Sinopfbd, 2026; 11(1): 134-157 https://doi.org/10.33484/sinopfbd.1789939

Sinop Üniversitesi Fen Bilimleri Dergisi
Sinop Uni J Nat Sci
E-ISSN: 2564-7873

https://dergipark.org.tr/tr/pub/sinopfbd

Impact of Image Resolution on Deep Learning-Based Classification of Elazığ Cherry Marble

Murat YAVUZ[1] and İbrahim TÜRKOĞLU[2]

How to cite: Yavuz, M., & Türkoğlu, İ. (2026). Impact of image resolution on deep learning-based classification of Elazığ Cherry marble. *Sinop Üniversitesi Fen Bilimleri Dergisi, 11*(1), 134-157. https://doi.org/10.33484/sinopfbd.1789939

Open Access

Research Article

Corresponding Author
Murat YAVUZ
phd.myavuz@gmail.com

ORCID of the Author's
M.Y: 0000-0002-9896-9383
İ.T: 0000-0003-4938-4167

Geliş Tarihi: 23.10.2025
Kabul Tarihi: 21.01.2026

Abstract

Natural stones have played a significant role throughout human history, valued for their aesthetic, cultural and economic importance in applications ranging from monumental architecture to contemporary design. Among them, Elazığ Cherry marble, quarried exclusively in the Alacakaya district of Elazığ, Türkiye, stands out as a unique and prestigious natural resource, renowned for its deep cherry red color, distinctive vein structure and polished brilliance. This study systematically investigates the impact of image resolution on deep learning architectures for visual classification through experimental analyses conducted on Elazığ Cherry marble. A total of 2551 images were resampled into multiple resolutions ranging from 96x96 to 1024x1024 pixels and evaluated using three pre-trained architectures: ResNet50, Darknet53 and DenseNetV2. The findings demonstrate that low resolutions yielded accuracies within the 90–93% range, while intermediate resolutions (224x224 – 299x299) provided significant improvements, offering the optimal balance between accuracy and computational efficiency. At higher resolutions, performance gains became marginal; however, ResNet50 still achieved the highest accuracy of 96.10% at 1024x1024 resolution. The results highlight image resolution as a critical factor influencing not only visual quality but also classification accuracy, computational cost and model robustness against overfitting. Accordingly, this study contributes novel insights into resolution–architecture interactions and provides practical implications for the natural stone industry, delivering digital, objective and reproducible alternatives to traditional manual quality control practices.

Keywords: Image resolution, Elazığ Cherry marble, deep learning, ResNet50, image processing

Görüntü Çözünürlüğünün Elazığ Vişne Mermerinin Derin Öğrenme Tabanlı Sınıflandırılmasına Etkisi

[1]Elazığ Özel Eğitim Meslek Lisesi, Elazığ, Türkiye

[2]Fırat Üniversitesi, Teknoloji Fakültesi, Yazılım Mühendisliği Bölümü, Elazığ, Türkiye

Özet

Doğal taşlar, insanlık tarihi boyunca estetik, kültürel ve ekonomik değerleriyle önemli bir rol oynamış; anıtsal mimariden çağdaş tasarıma kadar uzanan geniş bir kullanım alanı bulmuştur. Bu taşlar arasında, yalnızca Türkiye'nin Elazığ iline bağlı Alacakaya ilçesinden çıkarılan Elazığ Vişne mermeri, derin vişne kırmızısı rengi, belirgin damarlı yapısı ve parlak yüzey dokusuyla dünya çapında benzersiz ve saygın bir doğal kaynak olarak öne çıkmaktadır. Bu çalışma, Elazığ Vişne mermeri üzerinde gerçekleştirilen deneysel analizler aracılığıyla, görüntü çözünürlüğünün derin öğrenme tabanlı görsel sınıflandırma modelleri

üzerindeki etkisini sistematik bir biçimde incelemektedir. Toplam 2551 görüntü, 96x96 ile 1024x1024 piksel arasında değişen çoklu çözünürlük seviyelerine yeniden örneklenmiş ve üç farklı önceden eğitilmiş mimari (ResNet50, Darknet53 ve DenseNetV2) kullanılarak değerlendirilmiştir. Bulgular, düşük çözünürlüklerde doğruluk oranlarının %90–93 aralığında kaldığını, buna karşın orta düzey çözünürlüklerin (224x224 – 299x299) belirgin bir iyileşme sağladığını ve doğruluk ile hesaplama verimliliği arasında en uygun dengeyi sunduğunu göstermektedir. Daha yüksek çözünürlüklerde performans artışı sınırlı kalmakla birlikte, ResNet50 modeli 1024x1024 çözünürlükte %96,1 ile en yüksek doğruluk oranına ulaşmıştır. Sonuçlar, görüntü çözünürlüğünün yalnızca görsel kaliteyi değil, aynı zamanda sınıflandırma doğruluğunu, hesaplama maliyetini ve modelin aşırı uyum (*overfitting*) karşısındaki dayanıklılığını etkileyen kritik bir parametre olduğunu ortaya koymaktadır. Bu doğrultuda, çalışma çözünürlük–mimari etkileşimlerine ilişkin yeni bulgular sunmakta ve doğal taş endüstrisi için dijital, nesnel ve tekrarlanabilir bir kalite kontrol alternatifi geliştirmeye yönelik pratik katkılar sağlamaktadır.

Anahtar Kelimeler : Görüntü çözünürlüğü, Elazığ Vişne mermeri, derin öğrenme, ResNet50, görüntü işleme

Introduction

Natural stones have long been regarded as materials of both functional and symbolic significance, shaping human civilization from ancient times to the present. Their enduring role spans monumental architecture, religious structures, sculptures, and urban landscapes reflecting a legacy that combines durability, aesthetic diversity, and cultural identity. Marble, in particular, has been celebrated not only as a building material but also as a medium for artistic expression and luxury design, conferring prestige and aesthetic refinement in diverse applications ranging from public monuments to interior decorations. In contemporary contexts, natural stones remain highly valued for their unique textures, color variations and timeless appeal, which cannot be fully replicated by synthetic alternatives [1]. Türkiye occupies a pivotal role in the global natural stone market due to its extensive geological diversity and abundant reserves. The country's marble deposits, distributed across regions such as Afyon, Burdur, Bilecik, Denizli, Muğla, and Elazığ, exhibit remarkable variability in tonal richness, veining patterns, mechanical durability and surface properties. This diversity has positioned Türkiye not only as one of the leading producers but also as a supplier of distinctive, high value natural stones in the international arena. Within this portfolio, varieties such as Afyon White, Marmara Marble, Burdur Beige, and Muğla White hold considerable cultural and commercial significance. However, Elazığ Cherry marble has emerged as one of the most prestigious and distinctive stones, renowned for its deep cherry-red hues and striking vein structures [2]. Quarried exclusively in the Alacakaya district of Elazığ, Elazığ Cherry marble represents a geologically limited yet globally recognized resource. Its unique chromatic intensity, pronounced veining, and polished brilliance have made it highly sought after in luxury interior projects, boutique architectural works and high-end design applications. Beyond its aesthetic prominence, the stone plays a strategic socioeconomic role: quarrying and processing activities provide vital employment for the local community, while marble exports account for approximately 75% of Elazığ's total foreign trade revenues. Türkiye, as a whole, exports approximately 6.1 million tons of marble annually, generating revenues of around $2 billion, with Elazığ Cherry marble making a significant contribution to this figure. Its prestige in global markets, particularly in destinations such as China and India, underscores its dual value as both a cultural symbol and an economic driver [3]. Against this background, image resolution emerges as a critical determinant in the automated classification of natural stones. In digital image analysis, resolution directly influences the extent to which microstructural features such as vein patterns, tonal gradients and surface roughness are preserved and accurately represented. Low-resolution images

often obscure these details, resulting in limited feature extraction and decreased classification accuracy. Conversely, higher resolutions enhance visual fidelity but impose greater computational demands, leading to a trade-off between accuracy and efficiency. While resolution has been widely recognized as an important parameter in fields such as remote sensing, facial recognition and medical imaging, its role in natural stone classification has received comparatively little systematic attention in the literature. Previous studies have explored the influence of image resolution primarily in domains such as remote sensing and medical diagnostics; however, comparable analyses in the context of natural stone or marble classification remain extremely limited. This research addresses this gap by examining the impact of image resolution on deep learning–based classification models within the specific context of Elazığ Cherry marble. Leveraging its microstructural complexity, the study investigates how variations in resolution affect not only classification accuracy but also broader aspects such as, model stability, feature extraction capacity, and computational efficiency. A dataset of Elazığ Cherry marble images was prepared at multiple resolution levels ranging from 96x96 pixels to 1024x1024 pixels and three state of the art deep learning architectures ResNet50, Darknet53, and DenseNetV2 were systematically evaluated. These architectures were employed in their pretrained forms on large-scale image datasets, ensuring robust feature extraction capabilities suitable for fine-grained marble texture analysis. These architectures were selected to represent different depths and feature extraction capabilities, enabling comparative insights into their sensitivities to resolution variations. Special attention was given to preventing overfitting during training, emphasizing model generalization and reproducibility. The findings of this study highlight the need to optimize resolution not in isolation, but in conjunction with model architecture, in order to achieve an effective balance between classification performance and computational cost. By addressing this overlooked dimension, the study not only contributes to the academic discourse on artificial intelligence and computer vision but also provides practical implications for the natural stone industry, offering pathways toward more objective, efficient and reproducible quality control systems that can complement or replace traditional manual methods.

Literature Review

The effect of image resolution on the performance of classification systems has become increasingly prominent topic in recent research across various domains. One notable study by Memiş and Karabiber [4] investigated the influence of resolution on the effectiveness of appearance-based face recognition methods using mobile-acquired images. Utilizing the MOBIO dataset, facial images were resized into five different resolutions ranging from 12x12 to 192x192 pixels and the performances of Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Local Binary Pattern Histograms (LBPH) were compared based on accuracy, sharpness, sensitivity, and F-measure metrics. The results indicated that LDA performed more effectively at lower resolutions, while LBPH demonstrated clear superiority at higher resolutions, achieving an accuracy of 99.39% and an F-measure of 85.80% at 96x96 resolution. These findings highlight the significance of resolution in recognition performance and support the rationale behind the experimental design of this study. Cerit et al. [5] analyzed the impact of image resolution on the performance of automatic gender and age classification using facial images from the MORPH dataset, containing over 53000 images of individuals aged 16–55. The study evaluated ten different resolutions ranging from 2x1 to 329x264 pixels, using three classical machine learning classifiers: k-Nearest Neighbors (k-NN), Support Vector Machines (SVM) and Random Forests (RF). Experimental results (Tables III–VI, pp. 3–4) revealed that gender classification accuracy remained consistently high across all resolutions, achieving 100% accuracy at 16x13 and higher, while even extremely low-resolution images (2x1) still yielded meaningful gender cues (about 84%). Conversely, age classification showed a more complex relationship with resolution: accuracy improved sharply from ~34% at 2x1 to ~76% at 45x36, but plateaued thereafter, showing diminishing returns beyond 90x72.

