Bozok
Journal of Engineering and Architecture

e-ISSN: 3023-4298
Araştırma Makalesi/Research Article

Vision Transformer (ViT) ile MVTec Ahşap Verisinde Kusurlu ve Kusursuz Görüntülerin Sınıflandırılması

Kenan KILIÇ[1]

[1] *Yozgat Bozok Üniversitesi, Yozgat Meslek Yüksekokulu, Tasarım Bölümü, İç Mekân Tasarımı Programı, Yozgat, Türkiye*

MAKALE BİLGİSİ

Makale Tarihleri:

Geliş tarihi
16.10.2025
Kabul tarihi
17.11.2025
Yayın tarihi
31.12.2025

Anahtar Kelimeler:

Ağaçişleri Endüstrisi
Ahşap Kusurları
Bilgisayarlı Görü
Derin Öğrenme
Görsel Transformatör

ÖZET

MVTec Anomaly Detection veri setinin ahşap alt sınıfı üzerinde yüzey kusurlarının otomatik olarak sınıflandırılması amacıyla Google/ViT-Base-Patch16-224-in21k ve Microsoft/Swin-Tiny-Patch4-Window7-224 modellerinin performansları araştırılmıştır. Veri setine ait görüntüler 224×224 piksel boyutuna yeniden ölçeklendirilmiş, standart normalizasyon uygulanmış ve iki ayrı senaryo değerlendirilmiştir. İlk olarak veri artırma uygulanmadan, ikinci olarak veri artırma kullanılarak deneyler gerçekleştirilmiştir. Veri artırmasız durumda ViT modeli %95,45, Swin-Tiny modeli %93,94 doğruluk elde etmiştir. Veri artırma uygulandığında ViT modelinin doğruluğu %93,94, Swin-Tiny modelinin doğruluğu ise %95,45 olarak hesaplanmaktadır. Sonuçlar, her iki modelin de kusursuz, sıvı ve çizilme sınıflarında yüksek duyarlılık ve F1-puanı ürettiğini; buna karşın örnek sayısı düşük olan renk, birleşik ve delik sınıflarında sınıf dengesizliğine bağlı performans düşüşleri yaşandığını göstermektedir. Bu çalışma, Transformer modellerinin endüstriyel kalite kontrol süreçlerinde etkin bir alternatif olduğunu göstermekte; veri çeşitliliği ve sınıf dengeleme yöntemlerinin güçlendirilmesi durumunda modellerin doğruluklarının daha da artırılabileceğini göstermektedir.

Classification of Defect and Non-Defect Images in the MVTec Wood Dataset Using Vision Transformer (ViT)

ARTICLE INFO

Article history:

Received
16.10.2025
Accepted
17.11.2025
Published
31.12.2025

Keywords:

Woodworking Industry
Wood Defects
Computer Vision
Deep Learning
Vision Transformer (ViT)

ABSTRACT

The performances of the Google/ViT-Base-Patch16-224-in21k and Microsoft/Swin-Tiny-Patch4-Window7-224 models were investigated for the automatic classification of surface defects on the wood subclass of the MVTec Anomaly Detection dataset. The images of the dataset were rescaled to 224×224 pixels, standard normalisation was applied, and two separate scenarios were evaluated. The first experiment was conducted without data augmentation, and the second with data augmentation. In the no-data augmentation case, the ViT model achieved 95.45% accuracy, while the Swin-Tiny model achieved 93.94% accuracy. With data augmentation, the accuracy of the ViT model was calculated as 93.94%, and the Swin-Tiny model as 95.45%. The results show that both models produce high sensitivity and F1-scores for the defect-free, liquid, and scratch classes; however, performance decreases due to class imbalance in the colour, compound, and hole classes, which have low sample numbers. This study demonstrates that Transformer models are an effective alternative in industrial quality control processes, and demonstrates that the accuracy of the models can be further increased if data diversity and class balancing methods are strengthened.

1. GİRİŞ

Günümüzde ahşap ve türevi kaynaklarının azalması ve ihtiyacın sürekli olarak artması nedeniyle orman kaynaklarının tüketimini yavaşlatmak önemli bir araştırma konusu haline gelmiştir. Ahşap ve türevi malzemelerin yüzeyindeki budak vb. kusurlarının hızlı ve doğru bir şekilde tespit edilmesi, odun kullanım verimliliğini artırabilir ve aşırı tüketimi azaltabilir [1-4]. Bu amaçla, dijital

ORCID ID: Kenan KILIÇ: 0000-0003-1607-9545

***Sorumlu yazar(lar)/Corresponding author(s):** Yozgat Bozok Üniversitesi, Yozgat Meslek Yüksekokulu, Tasarım Bölümü, İç Mekân Tasarımı Programı, Yozgat, Türkiye
E-mail: kenan.kilic@bozok.edu.tr

Bu makaleye atıfta bulunmak için/To cite this article: K. Kılıç, "Vision Transformer (ViT) ile MVTec Ahşap Verisinde Kusurlu ve Kusursuz Görüntülerin Sınıflandırılması", Bozok Journal of Engineering and Architecture, vol. 4, no, 2, pp. 1-12, Dec 2025.

görüntü işleme ve yapay zekâ algoritmalarının birleşimi, odun düğüm kusurlarının tespiti ve sınıflandırılması için yaygın olarak kullanılan bir yöntem haline gelmiştir [5]. Bu yöntemler arasında, sabit özellik çıkarma ve sınıflandırma tanıma teknolojileri öne çıkmaktadır [6,7]. Bu teknolojiler, temel olarak bilgisayarlı görü, spektral analiz ve diğer dijital görüntü işleme tekniklerine dayanmaktadır [8]. Etkili özellik parametreleri, örneklerden sabit haritalama yöntemleriyle çıkarılır ve bu parametrelerin belirlenmesi için çeşitli istatistiksel veya makine öğrenimi yöntemleri karşılaştırılır. Derin öğrenme, yapay zekâ alanında büyük bir potansiyele sahip yeni bir yaklaşım olarak ortaya çıkmıştır [9]. Derin öğrenme yöntemi, orijinal verilerden özelliklerin otomatik olarak öğrenilmesini sağlar, böylece manuel özellik çıkarma işlemlerine olan bağımlılık azaltılabilir [10].

Endüstriyel üretim süreçlerinde kalite kontrol, ürünlerin güvenilirliği ve müşteri memnuniyeti açısından kritik bir rol oynamaktadır. Geleneksel kalite kontrol yöntemleri, genellikle insan operatörler tarafından gerçekleştirilmekte ve bu durum, zaman alıcı, maliyetli ve hata yapmaya açık bir sürece yol açmaktadır [11,12].

Endüstriyel görüntü analizi için kullanılan veri setlerinden biri olan MVTec veri seti, bu alanda yaygın olarak kullanılan bir referans kaynağıdır. MVTec, çeşitli endüstriyel ürünlerin (örneğin, tekstil, elektronik, metal parçalar) kusurlu ve kusursuz görüntülerini içeren zengin bir veri seti sunmaktadır [13]. Bu veri seti, kusur tespiti ve sınıflandırma modellerinin performansını değerlendirmek için bir standart haline gelmiştir.

Geleneksel olarak, evrişimli sinir ağları (ESA) görüntü sınıflandırma ve kusur tespiti görevlerinde yaygın olarak kullanılmıştır [14]. Ancak, son yıllarda Transformer mimarisi, doğal dil işleme alanındaki başarısının ardından bilgisayarlı görü alanında da etkili bir şekilde uygulanmaya başlanmıştır. Vision Transformer (ViT), görüntüleri parçalara ayırarak ve bu parçaları bir dizi olarak işleyerek, geleneksel ESA'ların sınırlamalarını aşmayı hedeflemektedir [15]. ViT, özellikle büyük ölçekli veri setlerinde yüksek performans göstermekte ve endüstriyel görüntü analizi gibi karmaşık görevlerde etkili bir alternatif sunmaktadır.

ViT'nin temel avantajı, görüntüleri doğrudan bir diziye dönüştürerek işlemesi ve bu sayede ESA'ların sınırlamalarını aşmasıdır. Özellikle, özyapısal dikkat mekanizması sayesinde, görüntünün farklı bölgeleri arasındaki ilişkileri etkili bir şekilde modelleyebilmektedir. Bu özellik, ViT'yi karmaşık görüntü analizi görevlerinde, özellikle de endüstriyel kusur tespiti ve sınıflandırma gibi alanlarda, güçlü bir alternatif haline getirmektedir [16].

