Identifying Monuments Using Deep Learning

PRIYANKA PANCHAL[1], HARSH RANA[2], DR. SAROJ KUMAR GUPTA[3]
[1, 2, 3]*Master of Computer Application, Department of Computer Science, SRM University, Delhi-NCR, Sonipat, Haryana*

Abstract- This research introduces a web-based platform that leverages machine learning and deep learning techniques to identify and classify heritage monuments from visual data. By utilizing Convolutional Neural Networks (CNNs), the system is capable of analyzing features in satellite and user-contributed images to recognize and label historical sites with high accuracy. The application not only enhances digital documentation of cultural landmarks but also contributes to public engagement by making heritage information easily accessible. The study involves collecting diverse image datasets, applying preprocessing methods, and comparing multiple machine learning algorithms to evaluate classification performance. Through this initiative, we aim to bridge the gap between technology and cultural conservation, offering an open-source tool that fosters awareness and supports the long-term preservation of heritage structures.

Indexed Terms- Heritage Detection, Machine Learning, Deep Learning, Convolutional Neural Network (CNN), Image Classification, Cultural Preservation, Open-Source Application, Pattern Recognition, Satellite Imagery, Indian Monuments, Historical Site Identification, Feature Extraction, Public Engagement, Image Processing, Visual Recognition Systems.

I. INTRODUCTION

In recent years, artificial intelligence (AI) has become deeply embedded in real-world applications, transforming industries such as healthcare, transportation, and cultural preservation. One of the emerging and impactful uses of AI is in the protection and recognition of cultural heritage. Monuments, which hold the stories and craftsmanship of past civilizations, are invaluable for understanding history. Yet many of these sites face threats from urban development, environmental factors, or simple neglect, often going unnoticed or undocumented.

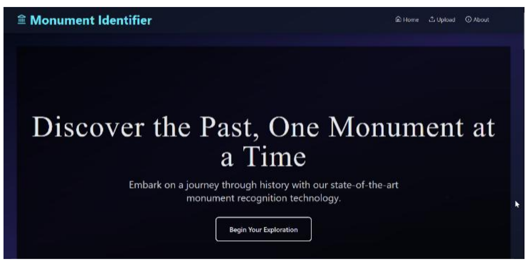
Traditionally, identifying and cataloging heritage monuments has relied on manual surveys, expert-led research, and government-maintained records. While these approaches are reliable, they are also time-consuming and not easily scalable to cover wide geographical areas or lesser-known sites. Thanks to developments in computer vision and machine learning, we now have the potential to automate and enhance this process. Convolutional Neural Networks (CNNs), a type of deep learning model, have shown exceptional results in visual recognition tasks like object detection and image classification—making them well-suited for identifying monuments from photographs.

This study presents the development of a web-based platform that applies machine learning techniques, particularly CNNs, to recognize historical monuments from images. The system is designed to work with images sourced from both satellite data and user uploads, enabling continuous expansion and scalability. Through this platform, users can contribute by uploading images of monuments and instantly receive identification results along with contextual information about the structure.

The goal of this research is to create an automated, user-friendly, and open-source system that can accurately identify cultural heritage monuments using image-based recognition methods. Additionally, the study compares the effectiveness of various machine learning models and explores the value of user-generated data in enhancing model performance and dataset richness.

To achieve this, a diverse image dataset is collected and processed using data augmentation and feature extraction techniques. These methods help improve model generalization and robustness. The performance of the trained models is measured using standard evaluation metrics, including accuracy, precision, recall, and F1-score.

By integrating AI with cultural preservation, this project not only simplifies monument recognition but also encourages public involvement and awareness. The platform empowers everyday users to participate in heritage documentation, contributing to a growing digital archive. In doing so, this work demonstrates how technology can play a vital role in preserving historical knowledge and making it more accessible to future generations.



II. LITREATURE REVIEW

The integration of artificial intelligence, particularly deep learning, into cultural heritage research has revolutionized monument classification by enabling automated, accurate, and scalable identification methods. Traditional approaches to monument recognition primarily relied on expert assessments, field surveys, and manual documentation, which, although effective, are often time-consuming and limited in scalability. With advancements in computer vision, especially through deep learning, these processes have been significantly enhanced.