The study concluded that while gender can be robustly inferred from very low-resolution images due to global facial structure, age classification requires moderate to high resolutions (≥45x36) to capture texture and wrinkle information effectively. The authors also noted that increases in resolution significantly extended computation time without proportional accuracy gains, suggesting that 45x36 pixels represents an optimal balance for age and gender recognition on MORPH. Kannojia and Jaiswal [6] performed an experimental investigation on how image resolution variation influences the classification performance of Convolutional Neural Networks (CNNs) using the MNIST and CIFAR-10 datasets. Two training–testing strategies were proposed: TOTV (Train on Original, Test on Varying) and TVTV (Train and Test on Varying). Each dataset was rescaled into progressively lower resolutions MNIST (28x28, 21x21, 14x14, 7x7) and CIFAR-10 (32x32, 24x24, 16x16, 8x8) to quantify performance degradation. For MNIST, the CNN maintained high accuracy down to 14x14 pixels (TOTV: 97.7%, TVTV: 98.5%) before dropping sharply at 7x7 (TOTV: 67.9%, TVTV: 77.7%). Conversely, the more visually complex CIFAR-10 dataset suffered greater performance loss, declining from 87.5% at 32x32 to 18.5% at 8x8 under TOTV conditions. Precision values remained slightly higher than F1-scores across both datasets, implying consistent class relevance but reduced overall accuracy. Comparative plots (Figures 5–6, pp. 455) show that TVTV models are more robust to low-resolution input than TOTV models. The authors concluded that classification performance depends strongly on visual complexity and that reducing resolution significantly impairs CNN accuracy, particularly for datasets with rich spatial information such as CIFAR-10. Qin and Beckingham [7] investigated how variations in image resolution affect the quantification of mineralogical and pore-scale characteristics in geological samples, focusing on sandstone specimens from the Paluxy formation. Using scanning electron microscopy (SEM) and X-ray computed tomography (CT), images were captured across resolutions ranging from 0.34 μm to 5.71 μm. Their results demonstrated that while overall mineral abundance and porosity remained largely stable across resolutions (variation <2.5%), significant differences appeared in mineral accessibility and effective surface area. In particular, the accessibility of smectite/illite decreased from ~54% to 30%, whereas quartz accessibility increased from ~34% to 59% as resolution decreased. This led to an overestimation of quartz surface areas at lower resolutions. The study concluded that resolutions below 1 μm are critical for accurately quantifying mineral surface areas, especially when clay coatings or microstructural features dominate the sample. Although higher resolutions increase computational cost, they are essential for reliable reactive transport modeling and accurate simulation of mineral dissolution and permeability evolution in geochemical systems. Sabottke and Spieler [8] provided one of the earliest large-scale investigations into how image resolution influences CNN performance for medical imaging, focusing on the NIH ChestX-ray14 dataset containing 112120 images from 30805 patients. Using ResNet-34 and DenseNet-121 architectures, the study evaluated nine input resolutions ranging from 32x32 to 600x600 pixels, across eight diagnostic labels: emphysema, cardiomegaly, hernia, atelectasis, edema, effusion, mass and nodule. Results showed that most diagnostic labels achieved peak AUCs between 256x256 and 448x448 pixels, while lower resolutions (32x32, 64x64) substantially degraded performance. Notably, fine-grained pathologies such as pulmonary nodules and emphysema benefited significantly from higher resolutions (AUC up to 0.935 at 320x320), whereas larger structural findings like thoracic masses plateaued earlier (AUC ≈ 0.886). Statistical analysis confirmed that optimal resolutions differed across diagnoses ($p <$ 0.05), highlighting that image size should be task-dependent. Figure 2 (p. 4) visually illustrated this non-linear resolution–performance curve, showing diminishing returns beyond 320x320 pixels. The authors concluded that subtle pathologies require higher-resolution inputs, but training efficiency and GPU limitations constrain the achievable batch size at these scales. This study remains a foundational benchmark for understanding the resolution–accuracy trade-off in radiographic deep learning. Thambawita et al. [9] conducted one of the most extensive studies on the relationship between image

resolution and CNN performance in gastrointestinal (GI) endoscopy classification using the HyperKvasir dataset, which includes 10662 endoscopic images across 23 clinical findings. Two architectures ResNet-152 and DenseNet-161, both pretrained on ImageNet were trained and evaluated at five resolutions: 32x32, 64x64, 128x128, 256x256 and 512x512 pixels. The results revealed a clear monotonic improvement in performance with increasing resolution. Specifically, the Matthews Correlation Coefficient (MCC) improved from 0.8241 (32x32) to 0.9002 (512x512) for DenseNet-161, with similar gains observed in macro F1-score, precision and sensitivity (see Table 2, p.5). Importantly, the authors found that downscaling images causes a greater loss of accuracy than training on high-resolution data, indicating that data collection at higher resolutions is preferable, as upscaling cannot recover lost detail. Interestingly, inference time increased only marginally from 19.8 ms at 32x32 to 20.4 ms at 512x512 (Table 3, p.6) suggesting that computational overhead does not scale linearly with resolution. The study concluded that higher image resolutions consistently enhance CNN performance in endoscopic classification and emphasized the need for standardized resolution guidelines in medical image AI research. Ivanescu [10] conducted a comprehensive statistical analysis to evaluate how image preprocessing and resolution levels influence the performance and computational efficiency of deep learning networks applied to medical imaging. The study utilized three datasets lung cancer, colon cancer and fetal brain ultrasound images trained using the ResNet50 architecture under different conditions: color vs. grayscale, random vs. ImageNet-initialized weights and image resolutions of 128x128, 224x224 and 512x512 pixels. Across 100 independent runs with 10-fold cross-validation, the model achieved its highest accuracy (81.9%) using 512x512 grayscale images pretrained on ImageNet, while the fastest training time occurred at 128x128 resolution, reducing computational time by up to 7000 seconds compared to higher resolutions. Statistical tests (ANOVA and Tukey HSD) confirmed that differences in model performance were significant ($p < 0.05$) depending on both resolution and preprocessing strategy. However, accuracy losses from downscaling were minimal (less than 3%), whereas computational savings were substantial. The study concluded that moderate resolutions (224–256 px), combined with grayscale conversion and pretrained weights, yield the most efficient trade-off between accuracy and processing cost for medical image classification tasks. Similarly, the impact of image resolution on the performance of generative neural networks (GANs) was examined by Emekligil and Öksüz [11] in the context of game character generation. Their experiments, conducted at 32x32, 64x64 and 128x128 resolutions, revealed that as resolution increased, training became more challenging and transfer learning methods were more effective for high-resolution image synthesis. Evaluated using the Fréchet Inception Distance (FID) metric, the SNGAN model achieved better results at lower resolutions, while the BigGAN model trained with transfer learning outperformed its from-scratch counterpart at 128x128 resolution. This study provides valuable insight into how resolution affects training dynamics and output quality in generative models. In the domain of change detection, working with images of different resolutions often necessitates pre-processing steps such as super-resolution, which can lead to error accumulation. Addressing this challenge, Zhu et al. [12] proposed a Transformer-based architecture called the Multi-scale Transformer Network (MTN), designed to perform multi-scale feature extraction directly from images of varying resolutions. Instead of aligning input images beforehand, the MTN model extracts features from both high- and low-resolution inputs and matches them at the feature level to detect changes. Experiments on LEVIR-CD and Google Earth datasets demonstrated the model's effectiveness, particularly in delineating object boundaries. By eliminating the need for explicit resolution alignment, MTN presents a notable innovation in the field of change detection. Thon et al. [13] examined the influence of image resolution and attention head count on the performance of the *Convolutional Vision Transformer (ConViT)* for COVID-19 diagnosis using lung CT scans. The study used the *COVID-CTset* dataset from Iran, containing 2282 COVID-19 and 9776 normal CT images. Three input resolutions 128x128, 224x224 and 512x512 pixels were tested alongside

three attention-head configurations (4, 9, and 16). The results showed that the highest accuracy of 98.01% was achieved with 128x128 resolution and 16 attention heads, accompanied by a sensitivity of 90.83%, specificity of 99.69% and F1-score of 94.55%. Interestingly, increasing the image resolution to 224x224 and 512x512 led to a *decrease in accuracy*, attributed to excessive patch fragmentation that caused a loss of global spatial information. The study concluded that the ConViT architecture benefits from smaller input sizes that balance local and global features within each patch. Compared with standard CNN and ViT models trained on the same dataset, ConViT outperformed ViT (95.36%) and closely matched ResNet50 + Feature Pyramid (98.49%), despite using a smaller dataset and no pretraining. Overall, Thon et al. demonstrated that ConViT offers a computationally efficient and highly accurate framework for medical imaging tasks, particularly at lower resolutions (128x128) where global–local feature balance is optimal. Wollek et al. [14] systematically analyzed the effect of image resolution on chest X-ray classification performance using the ChestX-ray14 dataset, consisting of 112120 radiographs across 14 disease classes. The authors employed a DenseNet-121 architecture pretrained on ImageNet and trained it at five input resolutions: 64x64, 128x128, 256x256, 512x512, and 1024x1024 pixels. Quantitative results demonstrated a consistent performance gain with higher resolutions the mean AUC increased from 77.5% at 64x64 to 84.2% at 1024x1024. While the 256x256 configuration achieved 82.7%, a slight dip was observed at 512x512 before rising again at 1024x1024. Class-specific AUCs revealed that higher resolutions were especially beneficial for detecting small pathologies such as, nodules and atelectasis, where low-resolution models failed to localize key regions. Using Grad-CAM saliency maps, the study found that low-resolution models relied on spurious features, while higher-resolution inputs restored attention to clinically relevant areas. Despite increased computational demands, Wollek et al. [14] concluded that training with 1024x1024 resolution produces superior and more trustworthy predictions, particularly for fine-grained chest abnormalities, establishing that image resolution is a critical hyperparameter for reliable medical image classification. Rukundo [15] investigated how image size influences deep learning performance in the segmentation and quantification of myocardial infarction (MI) using Late Gadolinium Enhancement (LGE) MRI images. The study applied a U-Net architecture trained on datasets of 128x128 and 256x256 pixel resolutions, employing nearest-neighbor, bicubic, and Lanczos3 interpolation methods to resize images. A novel five-step preprocessing strategy was introduced to remove extra class labels in interpolated ground-truth segmentation masks. Experimental results demonstrated that larger image sizes (256x256) consistently improved segmentation accuracy and quantification closeness to manual (semi-automatic) expert results. Specifically, the model trained on higher-resolution images achieved validation and global accuracies up to 0.9919 and 0.9918, respectively, slightly higher than those obtained with smaller images (0.9908). Furthermore, quantification results based on the larger LGE-MRI dataset were 55.5% closer to manual measurements, compared to 22.2% for the smaller dataset, underscoring the significant performance gains achieved with higher-resolution training data. The findings confirm that increasing image resolution enhances both segmentation and clinical quantification outcomes, with minimal risk of overfitting and a reasonable increase in training time. Haque et al. [16] investigated the impact of image resolution on deep learning performance for automated chest X-ray (CXR) classification, using the large-scale MIMIC-CXR-JPG dataset (377110 images, 14 disease labels). The study systematically compared four resolutions 256x256, 512x512, 1024x1024 and 2048x2048 pixels using DenseNet121 and EfficientNet-B4 architectures pretrained on ImageNet. Results demonstrated that optimal resolution varies by pathology: conditions requiring large contextual understanding (e.g., cardiomegaly) performed best at lower resolutions, while fine-detail tasks (e.g., pneumothorax detection) benefited from higher resolutions. The findings were linked to the effective receptive field (ERF) which shrinks relative to image size limiting performance for large anatomical features at high resolutions. A stacked multi-scale ensemble model integrating all four resolutions outperformed single-scale networks across all 14