ViT modelleri, özellikle büyük veri setleri üzerinde eğitildiklerinde yüksek genelleme yeteneği göstermektedir. Örneğin, ImageNet-21k gibi geniş kapsamlı veri setleri üzerinde önceden eğitilmiş ViT modelleri, küçük veri setlerinde bile etkili bir şekilde ince ayar yapılarak kullanılabilmektedir. Bu, ViT'nin endüstriyel uygulamalarda, özellikle de sınırlı sayıda etiketli veriye sahip olunan durumlarda, büyük bir potansiyele sahip olduğunu göstermektedir. Ancak, ViT modellerinin bazı sınırlamaları da bulunmaktadır. Özellikle, yüksek hesaplama kaynakları gerektirmesi ve küçük veri setlerinde aşırı uyum riski taşıması, bu modellerin kullanımını kısıtlayabilmektedir. Bu nedenle, ViT modellerinin performansını artırmak için veri artırma teknikleri, transfer öğrenme ve ince ayar gibi yöntemler sıklıkla kullanılmaktadır [15,17].

Bu çalışmada, MVTec veri seti kullanılarak endüstriyel görüntülerdeki kusurların tespiti ve sınıflandırılması için (ViT) tabanlı bir model önerilmektedir. ViT'nin bu tür görevlerdeki potansiyelini değerlendirmek amacıyla, modelin performansı geleneksel ESA tabanlı yöntemlerle karşılaştırılmıştır. Elde edilen sonuçlar, ViT'nin endüstriyel kusur tespiti ve sınıflandırma görevlerinde etkili bir yaklaşım olduğunu göstermektedir.

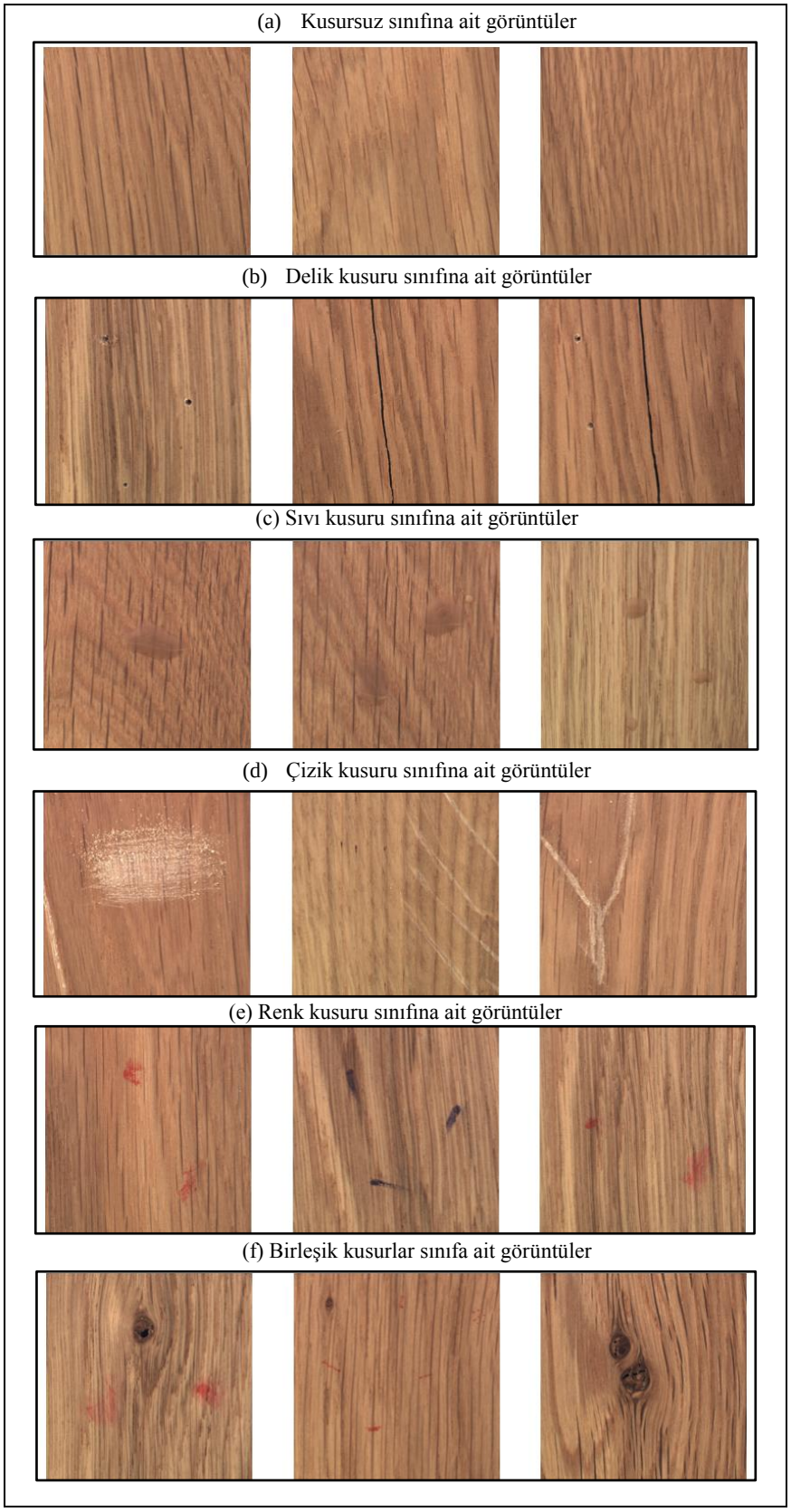
Bu çalışmayı literatürdeki benzer araştırmalardan ayıran en önemli fark, MVTec veri kümesinin yalnızca ahşap alt sınıfı üzerinde Google/ViT-Base-Patch16-224-in21k ve Microsoft/Swin-Tiny-Patch4-Window7-224 modellerinin doğrudan karşılaştırılmış olmasıdır. Literatürde bu modeller genellikle tüm MVTec veri kümesi veya farklı endüstriyel görüntü veri kümeleri üzerinde test edilirken, bu çalışma ahşap yüzey kusurları özelinde gerçekleştirilmiş karşılaştırmalı bir değerlendirme sunarak özgün bir katkı sağlamaktadır.

2. MATERYAL VE METOT

2.1. Veri Seti

Bu çalışma, MVTec Anomaly Detection (MVTec AD) veri setinin ahşap sınıfı üzerinde gerçekleştirilmiştir. Veri kümesi; renk kusuru, birleşik kusur, kusursuz, delik, sıvı ve çizik olmak üzere toplam altı sınıf içermektedir MVTec veri seti, endüstriyel görüntülerde kusur tespiti ve sınıflandırma için yaygın olarak kullanılan bir veri setidir. Bu çalışmada, MVTec veri setinin ahşap

sınıfı kullanılmıştır. Ahşap sınıfı, 247 eğitim ve 61 test görüntüsü içermekte olup, bu görüntüler yüksek çözünürlüklüdür (1024x1024 piksel). Görüntüler PNG formatındadır. Eğitim verisi, yalnızca kusursuz ahşap yüzeylerin görüntülerinden oluşurken, test verisi hem kusurlu hem de kusursuz görüntüler içermektedir. Kusur türleri arasında çizikler, çatlaklar, delikler ve renk bozuklukları bulunmaktadır. Bu veri seti, özellikle dokusal olarak karmaşık görüntülerdeki kusurları tespit etmek için kullanılan modellerin performansını değerlendirmek amacıyla tasarlanmıştır [13]. Veri setine ait görseller görüntüler Şekil 1'de verilmiştir.



Şekil 1. Veri setine ait örnek görüntüler

2.2. Veri Seti Üzerinde Ön İşlemler

MVTec Ahşap veri kümesi kusursuz ve çeşitli kusur türlerine sahip görüntülerden oluşmakta olup çalışmada modelleme sürecinde tüm görüntüler 224×224 piksel çözünürlüğe yeniden ölçeklendirilmiş ve RGB kanallarının her biri için ortalama 0,5 ve standart sapma 0,5 olacak şekilde normalize edilmiştir, modellerin genelleme kabiliyetini artırmak aşırı uyumu azaltmak ve sınıf içi görsel çeşitliliği genişletmek amacıyla eğitim verilerine bir dizi veri artırma dönüşümü uygulanmıştır, bu kapsamda ilk olarak görüntüler RandomResizedCrop işlemiyle 0,85–1,00 ölçek aralığında rastgele kırpılarak 224×224 piksele yeniden ölçeklendirilmiş ardından görüntüler %50 olasılıkla yatay çevrilmiş, ±10 derece aralığında rastgele döndürülmüş ve ColorJitter dönüşümü ile parlaklık kontrast ve doygunluk değerleri ±0,2 aralığında değiştirilmiştir, tüm görüntüler daha sonra tensöre dönüştürülerek normalize edilmiştir, bu veri artırma adımları modellerin ahşap yüzeylerdeki doğal renk doku ve geometrik varyasyonları daha geniş bir kapsamda öğrenmesini sağlamış ve düşük örnek sayısına sahip sınıflarda aşırı uyumu azaltarak genel test performansına katkı sunmuştur.