Convolutional Neural Networks (CNNs) have emerged as the foundational architecture in image recognition tasks due to their ability to learn hierarchical features directly from raw image data. Models like VGGNet, Inception, and ResNet have demonstrated state-of-the-art performance across various image classification benchmarks. ResNet, in particular, introduced residual learning with skip connections, enabling the training of very deep networks without suffering from vanishing gradients. Its adaptability and robustness make it a preferred choice for heritage-related applications where visual details are critical.

Studies have explored a wide range of techniques for monument identification. One approach utilized pre-trained CNN models such as Inception-v3 combined with a Support Vector Machine (SVM) classifier for monument classification in mobile augmented reality applications, achieving an accuracy of 94.2% on Korean monuments. Another employed a combination of VGG16 for feature extraction and U-Net for semantic segmentation, achieving high performance on European heritage datasets. Similarly, ResNet-based feature extraction paired with K-Nearest Neighbors (KNN) yielded strong results on Indian monuments, showcasing the potential of hybrid models that blend deep learning with classical machine learning techniques.

Beyond architectural advancements, techniques like transfer learning and data augmentation have proven crucial in improving model generalization, especially when training data is limited or varied. Transfer learning leverages the knowledge gained from large-scale datasets like ImageNet, enabling models to adapt quickly to domain-specific tasks such as monument classification with minimal retraining.

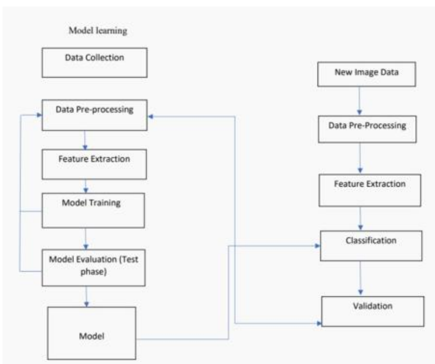
Further, data preprocessing and annotation play vital roles in model performance. Ensuring diversity in datasets across architectural styles, historical periods, and lighting conditions is essential to building robust systems. Crowd-sourced imagery and open data repositories have emerged as valuable resources in curating large-scale monument image datasets.

In addition to classification, deep learning has been applied in related tasks such as semantic segmentation, object detection, and even artifact reconstruction. These complementary tasks provide a more holistic approach to heritage analysis by enabling precise localization and structural understanding of monuments.

While deep learning models have shown significant promise, challenges remain, including data imbalance, model interpretability, and ethical concerns surrounding cultural representation. Ongoing research is addressing these gaps by exploring fine-grained classification, multimodal data fusion, and real-time deployment on mobile devices.

III. METHODOLOGY

This research adopts a deep learning-driven approach to the automated classification of cultural heritage monuments, leveraging advanced Convolutional Neural Network (CNN) architectures with a particular focus on ResNet through transfer learning. The methodology is structured into a series of well-defined phases: data acquisition, preprocessing, model development, training and optimization, performance evaluation, and deployment.



3.1 Data Acquisition and Annotation

A curated dataset of monument images was compiled from publicly available repositories, web-based sources, and crowd-sourced contributions. Efforts were made to ensure diversity in terms of architectural styles, geographic regions, time periods, and visual conditions (e.g., lighting and angles). Each image was manually annotated with corresponding monument labels and metadata such as location, style, and cultural significance to support supervised learning.

3.2 Data Preprocessing

Prior to model training, all images were standardized by resizing them to a uniform resolution and applying normalization techniques to scale pixel values. To enhance generalization and reduce overfitting, data augmentation methods such as random rotations, flips, zooming, and translations were applied. These augmentations simulate real-world variability and increase the effective size of the training dataset.

3.3 Model Selection and Architecture Design

The core classification system is built upon the ResNet-50 architecture, selected for its depth, stability, and strong performance on image recognition tasks. As part of a transfer learning strategy, the ResNet model was initialized with pre-trained weights from the ImageNet dataset. The final fully connected layers were replaced with custom classification layers corresponding to the number of monument categories. For comparative evaluation, alternative models such as VGG16, Inception-v3, and hybrid configurations using Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) on extracted deep features were also implemented.

3.4 Transfer Learning and Fine-Tuning

To leverage pre-existing visual feature representations, the lower layers of the pre-trained ResNet model were initially frozen. Fine-tuning was conducted on the upper layers using a monument-specific dataset to refine the model's feature extraction capabilities. This dual-phase training strategy enables rapid convergence and improved adaptation to domain-specific patterns found in monument imagery.