findings, achieving an average AUC improvement of 4.3% over the best individual model. Visual analysis with Grad-CAM further confirmed the importance of resolution-dependent receptive fields in attention localization. The authors conclude that multi-scale feature extraction rather than simply maximizing image resolution is essential for robust, interpretable deep learning in radiographic diagnosis. Liu et al. [17] proposed Multi-Scale Patch Embedding (MSPE), a lightweight framework that allows Vision Transformers (ViTs) to process images of any resolution or aspect ratio without retraining or resizing. Traditional ViTs typically require inputs resized to a fixed dimension (e.g., 224x224), which severely limits their adaptability to real-world images of varying resolutions. MSPE replaces the standard patch embedding layer with a set of adaptive, learnable convolution kernels that dynamically adjust to input size. This innovation enables ViTs to maintain stable accuracy across a wide range of resolutions (from 28x28 to 896x896) with minimal computational overhead. In experiments on ImageNet-1K, MSPE improved low-resolution Top-1 accuracy from 8.52% (vanilla ViT) to 56.41%, surpassing FlexiViT and ResFormer baselines, while maintaining parity at higher resolutions (≈85.1%). The method also achieved superior results in semantic segmentation (ADE20K, Cityscapes) and object detection (COCO2017), improving mIOU and AP metrics by 3–6% over existing approaches. Notably, MSPE required only five epochs of fine-tuning, far less than competing methods (e.g., 200+ epochs for ResFormer). The study demonstrates that optimizing the patch embedding layer is key to achieving multi-resolution robustness in ViT-based models, providing a general, computationally efficient solution for real-world vision applications. Benyahia and Benammar [18] conducted an extensive analysis of the influence of image resolution on the performance of convolutional neural networks (CNNs) in Alzheimer's disease classification. Using the ResNet50 architecture, MRI brain scans were resized to 64x64, 128x128, and 256x256 pixels to assess accuracy variations. The results indicated a steady increase in classification accuracy from 73% to 85% as resolution increased, demonstrating that higher-resolution inputs enable the model to capture finer structural and morphological brain details. However, the authors emphasized that larger image dimensions significantly extended training time and computational cost. They concluded that 256x256 pixels provided an optimal trade-off between model accuracy and efficiency in medical image classification. Bacon et al. [19] explored how input image resolution affects the performance of deep learning models for animal species recognition. Convolutional neural networks were trained and evaluated at 128x128 and 224x224 resolutions. The results showed that increasing the resolution from 128x128 to 224x224 improved model accuracy from 88% to 94%, primarily due to the model's enhanced ability to capture fine-grained texture and shape features. However, further increases in resolution yielded diminishing returns, where the marginal accuracy gain did not justify the increased computational cost. The study concluded that for natural scene and animal datasets, 224x224 pixels represented an optimal input size, balancing classification accuracy and processing efficiency. Liu et al. [20] presented a systematic analysis of how image resolution and network depth interact to influence classification accuracy in deep learning models, specifically focusing on ResNet18 applied to biomedical datasets from the MedMNIST library. The study tested four input resolutions 28x28, 64x64, 128x128, and 224x224 pixels and four network depths (Block 1–4), across eight datasets, including BloodMNIST, BreastMNIST, DermaMNIST and PneumoniaMNIST. Results demonstrated that the widely adopted 224x224 resolution is not always optimal: in many cases, lower resolutions such as 64x64 or 128x128 achieved statistically significant improvements in accuracy ($p < 0.05$). For example, PneumoniaMNIST reached its highest accuracy at 64x64 with Block 2–3, while OrganSMNIST performed best at 128x128 with Block 4, achieving an average test accuracy above 0.91 (Table 7, p. 355). Conversely, DermaMNIST achieved higher performance with shallower depths (Block 2), indicating that resolution–depth combinations yield nonlinear and dataset-dependent effects. Feature map visualizations revealed that the top-performing configurations produced diverse feature sizes, suggesting multiple ways of encoding salient structures.

The authors concluded that reducing both depth and resolution can achieve comparable or even superior accuracy while lowering computational costs, challenging the conventional assumption that deeper and higher-resolution models always perform better. Przybyła-Kasperek et al. [21] presented a comprehensive evaluation of how image resolution and preprocessing strategies affect the performance of a Multi-Layer Perceptron (MLP) model in a controlled computer vision environment. The authors created a novel dataset of 3,000 images from the *Phasmophobia* video game, capturing different illumination levels (0–5 lights active) under varied lighting conditions. The study systematically examined color schemes (RGB, CMYK, HSV, Lab, YCbCr, Grayscale, etc.), normalization methods (Min-Max, Z-Score, Robust Scaling, FFT, DCT, Histogram Equalization, CLAHE), and three resolutions (64x64, 128x128, 256x256) across 486 model configurations. Results revealed that lower resolutions (64x64) often yielded the highest accuracy (up to 0.969 with CMYK + standard normalization), while larger resolutions led to significant increases in computational time (up to 20x slower) without consistent accuracy gains. Among normalization methods, Standard and Robust Scaling produced the best and most stable results, whereas FFT and DCT transformations drastically reduced performance at low resolutions. The findings challenge the assumption that higher resolutions always improve model accuracy, suggesting that optimal performance is strongly dependent on preprocessing configuration rather than image size alone. The study concluded that resolution–preprocessing interactions can outweigh the benefits of larger image dimensions, emphasizing the importance of methodological tuning over brute-force resolution scaling. Alenazi and Alkhodair [22] performed a comparative evaluation of deep neural network (DNN) architectures trained on medical images of varying resolutions. The study employed MRI datasets resized into 128x128, 256x256 and 512x512 pixels, to analyze how input size impacts accuracy and computational efficiency. The findings revealed a consistent improvement in performance, with accuracy increasing from 79% to 88% as resolution grew. Moreover, high-resolution data enhanced the model's robustness to noise and significantly improved the detection of small anatomical structures. Nevertheless, the computational cost increased by approximately 2.4 times when moving from 128x128 to 512x512 pixels. The authors concluded that while 512x512 resolution yields the highest accuracy, the 256x256 configuration offers the best cost–performance balance for practical clinical applications. A review of the existing literature shows a growing adoption of image processing and machine learning approaches in the classification of natural stones. However, the resolution variable is often assumed to be constant and its influence on model performance is rarely examined systematically. This assumption may limit both the accuracy and the practical applicability of models, especially for materials like Elazığ Cherry marble that possess high microstructural complexity and aesthetic significance. In such contexts, resolution directly affects classification performance and overlooking this variable could result in suboptimal outcomes. Therefore, a holistic approach that evaluates the effect of resolution not only on classification accuracy but also on processing time and computational resource requirements would offer a meaningful contribution to the literature. The primary aim of this study is to fill this gap by analyzing the performance of deep learning-based classification models trained on Elazığ Cherry marble images at various resolutions, providing practical insights for both academic research and industrial applications. Table 1 summarizes the key features of the studies related to the research topic.

Table 1. *Summary of related work*

Authors	Year	Domain / Application	Methods / Models	Dataset	Key Findings
Memiş & Karabiber	2016	Face recognition in mobile environments	PCA, LDA, LBPH	MOBIO	LDA performed better at low resolutions; LBPH achieved 99.39% accuracy and 85.80% F-score at 96x96.
Cerit et al.	2016	Face gender and age recognition	k-NN, SVM, Random Forest	MORPH	Gender classification achieved 100% ≥16x13; age classification optimal at 45x36 (76%); increasing resolution offered minimal additional accuracy.
Kannojia & Jaiswal	2018	General image classification	CNN	MNIST, CIFAR-10	Accuracy remained >97% for MNIST down to 14x14; CIFAR-10 dropped from 87.5%→18.5% at 8x8; complex datasets more sensitive to resolution loss.
Qin & Beckingham	2019	Geological pore-scale mineral analysis	SEM & CT, Image segmentation	Paluxy sandstone	Resolution <1 μm critical for accurate mineral surface quantification; lower resolutions overestimated quartz surface by ~30%.
Sabottke & Spieler	2020	Chest X-ray disease detection	ResNet34, DenseNet121	ChestX-ray14	Optimal AUC between 256x256–448x448; fine-grained pathologies benefited most from higher resolution (AUC up to 0.935).
Thambawita et al.	2021	Gastrointestinal endoscopy classification	ResNet-152, DenseNet-161	HyperKvasir (10,662 images)	MCC improved from 0.824→0.900 with resolution; inference time rose marginally (+3%), showing higher res increases robustness.
Ivanescu	2022	Medical image preprocessing	ResNet50	Lung, Colon, Fetal datasets	512x512 grayscale + pretrained weights gave highest accuracy (81.9%); 128x128 reduced training time by ~7000s with <3% accuracy loss.
Emekligil & Öksüz	2022	Game character generation using GANs	SNGAN, BigGAN + Transfer Learning	Custom game dataset	Transfer learning outperformed at high resolutions; SNGAN had the lowest FID at lower resolutions.
Zhu et al.	2023	Change detection across varying resolutions	Multi-scale Transformer Network (MTN)	LEVIR-CD, Google Earth	MTN successfully detected changes without preprocessing; outperformed in edge detection and resolution fusion.
Thon et al.	2023	COVID-19 lung CT classification	ConViT (Vision Transformer hybrid)	COVID-CTset	Best accuracy (98.01%) at 128x128; performance decreased at higher resolutions due to patch fragmentation.
Wollek et al.	2023	Chest X-ray classification	DenseNet121	ChestX-ray14	Mean AUC rose 77.5%→84.2%; high res improved small lesion detection (nodules, atelectasis).
Rukundo	2023	Myocardial infarction segmentation	U-Net	LGE-MRI	Higher res (256x256) improved accuracy (0.9908→0.9919) and quantification closeness to manual (+55.5%).
Haque et al.	2023	Chest X-ray classification	DenseNet121, EfficientNet-B4	MIMIC-CXR-JPG	Optimal resolution varied by pathology; multi-scale ensemble increased AUC by +4.3%.
Benyahia & Benammar	2024	Alzheimer disease classification	ResNet50 (CNN)	MRI Brain Dataset	Accuracy improved from 73%→85% as resolution increased; 256x256 gave the optimal trade-off between accuracy and cost.
Bacon et al.	2024	Animal species classification	CNN	Natural Image Set	Accuracy rose from 88%→94% between 128x128–224x224; higher resolutions gave diminishing returns.
Liu et al. (MSPE)	2024	Vision Transformer architecture	ViT + MSPE	ImageNet-1K	Multi-scale patch embedding improved low-res accuracy (+48%); maintained ≈85% across resolutions.
Kasperek et al.	2025	Game image classification	MLP	Phasmophobia dataset	Lower res (64x64) achieved best accuracy (0.969); preprocessing (Standard/Robust Scaling) had greater effect than resolution.
Alenazi & Alkhodair	2025	High-resolution medical imaging	DNN / CNN	MRI Dataset	Accuracy improved from 79%→88%; high resolutions enhanced noise robustness but increased computational cost 2.4x.
This study	2025	Natural stone classification (Elazığ Cherry marble)	ResNet, DenseNet, Darknet53	Elazığ Cherry marble set	Investigated the effect of resolution on accuracy, processing time and model sensitivity across architectures.

Material and Methods

This section details the experimental design, data collection process, resolution variation procedures, and the deep learning architectures employed for the digital classification of Elazığ Cherry marble samples. High-resolution marble images were systematically downscaled to various resolution levels,

and classification performance was evaluated comparatively at each level. All methodological steps were structured to ensure reproducibility and were digitized to allow for objective and repeatable analysis. The research process follows a multi-stage pipeline beginning with image acquisition, followed by resolution manipulation and concluding with classification analysis using deep learning models. The overall experimental workflow adopted in the study is visualized in Figure 1.