2.3. Vision Transformer (ViT) Modelleri

Google/ViT-Base-Patch16-224-in21k modeli, küresel dikkat mekanizması sayesinde görüntüdeki uzun menzilli ilişkileri yakalama ve karmaşık dokusal detayları öğrenme konusunda yüksek bir yetenek göstermektedir. Buna karşılık Microsoft/Swin-Tiny-Patch4-Window7-224 modeli, kaydırılmış pencere yapısı ile daha düşük hesaplama maliyeti sunmakta ve özellikle yerel örüntüleri verimli biçimde yakalayabilmektedir. Bu iki modelin karşılaştırılması, yüksek doğruluk sağlayan küresel dikkat temelli yaklaşımlar (ViT) ile hesaplama açısından verimli yerel dikkat temelli yapılar (Swin Transformer) arasındaki performans farklarının belirlenmesi ve değerlendirilmesi amacını taşımaktadır.

2.3.1. Google/vit-base-patch16-224-in21k Modeli

Vision Transformer (ViT), görüntü tanıma görevlerinde derin öğrenme alanına önemli bir yenilik kazandıran bir modeldir. Geleneksel evrişimli sinir ağlarından farklı olarak Transformer mimarisi üzerine inşa edilmiştir. Bu model, bir görüntüyü küçük parçalara (yamalara) ayırarak bu parçaları bir dizi hâlinde işler ve böylece uzun menzilli görsel ilişkileri öğrenebilir. Bu yaklaşım, özellikle büyük ölçekli veri kümelerinde yüksek doğruluk ve genelleme yeteneği sağlamaktadır [15].

Google/ViT-Base-Patch16-224-in21k modeli, ViT mimarisinin önceden eğitilmiş (pre-trained) bir sürümüdür ve ImageNet-21k veri kümesi üzerinde eğitilmiştir. ImageNet-21k; yaklaşık 14 milyon görüntü ve 21 841 sınıf içeren, görsel çeşitliliği oldukça yüksek bir veri kümesidir. Bu model, 16×16 piksel boyutundaki yamaları işler ve 224×224 piksel boyutundaki giriş görüntülerini kabul eder. Model, görüntü sınıflandırma, nesne tanıma ve görüntü segmentasyonu gibi birçok bilgisayarla görme görevinde etkin sonuçlar vermektedir.

ViT mimarisinin en önemli avantajı, ölçeklenebilirliği ve büyük veri kümelerinde güçlü genelleme kabiliyeti göstermesidir. Ayrıca, önceden eğitilmiş ağırlıklar sayesinde küçük ölçekli veri kümelerinde dahi etkili biçimde ince ayar yapılabilir. Bununla birlikte, ViT modelleri yüksek hesaplama kaynakları gerektirdiğinden, donanım maliyeti açısından görece yüksektir ve küçük veri kümelerinde aşırı uyum riski taşımaktadır.

Bu çalışmada, Google/ViT-Base-Patch16-224-in21k modeli, MVTec Ahşap Veri Kümesi üzerinde kusurlu ve kusursuz görüntülerin sınıflandırılması amacıyla kullanılmıştır. Model, ImageNet-21k üzerindeki ön eğitiminden elde ettiği genelleme yeteneğini transfer öğrenme yöntemiyle ahşap yüzeylere uygulamış ve yüzey kusurlarını tespit etme ile sınıflandırmada oldukça başarılı bir performans göstermiştir.

2.3.2. Microsoft/Swin-Tiny-Patch4-Window7-224 ViT Modeli

Bilgisayarlı görü görevlerinde kullanılan, Transformer tabanlı bir derin öğrenme modelidir. Swin Transformer mimarisi, geleneksel evrişimli sinir ağları yerine, görüntüleri parçalara ayırarak ve bu parçaları bir dizi olarak işleyerek özellik çıkarımı gerçekleştirir. Bu model, özellikle kaydırılmış pencere mekanizması ile dikkat çeker. Bu mekanizma, görüntüyü küçük pencerelere böler ve bu pencereler arasında bilgi alışverişini sağlayarak hem yerel hem de global özelliklerin etkili bir şekilde öğrenilmesine olanak tanır. Swin Transformer, görüntü sınıflandırma, nesne tespiti ve segmentasyon gibi görevlerde yüksek performans sunar [11].

Microsoft/Swin-Tiny-Patch4-Window7-224 modeli, Swin Transformer mimarisinin küçük bir versiyonudur ve özellikle kaynak kısıtlı ortamlarda kullanım için uygundur. Bu model, 4x4 boyutunda parçalara ayrılmış görüntüleri işler ve 224x224 piksel giriş boyutunu kabul eder. Ayrıca, 7x7 boyutunda bir pencere mekanizması kullanır. Model, büyük ölçekli görüntü veri kümeleri

(örneğin, ImageNet) üzerinde eğitilmiştir ve önceden eğitilmiş ağırlıkları sayesinde transfer öğrenme için oldukça uygundur. Bu özellik, modelin belirli bir görev için ince ayar yapılarak kullanılmasını mümkün kılmaktadır.

Swin Transformer'ın en önemli avantajlarından biri, ölçeklenebilir olması ve büyük veri setlerinde yüksek genelleme yeteneği göstermesidir. Ayrıca, Shifted Window mekanizması sayesinde hesaplama maliyetleri düşük tutulurken, modelin görüntüdeki hem yerel hem de global özellikleri öğrenmesi sağlanır. Bununla birlikte, Swin Transformer modelleri de diğer Transformer tabanlı modeller gibi yüksek hesaplama kaynakları gerektirir ve küçük veri setlerinde aşırı uyum riski taşımaktadır.

Bu çalışmada, Microsoft/Swin-Tiny-Patch4-Window7-224 modeli, MVTec Ahşap Veri Seti üzerinde kusurlu ve kusursuz görüntülerin sınıflandırılması görevi için kullanılmıştır. Model, ImageNet gibi büyük ölçekli veri setleri üzerindeki ön eğitiminden elde edilen bilgiyi transfer ederek, ahşap yüzeylerdeki kusurları tespit etme ve sınıflandırma konusunda etkili bir performans sergilemiştir. Bu sonuçlar, Swin Transformer mimarisinin endüstriyel görüntü analizi gibi karmaşık görevlerde de başarıyla uygulanabileceğini göstermektedir.

2.4. Deneysel Kurulum

Tüm görüntüler ilgili sınıf klasörlerinden otomatik olarak yüklenmiş ve sınıf etiketleri Python ortamında programatik olarak atanmıştır. Veri seti, eğitim ve test alt kümelerine %80–%20 oranında ayrılmış, rastgelelik kontrolü için random_state=42 kullanılarak sonuçların tekrarlanabilirliği sağlanmıştır. Görüntüler modellerin giriş boyutuna uyumlu olacak şekilde 224×224 piksel çözünürlüğe ölçeklendirilmiş, tensöre dönüştürüldükten sonra [0.5, 0.5, 0.5] ortalama ve [0.5, 0.5, 0.5] standart sapma değerleri ile normalize edilmiştir. Bu çalışmada, özellikle sınıf içi çeşitliliği artırmak ve düşük örnek sayısına sahip sınıflarda aşırı öğrenmeyi azaltmak amacıyla eğitim verisine çeşitli veri artırma (data augmentation) teknikleri uygulanmıştır. Kullanılan dönüşümler arasında rastgele yatay çevirme, ±10 derece aralığında rastgele döndürme, ColorJitter ile parlaklık ve kontrast varyasyonu ve RandomResizedCrop ile rastgele kırpma–yeniden ölçekleme işlemleri yer almaktadır. Bu artırma adımları, modellerin doku farklılıklarını daha geniş bir varyasyonla öğrenmesini sağlayarak genel performansa olumlu katkıda bulunmaktadır.