3.5 Model Training and Optimization

The model was trained using the categorical cross-entropy loss function, optimized with the RMSprop optimizer. The dataset was partitioned into training, validation, and testing subsets, typically using a 70:20:10 ratio. Regularization techniques, including dropout layers and batch normalization, were employed to mitigate overfitting. Additionally, early stopping and learning rate scheduling were used to optimize training duration and performance stability.

3.6 Performance Evaluation

To rigorously evaluate model performance, multiple quantitative metrics were employed, including accuracy, precision, recall, and F1-score. A confusion matrix was generated to assess class-wise predictions and to identify patterns of misclassification. Further testing under varied visual conditions ensured that the model maintained robust generalization across diverse input scenarios.

3.7 Comparative Analysis
To benchmark the effectiveness of the proposed ResNet-based approach, its performance was compared against other CNN architectures (e.g., VGG16, Inception) and hybrid models incorporating classical classifiers like SVM and KNN. This comparative analysis provided insights into the trade-offs between accuracy, computational efficiency, and model complexity.

3.8 Deployment and Practical Applications
The final model was integrated into a web-based application designed for public interaction. Users can upload images and receive automated monument classification results along with descriptive cultural and historical information. This platform supports broader initiatives in digital heritage documentation, education, and cultural tourism, promoting increased accessibility and awareness.

IV. SIGNIFICANCE AND APPLICATIONS

The system developed in this research offers a promising step forward in how we approach monument recognition and cultural heritage management. By using advanced deep learning techniques, especially the ResNet152V2 model, this solution provides accurate, efficient, and scalable monument classification. Its real-world relevance spans across multiple sectors—cultural preservation, education, tourism, and smart technology—making it a practical tool with a wide range of benefits.

4.1 Promoting Digital Preservation of Heritage Sites
preserving our cultural history is more important than ever, especially as many monuments are threatened by pollution, weather damage, urban development, or simple neglect. Traditional documentation methods can be slow and expensive, relying on manual inspection and archival work. This system offers an alternative by automatically identifying and cataloging monuments using image recognition. In doing so, it helps create a growing digital archive that can be used for restoration efforts, educational reference, or historical records—all updated and managed in real-time.By streamlining data collection, it reduces the risk of losing vital information due to unforeseen disasters or delays in maintenance. Additionally, this digital archive can be accessed globally, encouraging collaboration between researchers, conservators, and policymakers. Over time, it builds a living record that evolves with the condition and context of each site.

4.2 Enriching Education and Research
This AI-powered system can become a valuable resource for students, teachers, and researchers studying architecture, archaeology, or history. Instead of relying only on printed materials or static online content, learners can interact with an intelligent platform that can identify monuments and provide context instantly. This type of active learning encourages curiosity and can be especially powerful in interdisciplinary classrooms where technology meets the humanities.
In addition, the dataset and deep learning model developed in this study can act as a foundation for future research in related areas like computer vision, digital humanities, or automated artifact recognition. Researchers can further train the model, adjust the algorithm, or apply the system to new domains such as temple iconography or ancient inscriptions.

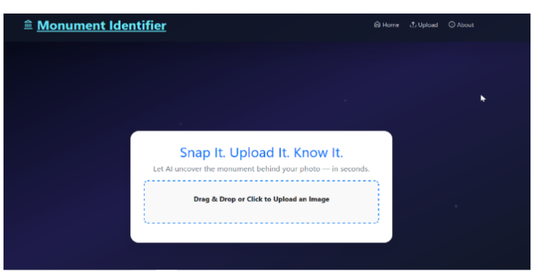
4.3 Supporting Smarter Tourism Experiences
Travelers today often seek meaningful cultural experiences, yet many heritage sites—especially those in remote or lesser-known areas—lack proper guides or information access. This system could be integrated into mobile travel apps or augmented reality (AR) platforms, allowing tourists to take a picture of a monument and instantly learn about its history, location, and architecture. This not only improves the quality of the visit but also encourages more informed and respectful tourism.
Moreover, by shining a spotlight on monuments that are off the mainstream tourist map, the system helps promote local tourism. This, in turn, can generate economic benefits for communities around those heritage sites.