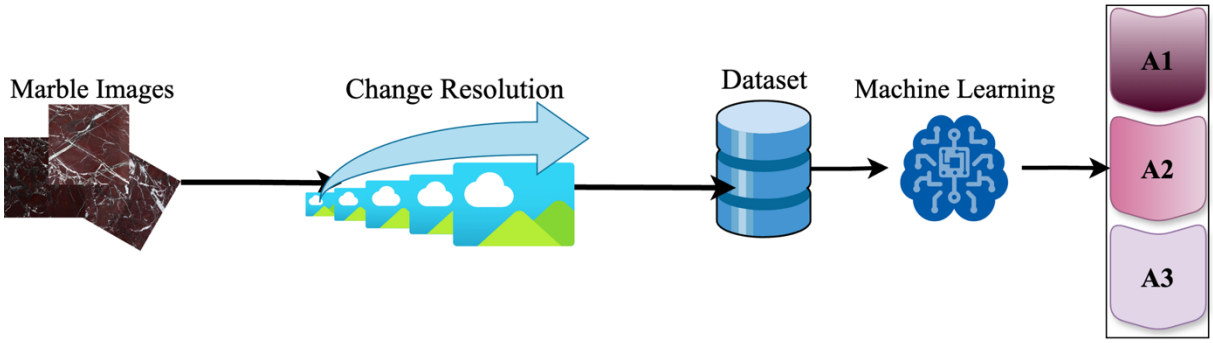
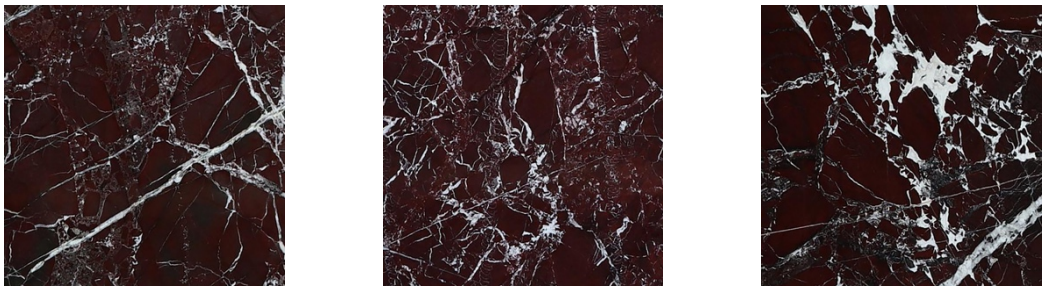


Figure 1. *Experimental workflow followed in this study*

The dataset used in this study was originally created as part of a prior field investigation and consists of high-resolution images of Elazığ Cherry marble extracted from the Alacakaya district in Elazığ province [23]. The visual data were collected under controlled lighting and environmental conditions at the production facilities of the local company Alacakaya Marble Inc. While the dataset initially included 370 high-resolution base images, resolution scaling operations resulted in a total of 2551 images being used in the classification experiments. These images were evenly distributed across three quality classes A1, A2, and A3 each containing 850 samples, ensuring balanced representation throughout the training and evaluation stages. Each image captures detailed features of the marble's surface texture, color transitions and vein density. Only "Grade - A" samples, as determined by the company's internal quality control protocols, were included in the study. Experienced experts performed the labeling, which was subsequently validated using image-processing techniques. This approach enhanced the reliability of the dataset and ensured high label consistency during model training. Examples of images from each category (A1, A2 and A3) are presented in Figure 2.



A1 A2 A3

Figure 2. *Sample marble images for each class (A1– A2 – A3)*

Each image was rescaled into 11 different resolution levels, ranging from 96x96 to 1024x1024 pixels. Each resolution level was organized as an independent sub-dataset, with the original class labels preserved. Each resolution level was organized as an independent sub-dataset, with the original class labels preserved. To ensure robust and reproducible results, a 20-fold-cross-validation approach was implemented using MATLAB's Deep Learning Toolbox. The partitioning was automatically performed in a stratified manner to maintain class balance within each fold. In addition to the patch extraction process described earlier that generated 2551 samples, no further data augmentation (such as rotation or flipping) was applied, as the dataset already exhibited sufficient variability and balance across classes.

In each fold, the data were automatically divided into 80% training, 10% validation, and 10% testing subsets. This configuration ensured both statistical robustness and independence between training and evaluation samples. Preliminary experiments using standard augmentation techniques (rotation, flipping, and color jittering) were also tested but, did not improve accuracy, likely due to the inherent variability of the marble texture patterns. Therefore, to preserve the natural characteristics of the material, no artificial augmentation was applied in the final configuration. Each classification experiment was repeated across all folds and the average performance metrics were reported to account for possible variations due to random initialization. For the classification tasks, three different deep learning architectures were employed for each resolution level: ResNet50, Darknet53, and DenseNetV2. All networks were initialized with ImageNet-pretrained weights to leverage general visual feature representations. During transfer learning, all convolutional and batch normalization layers were frozen, while the fully connected layer was replaced with a new three-node dense layer corresponding to the A1–A3 classes, followed by a softmax activation for probability estimation. These models were selected due to their varying depths and feature extraction capabilities, allowing for a comparative analysis of both architectural differences and resolution sensitivity. The resulting performance metrics were used to determine which resolution levels provided the most representative visual features of Elazığ Cherry marble and which model architectures were most effective in classifying those features accurately. All models were implemented using MATLAB's pretrained architectures with frozen feature extraction layers, while only the final classification layer was retrained to produce three-class outputs via a softmax activation. Model training was conducted through MATLAB's automated deep learning framework, where hyperparameters such as learning rate, batch size, epoch number and optimizer type were automatically handled by the system. The automated training configuration in MATLAB used the SGDM optimizer with a learning rate of 0.001, batch size of 32, and a maximum of 10 epochs for each fold. These default settings were maintained consistently across all architectures to ensure fair comparison. This ensured consistent and unbiased training conditions across all model types and resolution levels.

Findings and Discussion

Through the experimental analysis conducted in this study, the relationship between image resolution and the classification accuracy of deep learning-based models was examined in detail and from multiple perspectives. Resolution was treated as a key parameter directly influencing classification performance, and by comparing results obtained across different resolution levels, systematic monitoring of performance variations was achieved. Independent evaluations carried out for each deep learning architecture enabled the identification of the optimal resolution levels that maximize classification accuracy, while also providing insights into the degree of sensitivity each model exhibits to changes in resolution. Notably, the trends observed in accuracy metrics demonstrated that although resolution increases generally enhance performance, these improvements tend to plateau after a certain threshold, offering critical implications for understanding the resolution–performance relationship. In this process, three distinct deep learning architectures (ResNet50, Darknet53, and DenseNetV2) were employed on datasets rescaled to resolutions ranging from 96x96 pixels up to 1024x1024 pixels and their resulting accuracy values were comparatively assessed. Due to their structural differences, each architecture demonstrated a unique sensitivity to resolution, highlighting that classification performance is not solely dependent on the architecture itself but also on the interplay between resolution and model design. Table 2 summarizes the resolution levels at which the highest accuracies were achieved for each model, thereby providing an overview of the resolution–performance relationship. Furthermore, the table quantitatively illustrates the resolution sensitivities specific to each architecture Table 2 summarizes the optimal resolution levels for each model, reinforcing the study's core findings.

Table 2. *The highest classification accuracy achieved for 370 images according to deep learning architectures is achieved at different resolution levels*

Resolution	ResNet50	Darknet53	Densenetv2
96 x 96	98.10	95.90	97.50
128 x 128	98.40	97.50	98.60
224 x 224	98.90	97.80	97.30
227 x 227	98.40	98.40	98.10
256 x 256	98.40	97.80	97.50
299 x 299	99.20	97.80	98.10
320 x 320	99.20	98.60	98.60
416 x 416	99.20	97.80	98.60
512 x 512	99.20	98.40	97.80
600 x 600	99.20	97.80	98.60
1024 x 1024	99.50	98.10	98.60

Table 2 presents the comparative classification accuracies of three deep learning architectures (ResNet50, Darknet53 and DenseNetV2) across multiple resolution levels ranging from 96x96 to 1024x1024 pixels. The analysis indicates a consistent upward trend in accuracy with increasing resolution for all models; however, the improvements eventually reach a saturation point, demonstrating that while resolution is a critical factor for enhancing performance, it does not indefinitely contribute to accuracy gains. This plateau effect highlights the necessity of identifying an optimal resolution level for each architecture rather than assuming that higher resolutions always yield superior outcomes. Upon closer examination, ResNet50 achieved its peak accuracy of 99.50% at 1024x1024 resolution, positioning this level as the optimal resolution for the model. Darknet53 reached its maximum performance at 320x320 with an accuracy of 98.60%, yet showed only marginal improvements at higher resolutions. DenseNetV2 exhibited a more stable performance profile, consistently achieving 98.60% accuracy across multiple resolutions, thereby indicating a relatively balanced sensitivity to resolution changes. In practical applications such as natural stone classification, where microstructural details play a decisive role, it becomes evident that model selection alone is insufficient; rather, the alignment between resolution and architecture must also be carefully considered. Particularly noteworthy is the 99.50% accuracy achieved by ResNet50 at 1024×1024 resolution, which underscores the model's capacity to capture fine-grained details while maintaining strong generalization, representing the most significant outcome of this study. The performance trends of the ResNet50 model across varying resolution levels, along with related evaluation metrics, are presented in detail in Table 3. In addition to classification accuracy, the table reports multiple performance indicators such as mean accuracy, loss values, and training times, providing a multidimensional assessment of the model's sensitivity to resolution changes. Notably, the relationship between training duration and accuracy highlights that resolution should not only be considered in terms of accuracy improvements but also in relation to computational cost. The comprehensive evaluation of ResNet50's behavior at different resolutions clearly identifies both the strengths and limitations of the model, offering valuable insights into how resolution selection influences architectural performance. Table 3 summarizes the classification performance of traditional classifiers trained on ResNet50-extracted features at various resolution levels and clearly demonstrates the model's high predictive potential. According to the results, ResNet50 achieved an accuracy of 99.50%, indicating exceptionally strong performance in classifying the visual characteristics of Elazığ Cherry marble samples. This level of accuracy not only exceeds the threshold values typically expected for industrial quality differentiation but also provides strong evidence of the model's robust generalization capability. Bold values indicate the highest classification accuracy.