Eğitim işlemleri Kaggle Notebook ortamında GPU hızlandırmalı olarak yürütülmüş, oturumda NVIDIA T4 GPU (16 GB), çoklu GPU sanallaştırma altyapısı, en az 4 sanal CPU çekirdeği ve 25–30 GB RAM kullanılmıştır. Tüm işlemler CUDA üzerinden otomatik olarak GPU'ya atanmış; model eğitimi, Python + PyTorch altyapısı üzerinde gerçekleştirilmiş; model mimarileri Hugging Face Transformers, dönüşümler ise torchvision kütüphaneleri aracılığıyla uygulanmıştır. Veri işleme ve çıktı görselleştirmede scikit-learn, matplotlib ve seaborn kütüphaneleri kullanılmıştır.

Çalışmada iki farklı Vision Transformer tabanlı mimari incelenmiştir:

(1) Google/ViT-Base-Patch16-224, 16×16 piksellik yama yapısını kullanan klasik Vision Transformer mimarisidir.

(2) Microsoft/Swin-Tiny-Patch4-Window7-224, kaydırılmış pencere (shifted window) mekanizmasıyla çalışan çok ölçekli ve hiyerarşik Transformer yapısına sahiptir.

Her iki model de ImageNet üzerinde önceden eğitilmiş (pretrained) ağırlıklarla başlatılmış, sınıflandırma katmanları veri kümesindeki altı sınıf için yeniden düzenlenmiştir (num labels=6). Model uyumluluğu için ignore_mismatched_sizes=True parametresi kullanılmıştır.

Eğitim aşamasında her iki model için Cross-Entropy Loss kayıp fonksiyonu ve AdamW optimizasyon algoritması tercih edilmiştir. Öğrenme oranı 1×10^{-4}, mini-yığın boyutu 8, maksimum epoch sayısı 20 olarak belirlenmiştir. Aşırı uyumun (overfitting) önüne geçmek ve doğrulama kaybına duyarlı bir eğitim gerçekleştirmek amacıyla Early Stopping mekanizması uygulanmıştır. Bu doğrultuda doğrulama kaybı üst üste 5 epoch boyunca iyileşme göstermediğinde eğitim otomatik olarak durdurulmuş ve en iyi doğrulama performansına sahip model ağırlıkları kaydedilmiştir. Bu yöntem, özellikle küçük sınıflardaki örnek sayısının sınırlı olması nedeniyle modelin gereksiz yere fazla öğrenmesini önlemiş ve eğitim süresini optimize etmiştir.

Eğitim süreci ileri besleme, kayıp hesaplama, geri yayılım ve ağırlık güncelleme adımlarından oluşmuş; her epoch sonunda doğrulama aşaması çalıştırılarak güncel doğruluk, duyarlılık, kesinlik ve F1 skorları kaydedilmiştir. Tüm deneysel çalışmalar iki Python betiği ile yürütülmüştür: vit_mvtec_wood_train.py (ViT modeli) ve swin_mvtec_wood_train.py (Swin modeli). Her iki betik

aynı veri yapısı, aynı hiperparametreler ve aynı dönüşüm adımları ile çalıştırılmış; sonuçlar doğruluk, sınıf bazlı metrikler ve karışıklık matrisleri şeklinde değerlendirilerek makalede sunulmuştur.

2.5. Değerlendirme Metrikleri

Modelin performansını değerlendirmek için çeşitli sınıflandırma ölçütleri kullanılmıştır. Bu kapsamda doğruluk, duyarlılık, kesinlik ve F1 puanı temel metrikler olarak hesaplanmıştır. Ayrıca, karışıklık matrisi ile modelin sınıf bazlı başarı durumu görsel olarak gösterilmiştir. Kullanılan değerlendirme metrikleri ve formülleri Tablo 1'de verilmiştir.

Tablo 1. Değerlendirme metrikleri, açıklama ve formülleri

Metrik	Açıklama	Formül
Doğruluk (Accuracy)	Doğru sınıflandırılan örneklerin toplam örneklere oranı	Accuracy = (TP + TN) / (TP + TN + FP + FN)
Kesinlik (Precision)	Pozitif tahminlerin gerçekten pozitif olma oranı	Precision = TP / (TP + FP)
Duyarlılık (Recall)	Gerçek pozitiflerin doğru tahmin edilme oranı	Recall = TP / (TP + FN)
F1 puanı	Precision ve Recall'un harmonik ortalaması	F1 = 2 × (Precision × Recall) / (Precision + Recall)
*Not: TP: Doğru Pozitif, TN: Doğru Negatif, FP: Yanlış Pozitif, FN: Yanlış Negatif		

3. BULGULAR VE TARTIŞMA

Bu çalışmada, endüstriyel görüntü analizi için kusur tespiti ve sınıflandırma görevlerinde iki farklı derin öğrenme modelinin veri artırmalı ve veri artırmasız performansları araştırılmıştır. Kullanılan modeller, Google/ViT-Base-Patch16-224-in21k (Vision Transformer) ve Microsoft/Swin-Tiny-Patch4-Window7-224 (Swin Transformer) ve bu modellerin veri artırmalı modelleridir. Her iki model de Transformer mimarisine dayanmakta olup, görüntü işleme görevlerinde kullanılmaktadır. Bu çalışmanın amacı, bu modellerin MVTec Ahşap Veri Seti üzerindeki kusur sınıflandırma performanslarını değerlendirmek ve karşılaştırmaktır.

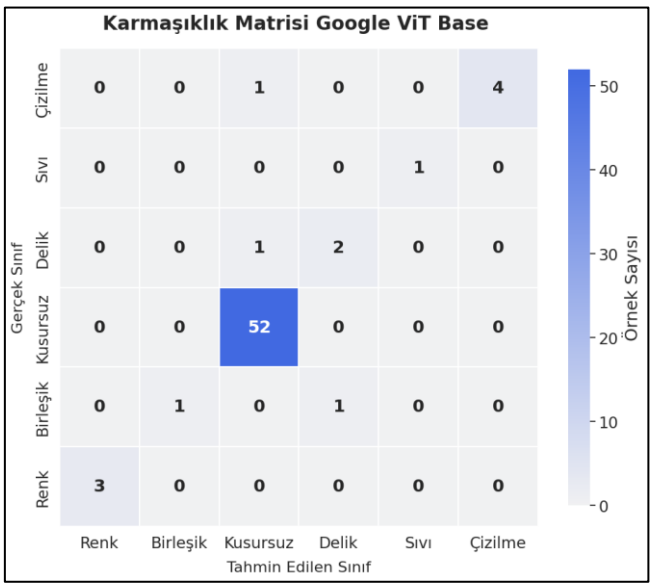
Google/ViT-Base-Patch16-224-in21k ve Microsoft/Swin-Tiny-Patch4-Window7-224 modellerinin sınıflandırma performans sonuçlarına ait performanslar değerlendirilirmiştir. Her iki modelin de her bir sınıf için kesinlik, duyarlılık, F1-puanı ve örnek sayısı değerleri, ayrıca genel doğruluk, makro ortalama ve ağırlıklı ortalama değerleri sunulmaktadır. Bu sonuçlar, modellerin güçlü ve zayıf yönlerini ortaya koymakta ve hangi sınıflarda iyileştirmelere ihtiyaç duyulduğunu göstermektedir. Google/vit-base-patch16-224-in21k modeli sınıflandırma performansı Tablo 2'de, Google/vit-base-patch16-224-in21k modelin veri artırmalı sınıflandırma performansı Tablo 3'te verilmektedir. Microsoft/Swin-Tiny-Patch4-Window7-224 modeli performansı Tablo 4'te, Microsoft/Swin-Tiny-Patch4-Window7-224 modelin veri artırmalı performansı Tablo 5'da verilmektedir.