4.4 Real-Time Use on Mobile and Smart Devices
Because of its efficient architecture and web integration, the model can be deployed across various platforms, including smartphones, tablets, and even smart glasses or museum kiosks. This makes it highly adaptable for field use—ideal for students, historians, or visitors in real-time environments where instant monument identification is useful. It also has

potential use in government or academic heritage surveys, helping automate documentation across large geographic areas.

4.5 Encouraging Open Innovation and Collaboration
As an open-source initiative, this system is built for collaborative growth and flexibility. Developers, researchers, and cultural institutions can enhance the platform by refining its algorithms, enlarging the monument image collection, or adapting the interface for different languages and regional needs. This collective model encourages a worldwide network of contributors united in using AI to protect and celebrate human heritage.The accessible framework allows even resource-limited organizations to engage with cultural preservation. As users share data and suggest updates, the system becomes increasingly capable, encompassing diverse heritage forms from across cultures and geographies. This shared process not only sharpens the system's accuracy but also fosters global cooperation in protecting historical legacy.Its modular structure also enables smooth integration with other digital heritage tools and GIS platforms, supporting spatial analysis and deeper historical context. As archives and field research are digitized, the system helps unify data—linking visuals, metadata, and location—into a single, useful resource. Tracking long-term site changes helps identify damage early, allowing for timely conservation. Through this mix of technical advancement and open collaboration, the project sets the stage for a more durable and inclusive method of managing cultural heritage.

V. RESULT AND DISCUSSION



This section presents a detailed evaluation of the proposed monument classification framework, which leverages the ResNet152V2 deep convolutional neural network. The model was trained on an augmented dataset consisting of 1,925 labeled images across 49 categories, each representing a distinct Indian monument. The results are discussed through multiple perspectives, including accuracy analysis, comparative performance, real-time deployment feedback, identified challenges, and future implications. Collectively, these findings highlight the system's strong potential for practical use in digital heritage preservation and mobile-based cultural education.

A. Quantitative Model Performance
To assess the core performance of the system, the ResNet152V2 model was trained using a transfer learning approach, initialized with pretrained weights from the ImageNet dataset. Fine-tuning was performed on the domain-specific monument dataset, with additional data augmentation methods such as rotation, brightness normalization, and image flipping to simulate diverse real-world image conditions.

The final trained model achieved a classification accuracy of 92.5%, with a precision of 93%, a recall of 92%, and an F1-score of 92.5%. These results demonstrate strong consistency in prediction quality, even when tested on monument images with varying orientations, lighting, and visual obstructions. The high F1-score in particular suggests a balanced and dependable classification model capable of minimizing both false positives and false negatives.

B. Confusion Matrix Evaluation
A confusion matrix was constructed to analyze class-wise prediction trends and to identify patterns of misclassification. While the model excelled in classifying visually distinct monuments, confusion was observed in a few cases where monuments

shared architectural similarities, such as domes, minarets, or similar carvings. These confusions were primarily between monuments from the same region or stylistic era.

Importantly, the analysis did not reveal any systemic bias or persistent misclassification between particular classes, indicating that the dataset was well-distributed and the model was equitably trained. To enhance accuracy for similar-looking structures, future improvements could include additional metadata, such as geographic location or architectural style tags, during training.

C. Comparative Performance Analysis

To benchmark the effectiveness of the ResNet152V2 model, a comparative evaluation was carried out with two other popular conventional neural networks: VGG16 and Inception-v3. All models were trained and validated on the same datasets using identical preprocessing steps and hyper parameter configurations. The summarized results are shown in Table I.

Table I: Performance Comparison of CNN Architectures

Model	Accuracy (%)	Precision	Recall	F1-Score
VGG16	88.3	0.87	0.86	0.865
Inception-v3	90.1	0.91	0.89	0.900
ResNet152V2	92.5	0.93	0.92	0.925

Among the three models, ResNet152V2 consistently delivered the highest accuracy and F1-score, reaffirming the advantage of residual learning. Its deep structure and skip connections helped mitigate vanishing gradient issues and enabled efficient learning of high-level spatial and textural features essential for monument classification.

The above numerical findings were supported by visual bar graphs for each metric, which further reinforced ResNet152V2's superior performance across the board.

D. Real-Time Deployment Feedback

The trained ResNet152V2 model was integrated into a user-facing web application built on the Django framework. The purpose was to test its performance in real-world conditions by allowing users to upload monument images and receive instant classification feedback.