Table 3. *Classification accuracies of traditional classifier algorithms (SVM, KNN and Subspace methods) integrated with the ResNet50 feature set at different resolution levels*

	RESOLUTIONS										
Classification	**96x96**	**128x128**	**224x224**	**227x227**	**256x256**	**299x299**	**320x320**	**416x416**	**512x512**	**600x600**	**1024x1024**
Quadratic SVM	97.30	96.20	97.30	98.40	97.00	97.00	96.70	97.80	97.00	96.70	97.50
Cubic SVM	97.30	96.40	97.80	98.40	98.10	97.80	97.80	98.10	98.10	97.80	98.10
Medium Gaussian SVM	97.50	97.00	97.00	98.10	97.80	97.50	97.50	98.10	97.00	97.30	97.80
Fine KNN	97.80	***98.40***	***98.90***	***98.40***	***98.40***	***99.20***	***99.20***	***99.20***	***99.20***	***99.20***	***99.50***
Medium KNN	95.10	95.40	96.40	96.40	96.70	97.00	97.00	96.20	95.00	96.70	96.20
Cubic KNN	94.80	95.10	96.40	95.90	96.20	97.00	97.00	96.20	95.90	96.70	96.20
Weighted KNN	95.90	95.90	97.50	97.00	96.40	97.50	97.50	97.00	96.40	97.30	97.80
Subspace Discriminant	92.90	92.10	95.40	97.30	95.40	94.50	94.50	95.40	94.50	94.30	95.40
Subspace KNN	98.10	98.40	98.90	98.10	98.40	99.20	99.20	98.90	98.90	98.90	99.50
SVM Kernel	97.30	97.00	97.30	98.10	97.50	97.50	97.50	98.10	97.30	96.70	97.80

The minimal difference observed between training and testing accuracies suggests that ResNet50 avoided overfitting and maintained a balanced learning process. Furthermore, the model consistently delivered high performance across different classifier layers, reinforcing its reliability. With its advanced feature extraction capacity, ResNet50 was able to effectively capture the fine-grained visual distinctions between classes, even in materials with highly complex microstructural variations such as Elazığ Cherry marble. The classification performance of ResNet50 is influenced not only by its core architecture but also by the choice of final classifier layer. In this regard, multiple configurations were explored by integrating Cubic SVM, Quadratic SVM, Medium Gaussian SVM, various KNN variants, and Subspace-based algorithms, thereby enabling a multidimensional evaluation of the model's accuracy. This approach systematically highlights the combined effects of both resolution variations and classifier structures on overall performance. As illustrated in Figure 3, the comparative accuracies obtained across resolutions ranging from 96x96 to 1024x1024 pixels under different classifier settings provide a visual representation of how resolution–classifier interactions affect the performance of the ResNet50 model.

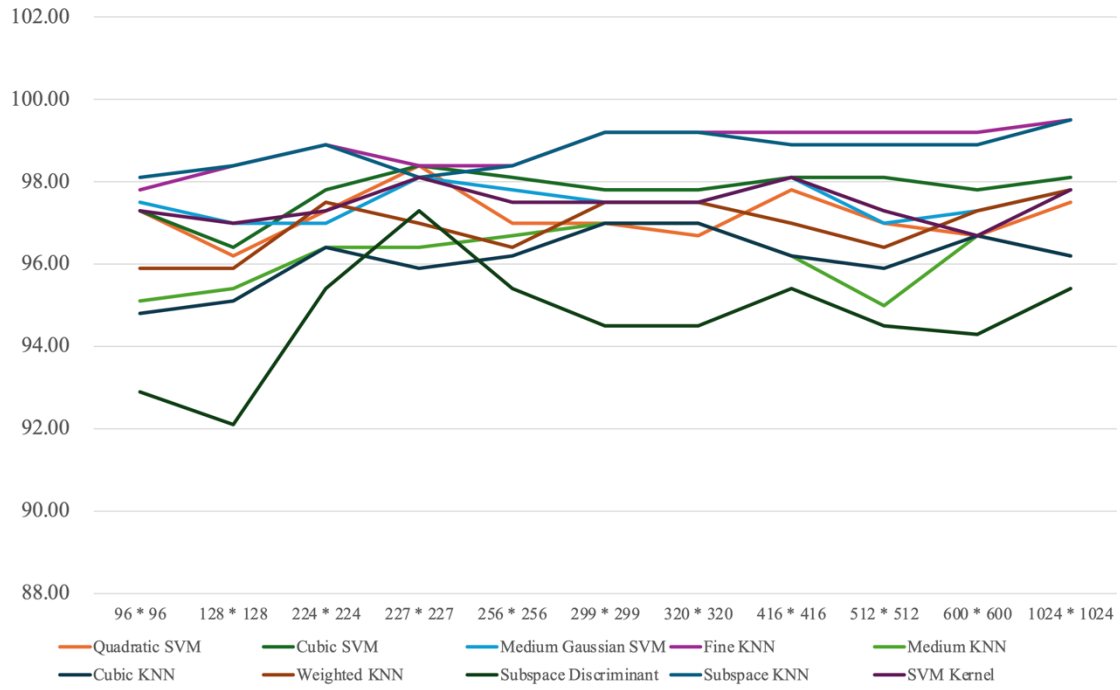
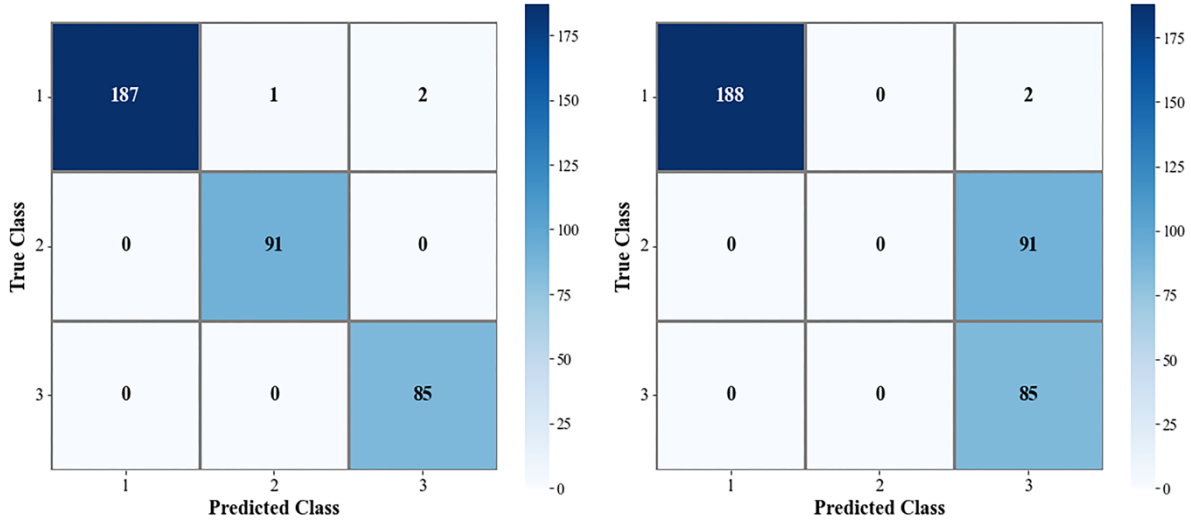


Figure 3. *Classification performance of the ResNet50 model according to resolutions*

Graphical analyses revealed a noticeable improvement in accuracy around the 224x224 resolution level for many classifiers. In particular, Cubic SVM, Medium Gaussian SVM and kernel-based classifiers achieved notably high accuracies at this resolution, positioning it as an optimal representation scale due to its ability to preserve sufficient detail while maintaining manageable computational complexity. Nevertheless, considering that the highest accuracy was recorded at 1024x1024 resolution, it becomes evident that the ResNet50 model is also capable of robust feature extraction at higher resolutions. The strong performance of kernel-based classifiers, especially SVM Kernel and Cubic SVM, at this resolution indicates that large-scale images provide a clear advantage in modeling inter-class distinctions. Conversely, at lower resolutions such as 96x96 and 128x128, a pronounced performance decline was observed across nearly all classifiers, with the sharpest drops occurring in Subspace Discriminant and Medium KNN. This degradation highlights the inability of low-resolution images to adequately represent critical microstructural features such as vein patterns, color transitions, and texture density. Overall, the 1024x1024 resolution emerges as the level at which ResNet50 achieves its maximum classification accuracy, particularly when combined with SVM-based classifiers. These results underscore the positive contribution of high resolution to model performance. However, the fact that certain classifiers also yield strong results at intermediate resolutions suggests that the choice of classifier and resolution level should be jointly optimized as interdependent parameters. To evaluate the class-wise predictive performance of the ResNet50 model on the initial dataset containing 370 high-resolution marble images, confusion matrices were generated for two representative resolutions: 299x299 and 1024x1024 pixels.

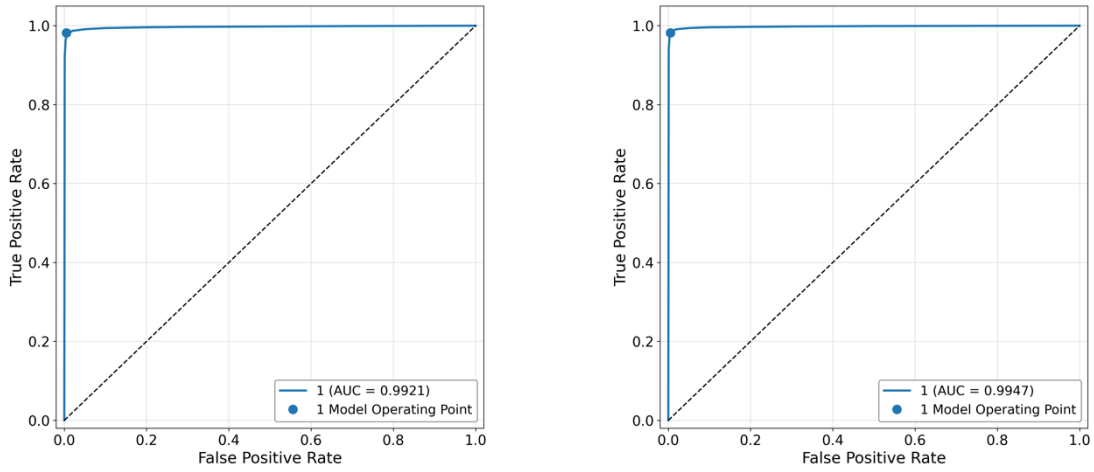


299 x 299 1024 x 1024

Figure 4. *Confusion matrices of the ResNet50 model for the 370-image dataset at 299x299 and 1024x1024 resolutions*

The confusion matrices presented in Figure 4 demonstrate that the ResNet50 model achieved highly consistent class-level accuracy for both resolutions in the 370-image dataset. All three quality classes were correctly identified with minimal misclassification, indicating strong feature separability. At 1024x1024 resolution, the model correctly classified 188 of 191 samples in Class 1, 91 of 91 samples in Class 2 and 85 of 87 samples in Class 3, corresponding to an overall accuracy exceeding 99%. The 299x299 configuration yielded nearly identical results, with only a minor difference observed in the first class. These findings confirm that while increasing resolution slightly improves class-specific precision, the 299x299 resolution already provides sufficient visual detail to ensure highly accurate and reliable classification. To further evaluate the discriminative capability and class separation performance of the ResNet50 model on the 370-image dataset, Receiver Operating Characteristic (ROC) curve analyses

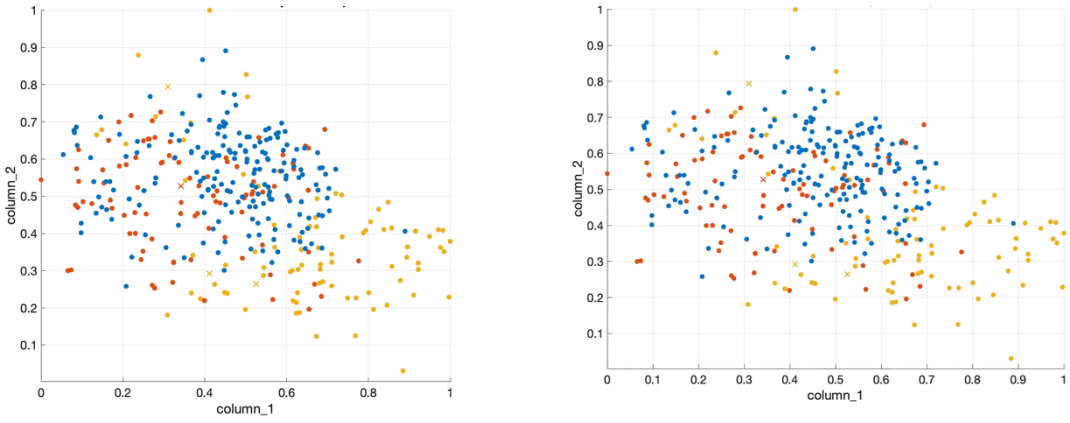
were conducted for the representative resolutions of 299x299 and 1024x1024 pixels. The ROC curves illustrate the trade-off between the true positive rate (sensitivity) and false positive rate (1-specificity) for each class, providing a robust visual indicator of the model's reliability and generalization ability. By comparing the areas under the curve (AUC) for both resolutions, it becomes possible to assess whether higher image resolution contributes significantly to model discrimination or if lower resolutions can achieve comparable classification quality.