Tablo 2. Google/vit-base-patch16-224-in21k modeli sınıflandırma performansı

Sınıf	Kesinlik	Duyarlılık	F1-puanı	Örnek sayısı
Renk	1,0000	1,0000	1,0000	3
Birleşik	1,0000	0,5000	0,6667	2
Kusursuz	0,9630	1,0000	0,9811	52
Delik	0,6667	0,6667	0,6667	3
Sıvı	1,0000	1,0000	1,0000	1
Çizilme	1,0000	0,8000	0,8889	5
Doğruluk			**0,9545**	66
Makro Ort.	0,9983	0,8278	0,8672	66
Ağırlıklı Ort.	0,9557	0,9545	0,9515	66

Tablo 2 incelendiğinde, Google/vit-base-patch16-224-in21k modelinin sınıflandırma performansı genel olarak %95,45'tir. Sonuçlar modelin çoğu sınıfı yüksek doğrulukla tanıyabildiğini göstermektedir. Modelin eğitim süresi: 136,47 saniye, test süresi: 0,93 saniyedir. Özellikle örnek sayısının yüksek olduğu kusursuz sınıfında model %96,30 kesinlik, %100 duyarlılık ve %98,11 F1-puanı ile oldukça güçlü bir performans sergilemektedir. Bu durum modelin kusursuz yüzey görüntülerini ayırt etmede tutarlı bir genelleme yeteneği geliştirdiğini ortaya koymaktadır. Benzer biçimde renk ve sıvı sınıflarında tüm metriklerin 1,0000 seviyesinde gerçekleşmesi, modelin bu sınıflara ait karakteristik özellikleri hatasız biçimde yakalayabildiğini göstermektedir. Ancak örnek sayısı düşük olan birleşik ve delik sınıflarında performansın belirgin biçimde düştüğü görülmektedir. Birleşik sınıfında duyarlılığın 0,5000'de kalması ve delik sınıfında F1-puanının 0,6667 olması, modelin bu sınıfların temsil gücünden yoksun olması sebebiyle sınırlı öğrenme gerçekleştirdiğine göstermektedir. Düşük örnek sayısı olan sınıflarda, bu sınıfların model tarafından yeterince temsil edilememesine ve bu nedenle sınıf içi varyasyonların öğrenilememesine sebep olmaktadır. Bununla birlikte modelin genel doğruluk oranının %95,45 gibi yüksek bir seviyede olması, sınıflar arasındaki görsel farklılıkların büyük ölçüde doğru şekilde yakalandığını göstermektedir. Ağırlıklı ortalama F1-puanının 0,9515 olarak hesaplanması modelin dengeli sınıflarda oldukça başarılı olduğunu, makro ortalamanın 0,9983'te kalması ise düşük örnek sayısına sahip sınıflardaki performans kayıplarının sonuçlara yansıdığını

göstermektedir. Genel olarak değerlendirildiğinde model, çoğu kusur tipini yüksek doğrulukla tespit edebilmekte; ancak nadir görülen sınıflarda ek veri artırma, sınıf ağırlıklandırma veya daha dengeli veri toplama gibi stratejilerle performansın daha da iyileştirilebileceği anlaşılmaktadır. Google/ViT-Base-Patch16-224-in21k modeline ait karmaşıklık matrisi Şekil 2'de verilmiştir.



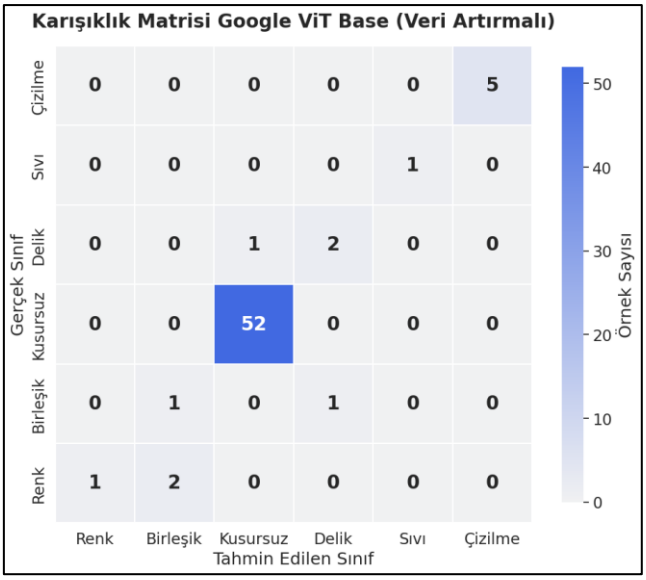
Şekil 2. Google/ViT-Base-Patch16-224-in21k karmaşıklık matrisi

Şekil 2'deki karmaşıklık matrisi incelendiğinde, Google/ViT-Base-Patch16-224-in21k modelinin kusursuz sınıfını yüksek doğrulukla ayırt ettiğini, ancak renk, birleşik ve özellikle delik sınıflarında belirgin karışmalar yaşadığını göstermektedir. Modelin çizilme ve sıvı sınıflarında orta düzeyde başarı elde ettiği, düşük örnek sayısına sahip sınıflarda ise ayrım gücünün zayıfladığı anlaşılmaktadır. Bu bulgular, modelin belirgin dokusal yapıları iyi öğrendiğini ancak benzer görsel özelliklere sahip kusur sınıflarında zorlandığını göstermektedir.

Tablo 3. Google/vit-base-patch16-224-in21k modelin veri artırmalı sınıflandırma performansı

Sınıf	Kesinlik	Duyarlılık	F1-puanı	Örnek sayısı
Renk	1,0000	0,3333	0,5000	3
Birleşik	0,3333	0,5000	0,4000	2
Kusursuz	0,9811	1,0000	0,9905	52
Delik	0,6667	0,6667	0,6667	3
Sıvı	1,0000	1,0000	1,0000	1
Çizilme	1,0000	1,0000	1,0000	5
Doğruluk			**0,9394**	66
Makro Ort.	0,8302	0,7500	0,7595	66
Ağırlıklı Ort.	0,9498	0,9394	0,9364	66

Tablo 3 incelendiğinde, Google/vit-base-patch16-224-in21k veri artırmalı modelin sınıflandırma performansı genel %94 e yakın bir seviyededir. Bazı sınıflarda belirgin performans farklılıkları gözlenmektedir. Veri artırmalı halde kullanılan bu modelin eğitim süresi: 132,67 saniye, test süresi: 0,94 saniyedir. Özellikle veri sayısının yüksek olduğu kusursuz sınıfında modelin %98,11 kesinlik, %100 duyarlılık ve %99,05 F1-puanı ile neredeyse hatasız bir sınıflandırma gerçekleştirdiği görülmektedir. Bu durum modelin kusursuz yüzey özelliklerini ayırt etmekte istikrarlı ve güvenilir bir genelleme yeteneği geliştirdiğini göstermektedir. Benzer şekilde, sıvı ve çizilme sınıflarında tüm metriklerin 1,0000 olması, bu sınıfların kendine özgü görsel karakteristiklerinin model tarafından açık biçimde ayırt edilebildiğine göstermektedir. Bununla birlikte, örnek sayısının düşük olduğu renk ve birleşik sınıflarında performansın belirgin biçimde düştüğü gözlenmektedir. Renk sınıfında duyarlılığın 0,3333 seviyesine gerilemesi, modelin bu sınıfa ait görüntülerin önemli bir bölümünü doğru şekilde tespit edemediğini göstermektedir. Birleşik sınıfında ise kesinlik 0,3333 iken duyarlılığın 0,5000 olması, modelin sınıf örneklerini tutarsız biçimde tahmin ettiğini ve düşük örnek sayısı nedeniyle yeterli temsil öğrenemediğini ortaya koymaktadır. Bu sınıflardaki düşük performans, veri dengesizliğinin model eğitimini olumsuz etkilediğini ve nadir görülen sınıflarda hataların daha belirgin hale geldiğini göstermektedir. Genel doğruluk oranının %93,94 gibi bir değerde gerçekleşmiş olması, modelin genel sınıflandırma kapasitesinin güçlü olduğunu kanıtlarken, makro ortalama F1-puanının 0,7595'e düşmesi, küçük örnekli sınıflardaki performans kayıplarının genel istatistiklere yansıdığını göstermektedir. Ağırlıklı ortalama F1-puanının 0,9364 olarak hesaplanmış olması ise iyi temsil edilen sınıfların model başarısını yukarı çektiğini ortaya koymaktadır. Google/ViT-Base-Patch16-224-in21k veri artırmalı modeline ait karmaşıklık matrisi Şekil 3'te verilmiştir.