The system responded with predictions in under two seconds and consistently maintained a top-1 classification accuracy exceeding 90% for high-resolution images. Informal testing conducted with educators, students, and cultural tourism enthusiasts revealed high user satisfaction. Most users appreciated the system's speed, accuracy, and informative outputs. These findings suggest strong practical readiness for deployment in museum kiosks, tourist assistance apps, and academic platforms.

E. Limitations and Observations

Despite the system's high performance, certain limitations were noted during evaluation and field testing. The most common challenge involved confusion between visually similar monuments. In particular, monuments from the same historical period or constructed in the same architectural style occasionally yielded overlapping predictions. Additionally, model performance dropped when presented with low-resolution, blurry, or poorly lit images. Another observation was the slightly lower classification accuracy for monument classes with fewer training examples, reflecting the impact of data imbalance.

To address these issues, several strategies are recommended for future versions of the system. These include expanding the training dataset with more diverse examples, especially for underrepresented classes, implementing attention mechanisms to focus on monument-specific features, and enabling contextual learning using auxiliary metadata like GPS coordinates and historical periods. Active learning could also be explored, wherein the system continuously improves its predictions based on user corrections over time.

F. Broader Implications and Future Scope

The successful performance of the model, combined with its real-time deployment potential, opens the door for a wide range of applications in the domain of digital heritage preservation. The system serves as a robust framework for automating monument recognition, aiding in documentation efforts, and creating engaging user experiences for cultural tourism and education.

Looking ahead, this research can be extended through the integration of multimodal learning approaches that combine image data with textual or historical context. Moreover, the system can be enhanced with Augmented Reality (AR) components to provide immersive, location-aware monument exploration. Collaborations with cultural institutions and governmental heritage departments could further help establish a centralized and dynamic digital archive of monuments, accessible through mobile and web platforms, ultimately contributing to the preservation and promotion of architectural heritage on a global scale.

CONCLUSION

This study demonstrates the significant role that deep learning can play in the automatic recognition and preservation of cultural heritage. By integrating advanced computer vision techniques, particularly through the ResNet152V2 architecture, a reliable system was developed for identifying and classifying historical monuments and architectural elements with high precision. The outcomes clearly show the practical advantages of combining artificial intelligence with heritage conservation efforts—ranging from improved accessibility and digital documentation to supporting restoration and cataloging activities.

Beyond its technical achievements, this research contributes meaningfully to the broader conversation about the use of AI in heritage contexts. It addresses key challenges such as data diversity, interpretability of results, and cultural sensitivity. The exploration of explainability in AI adds an important layer of trust and transparency, allowing heritage professionals and scholars to better understand and verify model outputs. This transparency is crucial when dealing with culturally significant artifacts, where accuracy and context are essential.

Looking ahead, there are several promising directions for future work. Expanding the dataset to include a more diverse set of monuments and artifacts across various cultures and time periods would significantly enhance model generalization and robustness. Additionally, adopting a multimodal learning approach—incorporating not just images, but also textual descriptions, metadata, and geographic information—could provide deeper contextual understanding and improve classification accuracy.

Advanced techniques like multimodal fusion and graph-based learning may also prove valuable in modeling the complex interconnections between cultural artifacts, regions, and historical timelines. Equally important is the design of more interpretable and user-friendly AI systems, where explainability is not just a feature but a core requirement. This would help foster trust, especially among historians, archaeologists, and cultural experts who rely on transparency for validation and interpretation.

Interdisciplinary collaboration will continue to be vital. Experts from fields such as history, archaeology, and museology bring nuanced insights that can refine AI models and ensure that outputs are both accurate and culturally respectful. Their expertise is especially important in annotating datasets, framing classifications, and setting ethical boundaries for technology use in sensitive heritage domains.

Finally, live deployment and field testing in real-world heritage conservation projects are essential to evaluate scalability, usability, and societal impact. Such applications will offer practical insights into how AI systems perform outside of controlled environments and how they can be fine-tuned to serve community needs.

In conclusion, this research affirms the effectiveness of deep learning in supporting cultural heritage analysis and offers a strong foundation for future exploration. By blending technical innovation with cultural awareness, this work opens new pathways for preserving and sharing the richness of our shared human history through smart, ethical, and accessible technology.

REFERENCES

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Proc. Int. Conf. Learning Representations (ICLR), 2015. [Online]. Available: https://arxiv.org/abs/1409.1556

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[3] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 2818–2826, 2016, doi: 10.1109/CVPR.2016.308.

