16 modeli, %82 genel doğruluk ve kritik Level_2 sınıfında %96 recall elde ederek klinik kullanım için güçlü bir performans göstermiştir. Buna karşılık, SimCLR tabanlı SSL yaklaşımı veri ve eğitim kısıtları nedeniyle %20 doğrulukta kalmış; ancak daha büyük etiketlenmemiş veri setleriyle genelleme potansiyeli sunduğu görülmüştür. Geliştirilen model, nadir ağır vakaları yüksek hassasiyetle tespit ederek dermatologlar için etkili bir karar destek aracı olma potansiyeli taşımaktadır.

Model	Accuracy	Recall
ViT	0.82	0.12
ViT + SSL	0.20	0.100

TEŞEKKÜRLER