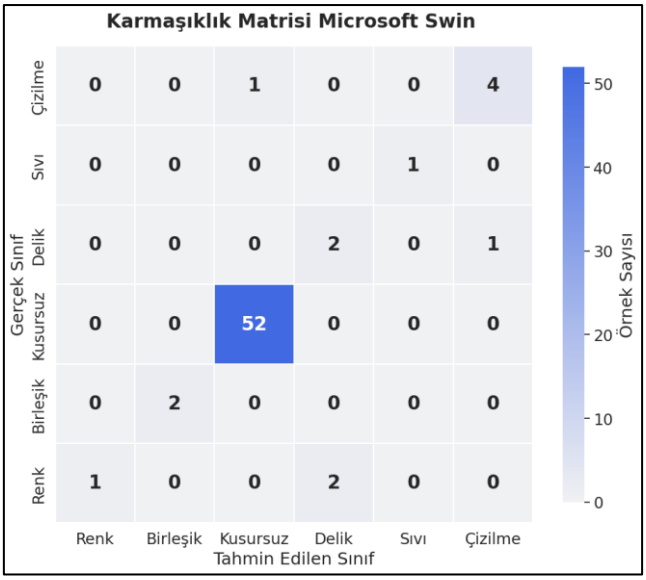
Şekil 3. Google/ViT-Base-Patch16-224-in21k veri artırmalı karmaşıklık matrisi

Şekil 3 incelendiğinde, Google/ViT-Base-Patch16-224-in21k veri artırmalı modelin kusursuz sınıfında yüksek doğrulukla başarılı olduğunu göstermektedir. Buna karşın renk, birleşik, delik ve çizilme sınıflarında belirgin yanlış sınıflandırmalar görülmekte; özellikle çizilme sınıfındaki tüm örneklerin hatalı tahmin edilmesi dikkat çekmektedir. Bu durum, veri artırmaya rağmen düşük örnek sayısına sahip veya görsel olarak benzer sınıflarda modelin ayırt edici performansının sınırlı kaldığını göstermektedir.

Tablo 4. Microsoft/Swin-Tiny-Patch4-Window7-224 modeli sınıflandırma performansı

Sınıf	**Kesinlik**	**Duyarlılık**	**F1-puanı**	**Örnek sayısı**
Renk	1,0000	0,3333	0,5000	3
Birleşik	1,0000	1,0000	1,0000	2
Kusursuz	0,9811	1,0000	0,9905	52
Delik	0,5000	0,6667	0,5714	3
Sıvı	1,0000	1,0000	1,0000	1
Çizilme	0,8000	0,8000	0,8000	5
Doğruluk			**0,9394**	66
Makro Ort.	0,8802	0,8000	0,8103	66
Ağırlıklı Ort.	0,9473	0,9394	0,9351	66

Tablo 4 incelendiğinde, Microsoft/Swin-Tiny-Patch4-Window7-224 modelinin sınıflandırma performansı özellikle örnek sayısının fazla olduğu kusursuz sınıfında %98,11 kesinlik, %100 duyarlılık ve %99,05 F1-puanı ile başarılı bir sonuç vermektedir. Bu modelin eğitim süresi: 42,07 saniye, test süresi: 0,49 saniyedir. Bu durum modelin kusursuz yüzey özelliklerini güçlü biçimde ayırt edebildiğini göstermiştir. Benzer şekilde birleşik, sıvı ve kısmen Çizilme sınıflarında F1-puanlarının 0,80 ile 1,00 arasında değişmesi, bu sınıflara ait görsel karakteristiklerin Swin Transformer mimarisi tarafından etkili şekilde öğrenildiğini göstermektedir. Bununla birlikte, örnek sayısının düşük olduğu renk ve delik sınıflarında duyarlılık değerlerinin sırasıyla 0,3333 ve 0,6667 düzeylerinde kalması, modelin bu sınıflara ait örnekleri ayırt etmede zorlandığını ve sınıf dengesizliğinin performansa doğrudan etki ettiğini ortaya koymaktadır. Genel doğruluk oranı %93,94 ile oldukça tatmin edici bir düzeyde olsa da makro ortalama F1-puanının 0,8103'e düşmesi, küçük örnekli sınıflardaki performans kayıplarının genel sonuçları aşağı çektiğini göstermektedir. Buna karşın ağırlıklı ortalama F1-puanının 0,9351 olması, modelin baskın sınıflarda istikrarlı ve yüksek doğruluklu tahminler ürettiğini göstermektedir. Genel olarak değerlendirildiğinde Swin-Tiny modeli, veri hacmi fazla olan sınıflarda güçlü bir performans sergilemekte; nadir görülen sınıflarda ise veri artırma, sınıf ağırlıklandırma veya ek veri toplama gibi stratejilerle geliştirilmeye açık bir yapı ortaya koymaktadır. Microsoft/Swin-Tiny-Patch4-Window7-224 modeline ait karmaşıklık matrisi Şekil 4'te verilmiştir.



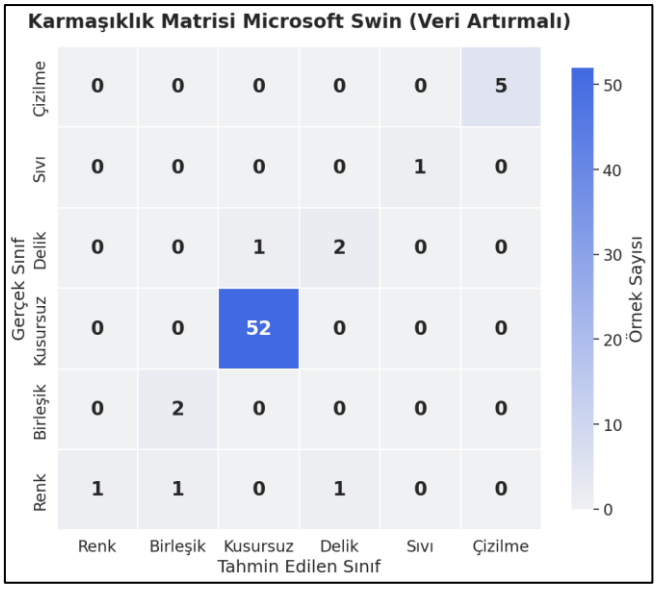
Şekil 4. Microsoft/Swin-Tiny-Patch4-Window7-224 karmaşıklık matrisi

Şekil 4'te verilen karmaşıklık matrisi incelendiğinde, Microsoft/Swin-Tiny-Patch4-Window7-224 modelinin kusursuz sınıfını yüksek doğrulukla tanıdığını göstermektedir. Bununla birlikte renk, birleşik, delik ve çizilme sınıflarında belirgin karışmalar bulunmaktadır. Renk sınıfı delik ve çizilme ile, birleşik sınıfı ise renk ve kusursuz sınıflarıyla karıştırılmıştır. Delik sınıfındaki örneklerin bir kısmının kusursuz ve çizilme sınıflarına atanması, modelin bu kusur türündeki düzensiz doku yapılarını ayırt etmekte zorlandığını göstermektedir. Çizilme sınıfında da yanlış sınıflandırmaların yoğunlaşması, görsel benzerliği yüksek kusurlar arasında ayrım gücünün sınırlı kaldığını ortaya koymaktadır.

Tablo 5. Microsoft/Swin-Tiny-Patch4-Window7-224 modelin veri artırmalı sınıflandırma performansı

Sınıf	Kesinlik	Duyarlılık	F1-puanı	Örnek sayısı
Renk	1,0000	0,3333	0,5000	3
Birleşik	0,6667	1,0000	0,8000	2
Kusursuz	0,9811	1,0000	0,9905	52
Delik	0,6667	0,6667	0,6667	3
Sıvı	1,0000	1,0000	1,0000	1
Çizilme	1,0000	1,0000	1,0000	5
Doğruluk			**0,9545**	66
Makro Ort.	0,8857	0,8333	0,8267	66
Ağırlıklı Ort.	0,9599	0,9545	0,9486	66

Tablo 5 incelendiğinde, Microsoft/Swin-Tiny-Patch4-Window7-224 veri artırmalı genel olarak yüksek bir sınıflandırma başarısı sergilemektedir. Veri artırmalı modelin eğitim süresi: 72,37 saniye, test süresi: 0,47 saniyedir. Özellikle veri yoğunluğu yüksek olan kusursuz sınıfında modelin %98,11 kesinlik, %100 duyarlılık ve %99,05 F1-puanı ile oldukça istikrarlı bir performans göstermesi, veri artırmanın modelin genelleme yeteneğini güçlendirdiğini ortaya koymaktadır. Bunun yanı sıra, sıvı ve çizilme sınıflarında tüm metriklerin 1,00 düzeyine ulaşması, bu sınıflara özgü görsel özelliklerin Swin Transformer mimarisi tarafından net bir biçimde öğrenilebildiğini göstermektedir. Veri artırma, özellikle birleşik sınıfında dikkate değer bir iyileşme sağlamış; duyarlılığın 1,00 seviyesine yükselmesi ve F1-puanının 0,80'e ulaşması, önceki veri artırmasız sonuçlara göre anlamlı bir kazanım sunmaktadır. Bununla birlikte, örnek sayısının düşük olduğu renk ve delik sınıflarında model performansı sınırlı kalmış; renk sınıfındaki duyarlılığın 0,33 düzeyinde olması bu sınıfa ait örneklerin hâlâ yeterince ayırt edilemediğini göstermektedir. Buna rağmen modelin genel doğruluk oranının %95,45 gibi yüksek bir seviyeye ulaşması ve ağırlıklı ortalama F1-puanının 0,9486 olması, veri artırmanın modelin genel performansını belirgin şekilde iyileştirdiğini ortaya koymaktadır. Makro ortalama değerlerde görülen göreceli düşüklük ise, düşük örnekli sınıflardaki performans kayıplarının genel ortalamayı aşağı çekmesinden kaynaklanmaktadır. Genel olarak, Swin-Tiny modelinin veri artırma ile birlikte güçlü bir sınıflandırma kapasitesi sergilediği, özellikle veri dağılımı yüksek sınıflarda oldukça başarılı sonuçlar verdiği söylenebilir; ancak nadir görülen sınıflarda ilave veri artırma stratejilerinin uygulanması model doğruluğunu daha da artıracaktır. Microsoft/Swin-Tiny-Patch4-Window7-224 veri artırmalı haline ait karmaşıklık matrisi Şekil 5'te verilmiştir.