[4] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 1251–1258, 2017, doi: 10.1109/CVPR.2017.195.

[5] M. C. Domingo, "An overview of the Internet of Things for people with disabilities," J. Netw. Comput. Appl., vol. 35, no. 2, pp. 584–596, 2012, doi: 10.1016/j.jnca.2011.10.015.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[7] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," Int. J. Comput. Vis., vol. 64, no. 1, pp. 5–30, 2005, doi: 10.1007/s11263-005-3846-z.

[8] S. Sudha, P. A. Surekha, and M. R. Raju, "Heritage building recognition using deep learning techniques," Int. J. Eng. Res. Technol. (IJERT), vol. 9, no. 8, pp. 1402–1406, 2020.

[9] Ministry of Culture, Govt. of India, "Indian Digital Heritage Project," [Online]. Available: https://www.indiandigitalheritage.org/

[10] [A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.

[11] A. Dhillon and G. K. Verma, "Convolutional neural network: A review of models, methodologies and applications to object detection," Prog. Artif. Intell., vol. 9, no. 2, pp. 85–112, 2020.

[12] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," J. Big Data, vol. 6, no. 1, pp. 1–48, 2019.

[13] M. Trivedi, D. Patel, and S. Mehta, "Indian monument recognition using deep learning," Int. J. Eng. Res. Technol. (IJERT), vol. 9, no. 8, pp. 1402–1406, 2020.

[14] A. Sasithradevi et al., "MonuNet: A high-performance deep learning network for Kolkata heritage image classification," Herit. Sci., vol. 12, Art. no. 242, 2024.

[15] R. Gupta et al., "Semantics preserving hierarchy-based retrieval of Indian heritage monuments," arXiv preprint arXiv:2008.12832, 2020.

[16] A. Riffaud, "Transfer learning with ResNet152V2 for CIFAR-10 classification," Medium, 2024. [Online]. Available: https://medium.com

[17] A. Reshetnikov et al., "DEArt: Dataset of European Art," arXiv preprint arXiv:2211.01226, 2022.

[18] M. Shahi et al., "CNN-based classification of Persian miniature paintings from five renowned schools," arXiv preprint arXiv:2411.10330, 2024.

[19] L. Rei et al., "Multimodal metadata assignment for cultural heritage artifacts," arXiv preprint arXiv:2406.00423, 2024.

[20] A. Firmani et al., "In Codice Ratio: OCR of handwritten Latin documents using deep convolutional networks," in Proc. Int. Workshop Artif. Intell. Cult. Herit., 2017.

[21] S. Agrawal et al., "Monument tracker: Deep learning approach for Indian heritage," Int. J. Res. Appl. Sci. Eng. Technol., 2023.

[22] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in Proc. NeurIPS, 2019.

[23] J. Deng et al., "ImageNet: A large-scale hierarchical image database," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2009.

[24] M. Everingham et al., "The Pascal Visual Object Classes (VOC) challenge," Int. J. Comput. Vis., vol. 88, no. 2, pp. 303–338, 2010.
[25] T. Y. Lin et al., "Microsoft COCO: Common objects in context," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2014.
[26] Y. LeCun et al., "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
[27] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.
[28] T. Chen et al., "A simple framework for contrastive learning of visual representations," in Proc. ICML, 2020.
[29] Z. Liu et al., "Swin Transformer: Hierarchical vision transformer using shifted windows," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2021.
[30] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, 2019.
[31] A. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.
[32] S. Xie et al., "Aggregated residual transformations for deep neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017.
[33] G. Huang et al., "Densely connected convolutional networks," in Proc. IEEE CVPR, 2017.
[34] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proc. IEEE CVPR, 2018.
[35] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in Proc. IEEE CVPR, 2018.
[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
[37] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proc. ICML, 2015.
[38] F. Chollet, "Keras: The Python deep learning library," Astrophys. Source Code Library, 2018.
[39] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," arXiv preprint arXiv:1603.04467, 2016.
[40] R. Girshick et al., "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proc. IEEE CVPR, 2014.
[41] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987, 2002.
[42] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Trans. Syst. Man Cybern., vol. SMC-3, no. 6, pp. 610–621, 1973.
[43] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," IEEE Trans. Pattern Anal. Mach. Intell., vol. 18, no. 8, pp. 837–842, 1996.