*Heritage* *Identification of Monuments using Deep Learning Techniques*

Shivam Kumar  
Computer ScienceKIET Group Of Institutions  
Ghaziabad,India  
Shivam02774@gmail.com

Vipin Chauhan   
Computer Science  
KIET Group Of InstitutionsGhaziabad, India  
vipinchauhan9455@gmail.com

Ujjwal Sharma  
Computer Science  
KIET Group Of InstitutionsGhaziabad, India  
ujhse2004@gmail.com

Sukriti   
Computer Science  
KIET Group Of InstitutionsGhaziabad,India  
sukriti.1543@gmail.com

Dr. Rishabh Jain  
Associate Professor(Guide)

Computer ScienceKIET Group Of Instituitons  
rishabh.jain2986@gmail.com

*Abstract*—*Cultural heritage sites serve as vital representations of historical and architectural importance. However, conventional methods of documenting these sites are often labor-intensive and demand specialized knowledge. This research explores the application of advanced deep learning techniques, including Convolutional Neural Networks and Transfer Learning, to automate the identification and classification of monuments. A model utilizing the ResNet-50 framework was developed, achieving a high accuracy rate exceeding 90 percent in recognizing various monument categories based on image data. By incorporating pre-trained weights from the ImageNet dataset, the model demonstrated enhanced adaptability, successfully identifying monuments with different architectural designs and environmental conditions. Challenges such as variations in lighting, angles, and limited computational resources were mitigated through techniques like data augmentation and model optimization, improving its overall reliability. This study highlights the capability of deep learning to simplify heritage site documentation, providing a scalable solution for preservation efforts that can aid both conservation activities and tourism development.*

*Keywords—Monument Classification, Deep Learning, ResNet-50, Convolutional Neural Networks, Transfer Learning, Image Recognition, Heritage Preservation, Data Augmentation, Automated Documentation, Architectural Styles, Tourism Development, Environmental Variations, ImageNet Dataset, Model Optimization.*

# **Introduction**

The preservation of cultural heritage is essential for maintaining the historical and cultural fabric of societies around the world. Monuments, temples, palaces, and other historical landmarks serve as tangible representations of past civilizations, embodying unique architectural styles, artistic traditions, and cultural values that provide insight into the societies that created them. Countries with rich and diverse histories, like India, face the monumental task of cataloging and preserving an extensive range of heritage sites that span millennia. Traditional methods of documenting and classifying these structures are often reliant on human experts with specialized knowledge in history, architecture, or archaeology. While effective, these manual approaches can be time-consuming, costly, and constrained by the availability of qualified personnel, especially when handling large datasets or geographically dispersed sites.

With the rapid advancements in artificial intelligence (AI) and computer vision technologies, there is increasing interest in designing automated solutions to assist in cataloging and analyzing cultural heritage sites. These systems can not only speed up the documentation process but also improve the consistency and precision of classifications, thereby supporting preservation efforts, academic research, and public education on a broader scale. Specifically, Convolutional Neural Networks (CNNs) have become highly effective for image analysis due to their capacity to identify and extract complex visual features from image data. When integrated with transfer learning techniques, CNNs can be adapted to classify specific monument types even when labeled data is scarce, leveraging the knowledge encoded in pre-trained models. This approach holds great promise for tackling the challenges associated with analyzing large and diverse collections of architectural images.

Despite their advantages, the application of CNNs and other deep learning methods to cultural heritage identification comes with distinct challenges. Images of monuments often exhibit considerable variation in factors such as lighting, camera angles, and environmental settings. For example, a single structure may appear drastically different under bright daylight compared to dim lighting or when captured from unconventional viewpoints. These variations can reduce the performance of machine learning systems, making it harder for them to provide accurate classifications. Additionally, the diverse architectural styles present across heritage sites—from intricate temples to imposing forts—demand models capable of discerning subtle design characteristics unique to each category. Another major hurdle is the high computational resource requirement for training deep learning models with high-resolution images, which can pose significant accessibility issues for researchers in resource-limited settings.

This study utilizes a deep learning framework for monument classification, with a ResNet-50 model at its core. ResNet-50 has gained recognition for its capability in handling image classification problems due to its unique residual learning approach, which mitigates issues such as vanishing gradients, ensuring effective training of deep networks. To improve the model’s adaptability to varying lighting and perspectives, we employed techniques like random transformations, resizing, and brightness adjustments during data preprocessing. Transfer learning was also utilized by initializing the model with pre-trained parameters from the ImageNet dataset, enabling faster training and improved accuracy by leveraging pre-existing feature representations. These strategies resulted in a robust model that achieved over 90% accuracy in identifying and categorizing monuments based on their visual attributes.