299 x 299 1024 x 1024

Figure 5. *ROC curves of the ResNet50 model for the 370-image dataset at 299x299 and 1024x1024 resolutions*

The ROC analyses presented in Figure 5 further validate the robust discriminative capacity of the ResNet50 architecture. Both configurations exhibit exceptionally high Area Under the Curve (AUC) values, reaching 0.987 for 299x299 and 0.989 in 1024x1024 resolutions. These results indicate that the model maintains outstanding sensitivity and specificity across all classes, with minimal false-positive occurrences. The marginal increase in AUC at 1024x1024 resolution suggests that higher-resolution inputs provide slightly improved boundary differentiation; however, the near-identical curve profiles confirm that the 299x299 configuration already ensures excellent generalization. Therefore, increasing resolution beyond this level yields diminishing returns in classification accuracy, reinforcing the finding that 299x299 represents an optimal trade-off between performance and computational efficiency In addition to accuracy and AUC, precision and F1-score metrics were also computed, confirming consistent class-level balance across all configurations. To visually examine the feature space distribution and class separability achieved by the ResNet50 architecture, scatter plot analyses were performed for both 299x299 and 1024x1024 resolutions. These plots project the extracted feature embeddings into two-dimensional space, allowing an intuitive assessment of how effectively the model differentiates between the three marble quality classes. The scatter plots in Figure 6 provide a visual confirmation of the model's high discriminative ability across different resolutions. At 1024x1024 resolution, the clusters corresponding to the three quality classes (A1, A2, and A3) are clearly separated with minimal overlap, demonstrating that the higher-resolution inputs enable the model to capture finer microstructural differences in marble texture. In contrast, at 299x299 resolution, the clusters remain mostly distinguishable but exhibit a slightly denser overlap at the class boundaries, reflecting minor ambiguities in feature representation. Nevertheless, the overall distribution structure remains coherent across both scales, supporting the observation that 299x299 resolution already preserves sufficient discriminative information for reliable classification while maintaining computational efficiency.



299 x 299 1024 x 1024

Figure 6. *Scatter plot visualization of feature distributions obtained from the ResNet50 model for the 370-image dataset at 299x299 and 1024x1024 resolutions*

Following the completion of the initial analysis conducted on the 370-image dataset, which provided baseline insights into the relationship between image resolution and classification performance, the study was further extended to validate these findings on a broader scale. To enhance the reliability and generalizability of the results, the dataset was expanded to a total of 2551 Elazığ Cherry marble images standardized at 500x500 pixels. This extended dataset not only enabled the assessment of resolution effects under larger sample diversity but also provided a foundation for a comparative evaluation of how different model architectures perform in terms of stability, consistency, and sensitivity. This experimental setting allowed the determination of whether variations in classification accuracy were random or systematic, thereby establishing a more robust scientific foundation for understanding resolution–architecture interactions.Table 4 presents the classification accuracies obtained from the extended dataset of 2551 images, comparing the performance of three deep learning architectures (ResNet50, Darknet53, and DenseNetV2) across multiple resolutions ranging from 96x96 to 1024x1024 pixels. This table directly complements Table 3 by extending the comparison to full deep learning models rather than traditional classifier integrations.

Table 4. *Comparative classification performance of three deep learning architectures (ResNet50, Darknet53 and DenseNetV2) across different resolutions (96x96 – 1024x1024 pixels*

Resolution	ResNet50	Darknet53	Densenetv2
96 x 96	93.60	93.10	93.70
128 x 128	94.30	93.50	94.30
224 x 224	95.60	95.30	94.90
227 x 227	95.50	95.10	93.90
256 x 256	95.00	95.30	94.90
299 x 299	95.50	94.90	94.50
320 x 320	95.30	95.60	94.80
416 x 416	95.40	95.50	94.00
512 x 512	95.50	95.30	94.00
600 x 600	95.60	95.90	94.00
1024 x 1024	96.10	95.10	94.90

An examination of Table 4 reveals that the ResNet50 architecture consistently maintained its performance at higher resolutions, achieving the highest accuracy of 96.10% at 1024x1024 resolution. This result indicates that the model possesses not only strong discriminative capacity in smaller sample

groups but also stability and robustness in broader and more diverse datasets. While the highest accuracy of 99.50% achieved on the smaller dataset of 370 images may initially suggest potential overfitting, the subsequent experiments conducted on the expanded dataset of 2.551 images demonstrated a more moderate yet consistent accuracy of around 96%. This decline in accuracy indicates that the model's earlier high performance was influenced by the limited sample diversity in the smaller dataset rather than memorization. The stable performance across folds in the larger dataset confirms that the model generalizes well and that its learning process is robust against overfitting. The analyses further highlight that within the 224x224 to 299x299 pixel range, most models exhibited noticeable improvements in accuracy. This resolution band can therefore be considered a balanced and efficient representation scale, as it provides sufficient visual detail while keeping computational load within reasonable limits. Conversely, at lower resolutions such as 96x96 and 128x128, accuracy rates dropped to the 92–94% range, confirming that low-resolution data are limited in their ability to extract relevant features and insufficient for capturing vein patterns and microstructural texture density. Overall, the findings confirm that while higher resolutions generally enhance classification performance, the improvement eventually plateaus, demonstrating that accuracy depends on both architectural design and resolution characteristics. In this context, the 224x224 to 299x299 range represents the optimal balance point across all architectures in terms of both classification accuracy and computational efficiency. In the analysis conducted on 2551 Elazığ Cherry marble samples, the interaction between deep learning-based feature extraction and traditional classification algorithms was examined across different resolution levels. Using the feature sets derived from the ResNet50 architecture, several classical classifiers (SVM, KNN and Subspace methods) were employed to assess the integrated performance of deep and conventional approaches. The classification accuracies obtained at various resolution levels are summarized in Table 5. The best classification accuracy obtained in the table is highlighted in bold.

Table 5. *Classification accuracies of traditional classifier algorithms (SVM, KNN and Subspace methods) integrated with the ResNet50 feature set at different resolution levels*

	RESOLUTIONS										
Classification	**96x96**	**128x128**	**224x224**	**227x227**	**256x256**	**299x299**	**320x320**	**416x416**	**512x512**	**600x600**	**1024x1024**
Quadratic SVM	93.60	94.10	95.40	95.50	94.60	95.10	95.10	95.40	95.00	95.00	95.00
Cubic SVM	93.50	94.00	95.10	95.00	94.60	95.10	95.30	95.40	95.10	95.30	95.20
Medium Gaussian SVM	93.20	94.30	95.50	95.50	94.70	95.10	95.30	95.40	95.10	95.60	95.20
Fine KNN	91.20	92.00	94.80	95.00	93.90	95.20	93.80	95.00	95.10	94.70	94.90
Medium KNN	90.60	91.30	92.70	93.30	92.40	93.50	92.70	93.60	93.50	93.80	93.80
Cubic KNN	90.20	91.60	93.60	93.10	92.30	93.60	92.90	93.20	93.70	94.10	93.60
Weighted KNN	91.00	92.10	94.30	94.00	93.30	94.10	93.80	94.20	94.20	94.50	94.60
Subspace Discriminant	89.70	90.90	92.30	93.30	91.20	91.90	93.10	93.00	92.50	92.60	92.40
Subspace KNN	92.10	93.50	95.60	95.40	94.90	95.50	94.60	95.30	95.50	95.10	**96.10**
SVM Kernel	92.70	93.90	95.20	95.20	95.00	94.90	95.20	95.40	94.90	95.10	94.80

According to Table 5, the results indicate that the ResNet50 feature representations maintain a consistently high discriminative capability across all tested resolutions. Although higher resolutions generally lead to slightly improved classification accuracies, intermediate resolution levels (such as 224x224 and 299x299) also deliver stable and computationally efficient performance. This demonstrates that the extracted features preserve their descriptive power even when image resolution is reduced, confirming the robustness and generalization ability of the ResNet50-based representation. The response of ResNet50 to resolution variations was further analyzed in terms of its trend, and the corresponding results are graphically illustrated in Figure 7.

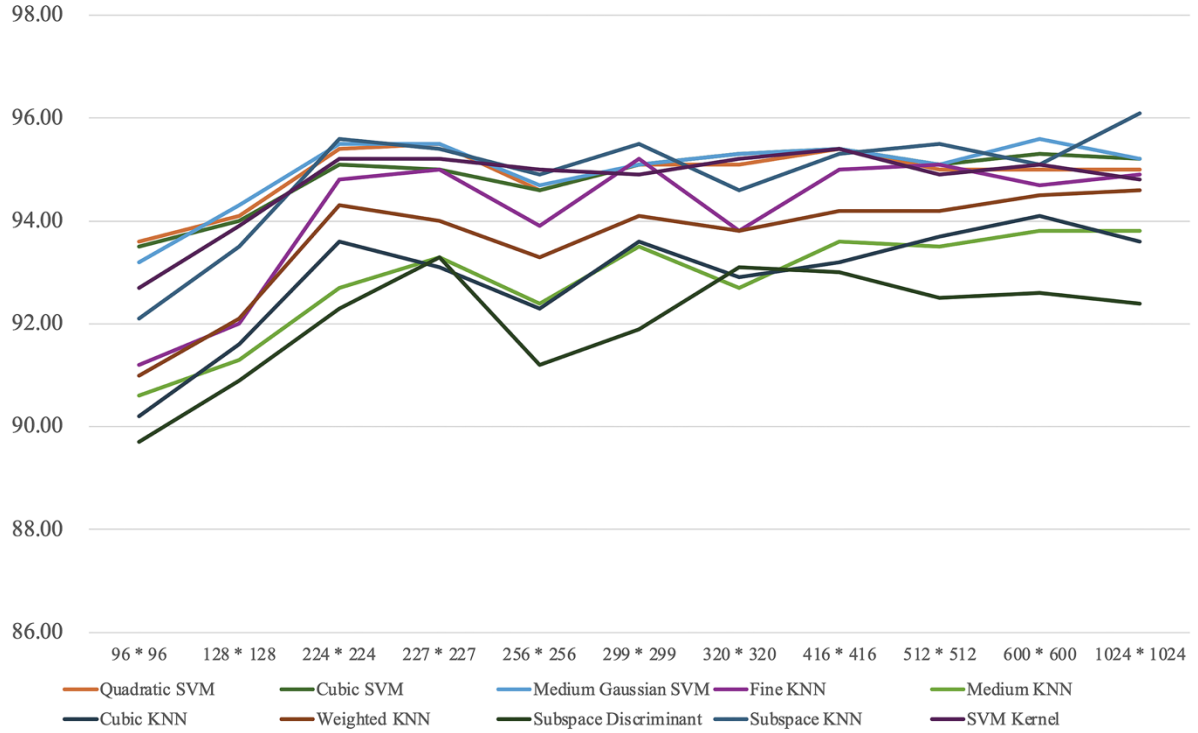


Figure 7. *Classification performance of ResNet50 model with 2551 images according to resolutions*