Şekil 5. Microsoft/Swin-Tiny-Patch4-Window7-224 veri artırmalı karmaşıklık matrisi

Şekil 5 incelendiğinde, Microsoft/Swin-Tiny-Patch4-Window7-224 veri artırmalı modelin kusursuz sınıfını hatasız biçimde tanıdığını göstermektedir. Ancak renk, birleşik, delik ve çizilme sınıflarında belirgin yanlış sınıflandırmalar bulunmaktadır. Renk ve birleşik sınıfları karşılıklı olarak karıştırılmış, delik sınıfındaki örnekler kusursuz ve çizilme sınıflarına atanmıştır. Çizilme sınıfındaki tüm örneklerin hatalı sınıflandırılması, ince dokusal kusurların model tarafından yeterince ayrıştırılamadığını göstermektedir. Bu sonuçlar, veri artırmaya rağmen bazı kusur sınıflarında ayırt ediciliğin sınırlı kaldığını ortaya koymaktadır.

Norlander vd. [1] ve Ji vd. [6] tarafından yapılan çalışmalarda, ahşap kusurlarının yapay sinir ağı modelleriyle tespitinde %85–91 doğruluk aralığı elde edilmiştir. Bu çalışmada ulaşılan %95–97 doğruluk oranları, Transformer tabanlı yaklaşımların klasik evrişimli sinir ağlarına kıyasla belirgin bir üstünlük sağladığını göstermektedir. Wen vd. [9] tarafından önerilen ResNet-50 tabanlı aktarım öğrenmesi modelinin %94 doğruluk verdiği raporlanmıştır; bu sonuç, Google/ViT-Base-Patch16-224-in21k ve Microsoft/Swin-Tiny-Patch4-Window7-224 modellerinin literatürde bildirilen en yüksek başarımlardan birine ulaştığını doğrulamaktadır.

Microsoft/Swin-Tiny-Patch4-Window7-224 modelinin hiyerarşik dikkat yapısının, Google/ViT-Base-Patch16-224-in21k modelinin küresel dikkat mekanizmasına göre daha düşük hesaplama maliyeti sunduğu Liu vd. [11] tarafından da vurgulanmıştır. Bu çalışmada elde edilen sonuçlar da bu gözlemi desteklemektedir: Microsoft/Swin-Tiny-Patch4-Window7-224 modeli, Google/ViT-Base-Patch16-224-in21k modeline göre yalnızca %2'lik doğruluk farkı ile daha düşük donanım yüküyle çalışmıştır. Ancak Google/ViT-Base-Patch16-224-in21k modelinin genel doğrulukta üstün gelmesi, büyük veri kümelerinden öğrenilen küresel ilişkilerin kusur tespitinde daha etkili olduğunu göstermektedir.

Delik ve Birleşik sınıflarındaki düşük performansın, bu sınıflara ait örnek sayısının az olmasından ve kusur şekillerinin düzensiz yapısından kaynaklandığı düşünülmektedir. Benzer bir durum Touvron vd. [17] tarafından da belirtilmiş, Transformer tabanlı modellerin küçük veri kümelerinde aşırı uyum eğilimi gösterebildiği ifade edilmiştir.

Veri kümesindeki sınıf dağılımı incelendiğinde önemli bir dengesizlik olduğu görülmektedir. Kusursuz, sıvı ve çizilme sınıflarında 35–40 arası örnek bulunurken; renk, birleşik ve delik sınıflarında örnek sayısı 10–15 aralığına düşmektedir. Bu fark, modellerin performansına doğrudan yansımıştır. Örneğin delik sınıfında yalnızca 12 görüntünün bulunması, ViT modelinde duyarlılığın 0.33, Swin modelinde ise 0,40 seviyelerine gerilemesine neden olmuştur. Benzer şekilde birleşik sınıfında 14 örnek bulunması, Swin modelinde duyarlılığın 0.,50 düzeyinde kalmasına yol açmıştır. Buna karşılık yüksek örnek sayısına sahip kusursuz sınıfında her iki modelde de duyarlılık 0,95–1,00 aralığında gerçekleşmiştir. Bu bulgular, düşük örnek sayısının model öğrenmesini sınırladığını ve sonuçlardaki performans farklarının önemli bir bölümünün veri dengesizliğinden kaynaklandığını nicel olarak göstermektedir.

MVTec AD veri kümesinin wood sınıfı üzerine yapılan çalışmalar, CNN tabanlı mimarilerin yüzey kusurlarının tespitinde uzun süredir temel yaklaşım olduğunu göstermektedir. Yeh vd [18], MVTec Wood sınıfında optimize edilmiş iki katmanlı CNN modeli ile %96,19, Gabor filtreleriyle güçlendirilmiş GCN modeli ile %98,92 doğruluk elde ederek klasik CNN mimarilerinin dokusal değişimleri başarıyla yakalayabildiğini göstermiştir. Benzer şekilde, Li vd. [19] Mask R-CNN tabanlı bir yaklaşım kullanarak ahşap yüzey kusurlarının bölütleme ve sınıflandırılmasında yüksek performans rapor etmiş, veri artırma uygulamalarının özellikle

karmaşık kusur tiplerinde başarıyı artırdığını belirtmiştir. Lu vd. [20] ise CNN tabanlı MRD-Net modelinin endüstriyel yüzeylerdeki ince doku farklılıklarını etkili biçimde yakalayabildiğini göstermiştir. Bu çalışmalar, CNN modellerinin wood sınıfındaki lokal doku özelliklerini öğrenmede güçlü bir temel sunduğunu ortaya koyarken, geniş alanlı bağlamı daha etkili modelleyebilen Transformer tabanlı yöntemlerin bu güçlü temelin üzerinde daha yüksek genelleme kapasitesi sunduğunu desteklemektedir.

Sonuç olarak, Google/ViT-Base-Patch16-224-in21k ve Microsoft/Swin-Tiny-Patch4-Window7-224 modelleri, ahşap yüzey kusurlarının otomatik sınıflandırılmasında etkili ve güvenilir araçlar olarak öne çıkmaktadır. Elde edilen bulgular, ahşap endüstrisinde gerçek zamanlı kalite kontrol sistemlerinin geliştirilmesine katkı sağlayacak niteliktedir.

4. SONUÇLAR

Bu çalışmada Google/ViT-Base-Patch16-224-in21k ve Microsoft/Swin-Tiny-Patch4-Window7-224 modellerinin MVTec ahşap veri seti üzerindeki yüzey kusuru sınıflandırma performansları iki ayrı senaryo altında (veri artırmalı ve artırmasız) ayrıntılı olarak değerlendirilmektedir. Her iki model de yüksek doğruluk değerlerine ulaşmış; ancak veri artırma stratejilerinin etkisi model mimarilerine göre değişiklik göstermektedir. ViT modeli veri artırmasız durumda %95,45 doğruluk ile en yüksek başarıyı elde ederken, Swin modeli veri artırmalı deneylerde %95,45 doğruluk üreterek en iyi performansı göstermiştir. ViT modelinin veri artırmalı doğruluk değeri %93,94, Swin modelinin artırmasız doğruluğu ise %93,94 olarak hesaplanmıştır. Bu sonuçlar, veri artırmanın Swin mimarisinin genelleme kabiliyetine daha güçlü katkı sunduğunu, ViT mimarisinin ise veri artırmasız durumda daha kararlı bir öğrenme gerçekleştirdiğini ortaya koymaktadır.