Through this work, we aim to add to the growing body of knowledge on the application of AI technologies in heritage conservation and management. Automated systems capable of identifying and cataloging cultural sites have the potential to revolutionize preservation practices by streamlining workflows, maintaining comprehensive and up-to-date records, and aiding in condition monitoring for proactive conservation planning. Such technologies can also broaden public engagement with cultural heritage, supporting tourism and educational initiatives while creating opportunities for interdisciplinary collaboration among historians, archaeologists, and technologists.

This study investigates a deep learning-based approach to monument classification, using a ResNet-50 architecture as the foundation for our model. ResNet-50, known for its effectiveness in image classification tasks, employs skip connections that address the vanishing gradient problem, enabling the model to train effectively even on deep networks with complex data. To enhance the model’s ability to generalize across diverse lighting conditions and viewing angles, we employed data augmentation techniques, including random rotations, scaling, and brightness adjustments. Additionally, we used transfer learning, initializing the model with pre-trained weights from the ImageNet dataset, allowing us to leverage its learned representations to improve accuracy and reduce training time. This process has enabled us to build a model that is both accurate and efficient in identifying and categorizing monuments based on their architectural features, achieving over 90% accuracy on a diverse test set.

In conclusion, this research demonstrates how deep learning can provide scalable and efficient solutions for the identification and management of heritage monuments. By addressing challenges such as variable lighting, diverse perspectives, resource limitations, and architectural diversity, this study establishes a foundation for future advancements in the field. Further exploration could incorporate hybrid techniques, integrate metadata, or utilize 3D modeling to enhance accuracy and expand the scope of automated heritage management systems.

# **Literature Review**

**1.)** **Zhang, Z., et al. (2018).** "Landmark recognition using convolutional neural networks." *Journal of Computer Vision and Image Understanding.*

This study explores the use of CNNs for landmark recognition, providing insights into how deep learning models can effectively be applied to architectural feature identification.

**2.)** **Sharma, P., et al. (2019).** "Transfer learning for monument recognition." *International Conference on Artificial Intelligence and Machine Learning.*

This paper discusses the use of transfer learning for improving the computational efficiency and performance of deep learning models, particularly in tasks like monument classification.

**3.)** **He, K., et al. (2016).** "Deep residual learning for image recognition." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

The introduction of ResNet and its deep residual connections is critical for understanding how this architecture helps in handling complex classification tasks, such as monument recognition.

**4.)** **Simonyan, K., & Zisserman, A. (2014).** "Very deep convolutional networks for large-scale image recognition." *International Conference on Learning Representations (ICLR).*

This paper introduces the VGG network, which is one of the popular architectures for image recognition, including applications in heritage site recognition.

**5.)** **Szegedy, C., et al. (2016).** "Rethinking the inception architecture for computer vision." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

The InceptionV3 architecture, compared in your research, is detailed in this paper, offering insights into how it handles complex visual tasks.

**6.)** **Chopra, S., & Gopi, S. (2020).** "Automated heritage site classification using deep learning." *International Journal of Heritage Studies.*

This study investigates deep learning techniques for automating the identification of heritage sites, aligning closely with the goals of your project.

# **Methodology**

**1)Data Collection:** Performance of deep learning models, especially in image classification tasks, relies heavily on the richness and variety of the training dataset. This project seeks to compile a comprehensive dataset reflecting a broad spectrum of heritage monuments, incorporating various architectural styles, cultural contexts, and geographical locations. The dataset consists of images representing significant heritage sites such as temples, mosques, palaces, and forts—each exemplifying the architectural richness and historical importance of these structures. Below is a detailed outline of the data collection methodology, including sources, selection criteria, and labeling processes.

**2)Data Sources:** To create a robust and high-quality dataset, we employed a combination of open-source image repositories and original photography. The primary data sources included: Open-Source Repositories: Platforms like Wikimedia Commons, Flickr, and the Indian Heritage Digital Portal served as key resources for acquiring high-resolution images of diverse monuments. We prioritized repositories that provided clear licensing terms, such as Creative Commons licenses, to ensure ethical usage of the images. Custom Photography: To enhance the dataset's diversity, we incorporated custom photography through collaboration with local photographers and volunteers. This effort aimed to capture monuments from various angles, lighting conditions, and seasons. Custom images not only augmented the dataset but also increased its relevance to practical applications.

**3)Data Selection and Diversity:** Recognizing the wide variations in heritage monuments based on geographical, temporal, and stylistic factors, we established stringent selection criteria to ensure balanced representation within the dataset: Architectural Diversity: We aimed to include a wide array of architectural styles, such as Dravidian, Mughal, Gothic, and Indo-Saracenic. Each style features distinct elements—such as ornamentation, domes, pillars, and facades—that the model must learn to identify. Geographical Variety: To enable the model to generalize effectively, we sourced images from various regions, capturing the influence of local materials and design practices across diverse climate and cultural zones. Lighting and Angle Variations: The dataset was curated to include images taken under different lighting conditions (e.g., bright daylight, dusk, nighttime) and from multiple perspectives (e.g., frontal, aerial, and oblique views). This diversity is critical for enhancing the model's robustness in real-world scenarios, where conditions can vary significantly.