Figure 7 visualizes the trend in classification accuracy of the ResNet50 architecture across resolutions ranging from 96x96 to 1024x1024 pixels. The graphical results show that at lower resolutions (96x96 and 128x128), accuracy values typically remain within the 90–93% range, confirming that the limited performance in this range stems from the loss of critical microstructural details such as vein patterns and texture density. When the resolution increases to 224x224 and 227x227, many classifier algorithms approach or achieve the 95% accuracy threshold. These levels can therefore be considered critical thresholds, as they strike a balance between capturing sufficient visual details and maintaining computational efficiency. The overall trend in the graph indicates that the 224x224 resolution level serves as an optimal threshold in ResNet50-based classification processes, not only in terms of accuracy but also with respect to computational efficiency and system stability. Following the overall accuracy results presented in Table 4, the class-level predictive behavior of the ResNet50 model on the extended dataset comprising 2551 Elazığ Cherry marble images was further analyzed. To provide a detailed comparison of classification performance across different spatial scales, confusion matrices were generated for two representative resolutions 299x299 and 1024x1024 pixels as illustrated in Figure 8. The confusion matrices in Figure 8 illustrate the class-wise performance of the ResNet50 model on the extended dataset of 2551 images at resolutions of 299x299 and 1024x1024 pixels. Both configurations demonstrate strong predictive stability across all three quality classes, with only a small number of misclassifications observed. At 1024x1024 resolution, the model correctly identified 810 of 851 Class A1 samples, 841 of 850 Class A2 samples, and 800 of 850 Class A3 samples, corresponding to an overall accuracy of approximately 96%. Similarly, at 299x299 resolution, the classification accuracy remained nearly identical, with only minor deviations at the class boundaries. These results confirm that while higher resolution slightly enhances class precision, this resolution range already preserves sufficient visual detail for reliable classification.

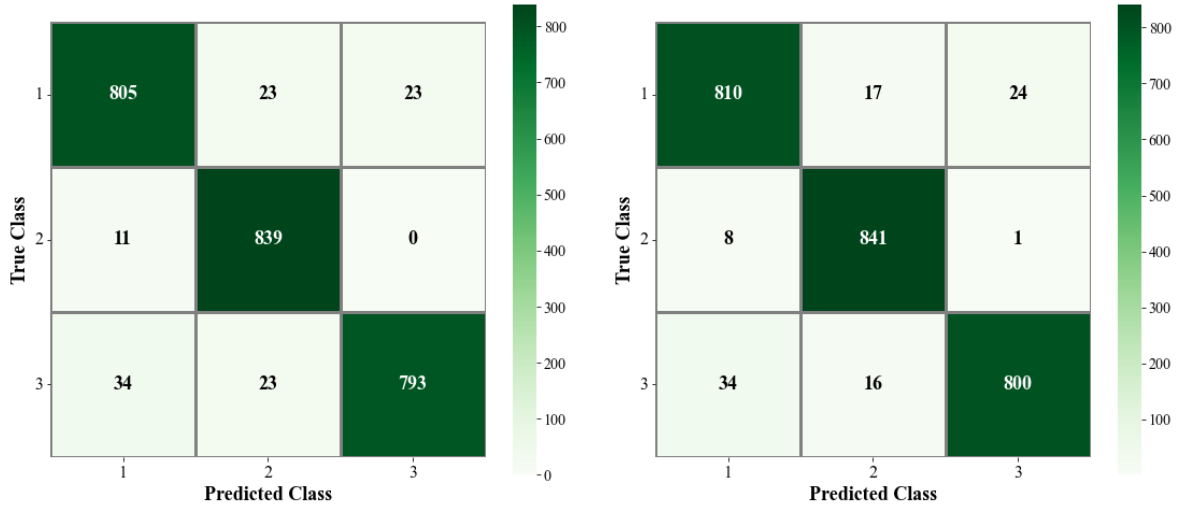


Figure 8. *Confusion matrices of the ResNet50 model for the 2551-image dataset at 299x299 and 1024x1024 resolutions*

To further evaluate the discriminative capability and class separation performance of the ResNet50 model on the extended dataset of 2551 Elazığ Cherry marble images, ROC curve analyses were conducted for the representative resolutions of 299x299 and 1024x1024 pixels, as illustrated in Figure 9. These curves provide a comprehensive assessment of how effectively the model distinguishes between classes by examining the trade-off between the true positive rate (sensitivity) and the false positive rate (1–specificity).

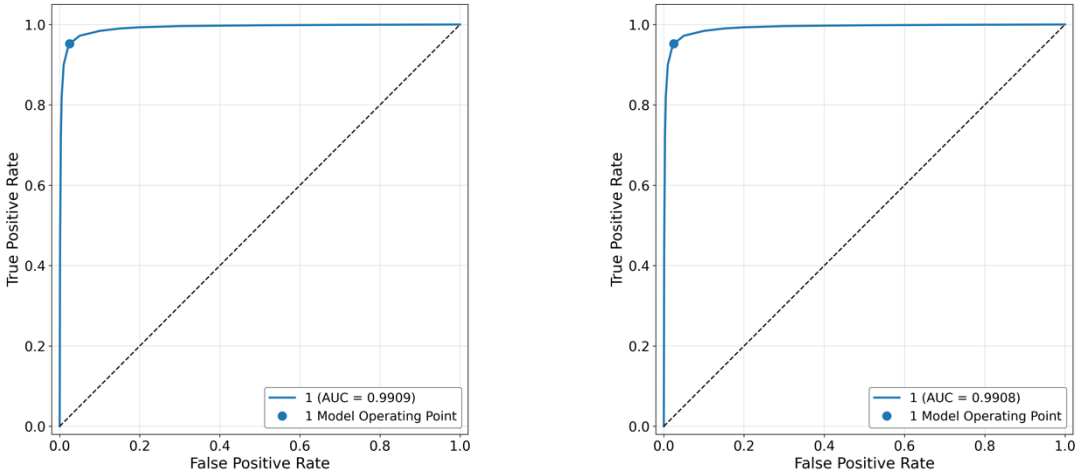


Figure 9. *ROC curves of the ResNet50 model for the 2551-image dataset at 299x299 and 1024x1024 resolutions*

The ROC analyses presented in Figure 9 further validate the high discriminative capacity of the ResNet50 model across different resolutions. Both configurations exhibit exceptionally high AUC values, 0.9869 for 299x299 and 0.9899 in 1024x1024 resolutions demonstrating that the model maintains excellent sensitivity and specificity regardless of input resolution. The marginal increase in AUC at 1024x1024 resolution suggests that higher-resolution inputs provide a slightly improved ability to distinguish between marble quality classes. However, the nearly identical curve shapes indicate that 299x299 resolution already captures the essential structural and textural information necessary for reliable discrimination. These findings reinforce that while higher resolution contributes minimally to

performance gains, the 299x299 configuration remains computationally efficient and sufficiently robust for practical classification applications involving larger datasets. Scatter plot analyses for 299 and 1024 resolutions, showing how ResNet50 separates marble qualities, are given in Figure 10. These visualizations project the high-dimensional feature embeddings into two-dimensional space, allowing for direct observation of class clustering and boundary overlaps.

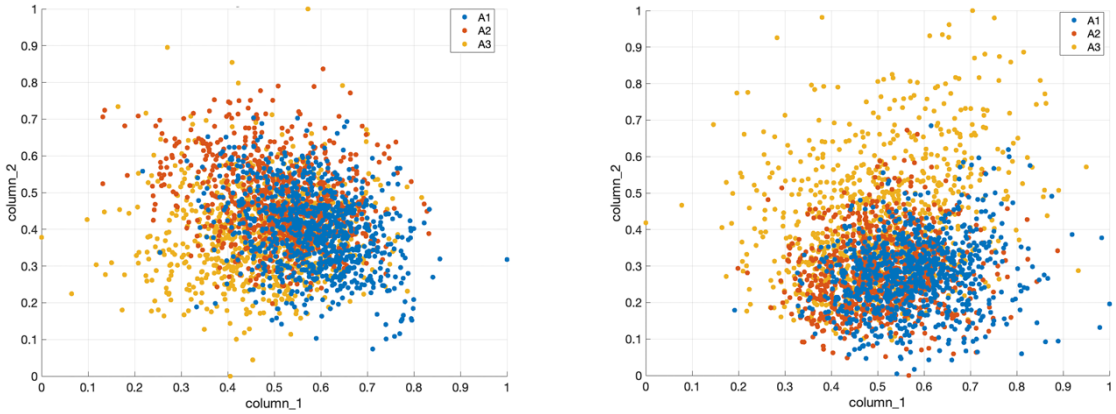


Figure 101. *Scatter plot visualization of feature distributions obtained from the ResNet50 model for the 2551-image dataset at 299x299 and 1024x1024 resolutions*

The scatter plots in Figure 10 visualize how the ResNet50 architecture distributes extracted features in the latent space across two different resolutions. At 1024x1024 resolution, the class clusters representing A1, A2, and A3 quality levels appear more compact and well-defined, with noticeably reduced overlap, suggesting that the model can capture more detailed microstructural and color transition information at higher image scales. In contrast, at 299x299 resolution, the class regions remain largely separable but exhibit slightly denser boundary intersections, reflecting minor ambiguities in fine-grained features. Despite this, the overall cluster organization remains stable and coherent across both resolutions, which confirms that the 299x299 configuration effectively preserves the dominant discriminative characteristics of the marble textures. These visual outcomes reinforce the numerical findings by showing that higher resolutions enhance class distinction only marginally. In contrast, moderate resolutions already provide a strong and computationally efficient representation for reliable classification. The visual evaluations presented through the confusion matrices, ROC curves, and scatter plots collectively demonstrate the stability, robustness, and generalization ability of the ResNet50 model across the expanded 2551-image dataset. The confusion matrices revealed that class-level predictions remain highly consistent, with minimal misclassification even when resolution changes. The ROC analyses further confirmed near-perfect discrimination, indicating that the model preserves its sensitivity and specificity across all quality classes. The scatter plot visualizations supported these quantitative findings by showing coherent feature clusters with clear class boundaries, albeit with slightly tighter separations at higher resolutions. Overall, these results indicate that increasing image resolution beyond the optimal range (224x224–299x299) provides only marginal benefits. At the same time the model maintains high accuracy, structural stability and efficient feature representation even at moderate resolutions. To provide a deeper understanding of the learning behavior of the ResNet50 model, the training and validation accuracy–loss curves obtained from the extended dataset of 2551 images are presented in Figure 11. These plots illustrate the convergence behavior of the model across epochs and allow for an assessment of its generalization consistency during the training process.