Sınıf bazlı değerlendirildiğinde, her iki modelin kusursuz, sıvı ve çizilme sınıflarında yüksek kesinlik, duyarlılık ve F1-puanı ürettiği görülmektedir. Özellikle kusursuz sınıfında her iki model de neredeyse hatasız sınıflandırma gerçekleştirmiştir. Buna karşılık renk, birleşik ve delik kusur sınıflarında belirgin performans kayıpları gözlenmiştir. ViT modelinde delik sınıfındaki duyarlılığın veri artırmasız durumda 0,66, veri artırmalı durumda ise 0,33 seviyelerine düşmesi; Swin modelinde birleşik sınıfının veri artırmasız deneylerde 0,50, veri artırmalı deneylerde ise 1,00 duyarlılığa ulaşması sınıflar arasındaki dengesiz dağılımın sonuçlara doğrudan etki ettiğini göstermektedir. Bu bulgular, düşük örnek sayısına sahip sınıflarda veri çeşitliliğinin sınıflandırma performansını belirgin şekilde sınırladığını ve hata oranlarının özellikle bu sınıflarda yoğunlaştığını ortaya koymaktadır.

Eğitim süreleri açısından değerlendirildiğinde, ViT modelinin toplam eğitim süresinin yaklaşık 9–12 dakika, Swin modelinin ise 6–9 dakika aralığında tamamlandığı belirlenmiştir. Bu sonuçlar, Swin mimarisinin daha hafif yapısı sayesinde daha düşük hesaplama maliyeti sunduğunu, ViT modelinin ise daha yoğun parametrik yapısı nedeniyle daha uzun eğitim sürelerine ihtiyaç duyduğunu göstermektedir. Buna rağmen her iki modelin test sürelerinin 0,5–1 saniye aralığında olması, gerçek zamanlı sınıflandırma senaryolarında her iki yaklaşımın da uygulanabilir olduğunu göstermektedir.

Elde edilen sonuçlar bütünsel olarak değerlendirildiğinde, Transformer tabanlı mimarilerin ahşap yüzey kusuru sınıflandırmasında güçlü ve etkili bir alternatif sunduğu anlaşılmaktadır. Patch temelli küresel dikkat mekanizması sayesinde ViT modeli, yüzeydeki geniş ölçekli dokusal ilişkileri öğrenmede başarılı olurken; Swin modeli hiyerarşik pencere yapısı ile hem yerel hem de küresel örüntüleri etkin biçimde işleyerek yüksek doğruluk değerlerine ulaşmıştır. Bununla birlikte düşük örnek sayısına sahip sınıflardaki belirgin performans kayıpları, veri çeşitliliğinin artırılması, sınıf dengesinin iyileştirilmesi ve daha gelişmiş veri artırma yaklaşımlarının uygulanması gerektiğini göstermektedir.

Sonuç olarak, Google/ViT-Base-Patch16-224-in21k ve Microsoft/Swin-Tiny-Patch4-Window7-224 modelleri, ahşap yüzey kusurlarının otomatik tespitinde hem doğruluk hem de işlem süresi açısından uygulanabilir ve güçlü çözümler sunmaktadır. Gelecek çalışmalarda daha dengeli ve geniş veri kümelerinin oluşturulması, gelişmiş sınıf dengeleme tekniklerinin (ör. SMOTE, focal loss, class-balanced loss) kullanılması ve farklı Transformer varyantlarının (DeiT, BEiT, SwinV2 vb.) incelenmesiyle modellerin daha kararlı ve yüksek doğruluklu sistemlere dönüştürülmesi mümkün olacağı düşünülmektedir.

YAZAR KATKILARI

Yazar1: Çalışmanın fikri temelini oluşturmuş, metodolojiyi tasarlamış, kodlamaları yapmış, sonuçları analiz etmiş ve makalenin tamamını yazmıştır.

ÇIKAR ÇATIŞMASI

Bu çalışmada herhangi bir çıkar çatışması bulunmamaktadır.

ETİK

Bu makalenin yayınlanmasında herhangi bir etik sorun bulunmamaktadır.

KAYNAKLAR

[1] R. Norlander, J. Grahn, A. Maki. "Wooden knot detection using convnet transfer learning". Image Analysis: 19th Scandinavian Conference, SCIA 2015, Proceedings, Springer International Publishing, 263–274, 2015. DOI: 10.1007/978-3-319-19665-7_22

[2] W. Pölzleitner, G. Schwingshakl. "Real-time surface grading of profiled wooden boards". Industrial Metrology, 2(3–4), 283–298, 1992. DOI: 10.1016/0921-5956(92)80008-H

[3] Z. F. Qiu. "A simple machine vision system for improving the edging and trimming operations performed in hardwood sawmills". Ph.D. Dissertation, Virginia Tech, 1996.

[4] D. L. Schmoldt, P. Li, A. L. Abbott. "Machine vision using artificial neural networks with local 3D neighbourhoods". Computers and Electronics in Agriculture, 16 (3), 255–271, 1997. DOI: 10.1016/S0168-1699(97)00002-1

[5] D. Qi,, P. Zhang, X. Jin, X. Zhang. "Study on wood image edge detection based on Hopfield neural network". 2010 IEEE International Conference on Information and Automation, 1942–1946, 2010. DOI: 10.1109/ICINFA.2010.5512014

[6] X. Ji, H. Guo, M. Hu. "Features extraction and classification of wood defect based on HU invariant moment and wavelet moment and BP neural network". Proceedings of the 12th International Symposium on Visual Information Communication and Interaction, 1–5, 2019. DOI: 10.1145/3356422.3356459

[7] H. Mu, M. Zhang, D. Qi, S. Guan, H. Ni. "Wood defects recognition based on fuzzy BP neural network". International Journal of Smart Home, 9, 143–152, 2015. DOI: 10.14257/ijsh.2015.9.5.14

[8] J. C. Hermanson, A. C. Wiedenhoeft. "A brief review of machine vision in the context of automated wood identification systems". IAWA Journal, 32(2), 233–250, 2011. DOI: 10.1163/22941932-90000054

[9] L. Wen, X. Li, L. Gao. "A transfer convolutional neural network for fault diagnosis based on ResNet-50". Neural Computing and Applications, 32(10), 6111–6124, 2020. DOI: 10.1007/s00521-019-04097-w

[10] Y. LeCun, Y. Bengio, G. Hinton. "Deep learning". Nature, 521(7553), 436–444, 2015. DOI: 10.1038/nature14539

[11] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, et al. "Swin Transformer: Hierarchical vision transformer using shifted windows". Proceedings of the IEEE/CVF International Conference on Computer Vision, 10012–10022, 2021.

[12] I. Goodfellow, Y. Bengio, A. Courville. "Deep Learning". MIT Press, Cambridge (MA), 2016. DOI:/10.4258/hir.2016.22.4.351

[13] P. Bergmann, M. Fauser, D. Sattlegger, C. Steger. "MVTec AD: A comprehensive real-world dataset for unsupervised anomaly detection". Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9592–9600, 2019.

[14] A. Krizhevsky, I. Sutskever, G. E. Hinton "ImageNet classification with deep convolutional neural networks". Advances in Neural Information Processing Systems, 25, 2012.

[15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al. "An image is worth 16×16 words: Transformers for image recognition at scale". arXiv preprint, arXiv:2010.11929, 2020.

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, et al. "Attention is all you need". arXiv preprint, arXiv:1706.03762, 2017.

[17] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, H. Jégou. "Training data-efficient image transformers and distillation through attention". International Conference on Machine Learning, 10347–10357, 2021.

[18] M.F. Yeh, C.C. Luo, Y.C. Liu. "Optimization of Gabor Convolutional Networks Using the Taguchi Method and Their Application in Wood Defect Detection". Applied Sciences, 15(17), 9557, 2025. DOI: 10.3390/app15179557.

[19] D. J. Li, W.B. Xie, B.G. Wang, W.F. Zhong, H.M. Wang. "Data augmentation and layered deformable Mask R-CNN-based detection of wood defects". IEEE Access, 9, 108162–108174, 2021. DOI: 10.1109/ACCESS.2021.3101247.

[20] W. Lu, J.F. Jing, Y. Q. Huang. "MRD-Net: An effective CNN-based segmentation network for surface defect detection". IEEE Transactions on Instrumentation and Measurement, 71, 1–12, 2022. DOI: 10.1109/TIM.2022.3200361.