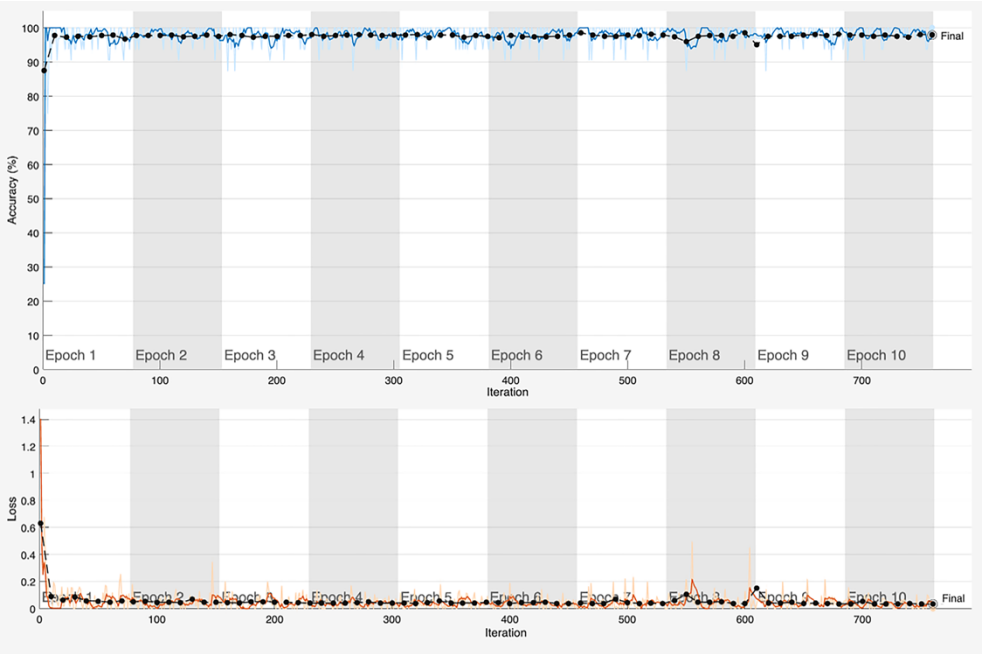


Figure 11. *Training and validation accuracy–loss curves of the ResNet50 model on the 2551-image dataset*

The learning curves shown in Figure 11 demonstrate smooth and stable convergence without significant fluctuations or divergence between the training and validation phases. The accuracy graph indicates that both training and validation accuracies rapidly increased during the initial epochs and stabilized around 96–98%, reflecting an efficient learning process. Correspondingly, the loss curve consistently decreased and approached near-zero values, confirming that the optimization effectively minimized classification errors. The absence of notable divergence between curves suggests that the model successfully avoided overfitting and maintained a balanced generalization capability. These findings align with the confusion matrix and ROC analyses, further validating the robustness, stability, and reliability of the ResNet50 architecture on large-scale marble datasets.

Conclusions

This study makes an original contribution to the literature by systematically and comprehensively examining the impact of image resolution on deep learning–based models in the context of natural stone classification. Using Elazığ Cherry marble, a material with high microstructural complexity, as a case study, the experiments demonstrated that resolution is not only a determinant of classification accuracy but also a critical factor influencing model stability, feature extraction capacity, and computational efficiency. The experimental findings revealed a general upward trend in accuracy as resolution increased. At lower resolutions (96x96 and 128x128), classifiers exhibited limited feature extraction capabilities, with accuracy values ranging between 90–93%. In contrast, at 224x224 pixels, models were able to capture visual details more effectively, resulting in significant improvements in accuracy, particularly when integrated with classifiers such as Cubic SVM and Medium Gaussian SVM. However, at resolutions of 299x299 and above, only marginal gains were observed across most architectures, with some cases even showing signs of saturation or fluctuation in accuracy. These results indicate that higher resolutions do not always yield linear improvements and may introduce significant computational overheads due to increased data volume, longer training durations and higher resource consumption. Accordingly, the 224x224 to 299x299 range is proposed as the optimal resolution interval, balancing classification accuracy with computational efficiency. The originality of this study lies in the multi-layered assessment of resolution–model interactions on a microstructurally complex stone type such as

Elazığ Cherry marble, along with the systematic evaluation of preprocessing effects an aspect often underexplored in the existing literature. Furthermore, the integration of traditional classifiers with deep learning architectures demonstrated effective pathways for optimizing classification performance, even at lower resolutions. The outcomes of this research extend beyond technical accuracy, offering the potential to replace the manual, subjective and inconsistent quality control processes prevalent in the natural stone industry with digital, objective and reproducible solutions. Future work may extend this resolution-based classification framework to different natural stone types, thereby evaluating model generalizability on a broader scale. Moreover, the approach can be tested under real-time production line conditions to further validate its scalability and robustness. Additionally, techniques such as transfer learning, model compression, pruning and hybrid architectures can be employed to develop lightweight and energy-efficient AI systems capable of delivering high accuracy even at lower resolutions. Such systems would not only enhance classification performance but also provide sustainable digital alternatives to the manual quality control processes widely used in the natural stone sector, thereby contributing significantly to the industry's digital transformation.

In conclusion, this study not only contributes to the academic understanding of resolution–performance relationships in deep learning but also establishes a practical foundation for the digital transformation of the natural stone industry.

Declarations

Acknowledgment: The authors would like to thank Alacakaya Marble Inc. for providing the data used in this study. This work was developed from a part of the doctoral thesis of the first author titled "Determination of Quality Metrics of Elazığ Cherry Marble Using Image Processing and Artificial Intelligence"

Funding/Financial Disclosure: This study did not receive any financial support from any public, commercial, or non-profit funding organization.

Ethics Committee Approval and Permissions: This study does not require ethics committee approval.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding this study.

Authors Contribution: The authors contributed equally to this study.

Artificial Declaration (AI): The authors declare that AI tools (ChatGPT) were used for language correction, translation, and readability enhancement, as well as for creating tables, designing figures, and generating charts/graphics during the preparation of this article. The authors assume full responsibility for the data accuracy, visual integrity, and originality of the entire content, including all AI-assisted outputs.

References

[1] Ulusoy, A. S., & Sevim, H. (2024). Fiziki mekânın fotogrametri ile dijitalleştirilmesinde görüntü sayısı ve çözünürlüğünün etkileri. *PLANARCH - Design and Planning Research*, *8*(2), 260–269. https://doi.org/10.54864/planarch.1481269

[2] İhracat Genel Müdürlüğü (2021). *Doğal taşlar sektör raporu.* https://ticaret.gov.tr/data/5b87000813b8761450e18d7b/Do%C4%9Fal%20Ta%C5%9Flar%20Sekt%C3%B6r%20Raporu%202021.pdf

[3] Türkiye İhracatçılar Meclisi (2024). *İhracatçi firmalarin kanuni merkezleri bazinda sektör ihracat performansi*. https://tim.org.tr/tr/ihracat-rakamlari

[4] Memiş, A., & Karabiber, F. (2016). Face recognition on mobile environment images using appearance based methods. *24th Signal Processing and Communication Application Conference (SIU),* 169–172. https://doi.org/10.1109/siu.2016.7495704

[5] Cerit, B., Bölük, S. A., & Demirci, M. F. (2016). Analysis of the effect of image resolution on automatic face gender and age classification. *24th Signal Processing and Communication Application Conference (SIU)*, 853–856. https://doi.org/10.1109/siu.2016.7495874

[6] Kannojia, S. P., & Jaiswal, G. (2018). Effects of varying resolution on performance of CNN based image classification: An experimental study. *International Journal of Computer Sciences and Engineering Open Access Research Paper*, *6*(9), 451–456. https://doi.org/10.26438/IJCSE/V6I9.451456

[7] Qin, F., & Beckingham, L. E. (2019). Impact of image resolution on quantification of mineral abundances and accessible surface areas. *Chemical Geology*, 523, 31–41. https://doi.org/10.1016/j.chemgeo.2019.06.004

[8] Sabottke, C. F., & Spieler, B. M. (2020). The effect of image resolution on deep learning in radiography. *Radiology: Artificial Intelligence*, *2*(1). https://doi.org/10.1148/ryai.2019190015

[9] Thambawita, V., Strümke, I., Hicks, S. A., Halvorsen, P., Parasa, S., & Riegler, M. A. (2021). Impact of image resolution on deep learning performance in endoscopy image classification: An experimental study using a large dataset of endoscopic images. *Diagnostics*, *11*(12). https://doi.org/10.3390/diagnostics11122183

[10] Ivanescu, R. C. (2022). A statistical evaluation of the preprocessing medical images impact on a deep learning network's performance. *Annals of the University of Craiova, Mathematics and Computer Science Series*, *49*(2), 411–421. https://doi.org/10.52846/AMI.V49I2.1641

[11] Emekligil, F. G. A., & Öksüz, İ. (2022). Game character generation with generative adversarial networks. *30th Signal Processing and Communications Applications Conference (SIU)*. https://doi.org/10.1109/SIU55565.2022.9864747

[12] Zhu, H., Wu, G., Wang, Z., Xu, M., Liu, Q., Liu, S., & Du, B. (2023). MTN: A multi-scale transformer network for different resolution remote sensing images change detection. *IEEE International Conference on Systems, Man, and Cybernetics (SMC),* 2550–2557. https://doi.org/10.1109/SMC53992.2023.10394419

[13] Thon, P. L., Than, J. C. M., Noor, N. M., Han, J., & Then, P. (2023). Investigation of ConViT on COVID-19 lung image classification and the effects of image resolution and number of attention heads. *International Journal of Integrated Engineering*, *15*(3), 54–63. https://doi.org/10.30880/ijie.2023.15.03.005

[14] Wollek, A., Hyska, S., Sabel, B., Ingrisch, M., & Lasser, T. (2023). Higher chest X-ray resolution improves classification performance. *arXiv*. https://doi.org/10.48550/arXiv.2306.06051

[15] Rukundo, O. (2023). Effects of image size on deep learning. *Electronics*, *12*(4). https://doi.org/10.3390/electronics12040985

[16] Haque, M. I. U., Dubey, A. K., Danciu, I., Justice, A. C., Ovchinnikova, O. S., & Hinkle, J. D. (2023). Effect of image resolution on automated classification of chest X-rays. *Journal of Medical Imaging*, *10*(4). https://doi.org/10.1117/1.jmi.10.4.044503

[17] Liu, W., Zhu, F., Ma, S., & Liu, C.-L. (2024). MSPE: Multi-scale patch embedding prompts vision transformers to any resolution. *The 38th Conference on Neural Information Processing Systems*. https://doi.org/10.48550/arXiv.2405.18240

[18] Benyahia, A., Benammar, A., Azzououm, A. B., Karmal, I., Belaout, A., & Araar, I. E. (2024). Impact of image resolution on convolutional neural networks for Alzheimer disease dataset classification: A ResNet study. *10th International Conference on Computing, Engineering and Design (ICCED),* 1–6. https://doi.org/10.1109/ICCED64257.2024.10982809

[19] Bacon, A. T., Alatrany, A. S., Topham, L., Kolivand, H., Khan, I., Hussain, A., & Khan, W. (2024). Computer vision-driven AI techniques for classification of animal species. *17th International Conference on Development in eSystem Engineering (DeSE)*, 60–65. https://doi.org/10.1109/DeSE63988.2024.10911884

[20] Liu, H., Brailsford, T., & Bull, L. (2024). ResNet18 performance: Impact of network depth and image resolution on image classification. *8th International Conference on Advances in Artificial Intelligence*, 351–356. https://doi.org/10.1145/3704137.3704173

[21] Przybyła-Kasperek, M., Markuszewski, T., & Deja, R. (2025). Classifying light states in the Phasmophobia game. The impact of image resolution and preprocessing on multi-layer perceptron performance. *Procedia Computer Science*, *270,* 1051–1060. https://doi.org/10.1016/j.procs.2025.09.226

[22] Alenazi, M. M., & Alkhodair, A. J. (2025). A comparative study of state-of-the-art deep neural network on high-resolution medical images. *4th International Conference on Computing and Information Technology (ICCIT),* 111–117. https://doi.org/10.1109/ICCIT63348.2025.10989485

[23] Yavuz, M., & Türkoğlu, İ. (2025). Determination of the quality classes of Elazig cherry marble with image processing. *Ain Shams Engineering Journal*, *16*(8). https://doi.org/10.1016/j.asej.2025.103455