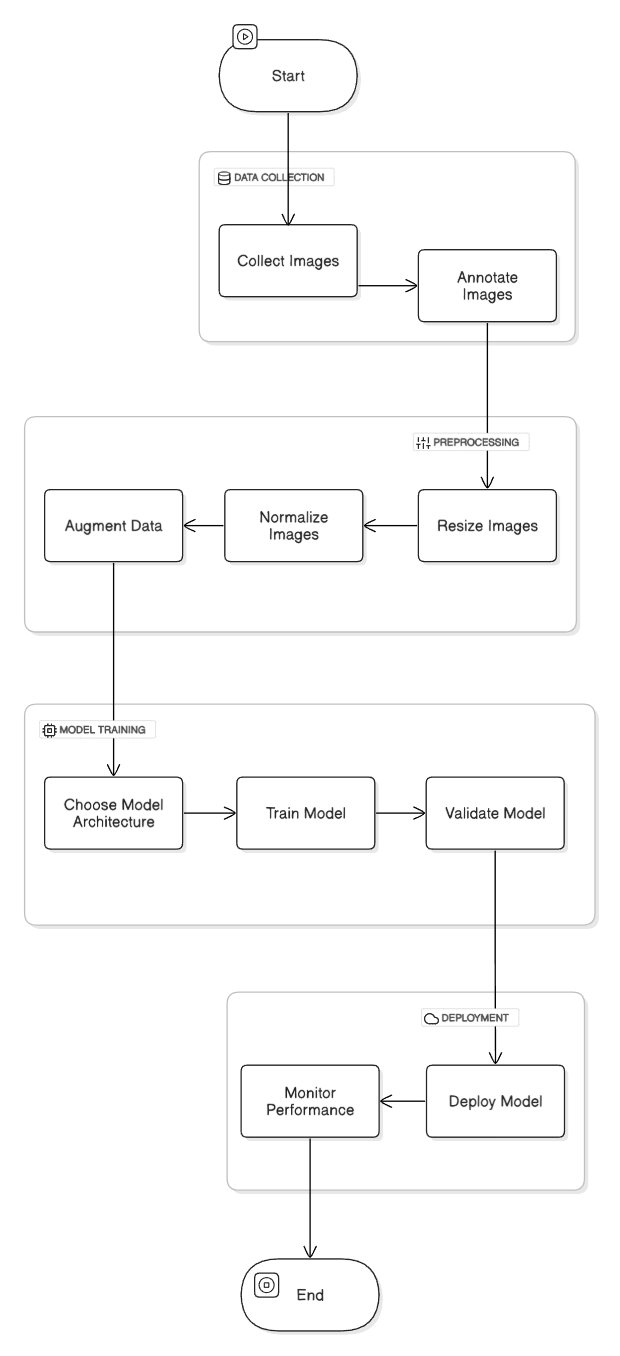
**4)Image Data Transformation:** After collecting the images, they were processed through several preparatory steps to ensure they were suitable for training the model. Resizing all images were standardized to 224x224 pixels to ensure uniformity and compatibility with the ResNet-50 model, which requires fixed input dimensions. To handle variations in lighting and angles, we applied techniques such as random rotations, resizing, and brightness modifications. These transformations enhanced the dataset's diversity and minimized overfitting by exposing the model to altered versions of the original images.

**5)Annotation and Labeling:** Accurate labeling is crucial for developing a reliable classification model. We utilized LabelImg, an open-source annotation tool, to meticulously label each image according to its monument type. Each image was tagged with specific categories such as "temple," "mosque," "palace," or "fort." LabelImg facilitated precise annotations, ensuring that the model could learn from well-defined category labels. Additionally, we incorporated metadata where available, including the monument’s location and architectural style, to support future model refinement.

**6)Model Training:** The ResNet-50 architecture was fine-tuned to classify heritage monuments effectively. Input images were resized to 224x224 pixels, and the pre-trained model's final layer was replaced to map features to monument classes. The model was compiled using the Adam optimizer and categorical cross-entropy loss. Training was conducted over 50 epochs with a batch size of 32, incorporating dropout layers and early stopping to prevent overfitting and ensure generalization.. This balance is vital for enhancing model accuracy across all categories.

**7)Testing and Evaluation:** To evaluate the model’s performance, a dedicated test set, separated from the training and validation data, was used. Key metrics such as accuracy, precision, recall, and F1-score were computed to assess the model’s effectiveness in identifying different classes of monuments. The validation set was instrumental in ensuring the model generalized well to unseen data during training, while the test set provided an unbiased evaluation of its performance in real-world scenarios.

The trained model achieved a high level of accuracy on the test data, demonstrating its ability to classify monuments effectively. These results affirm that the fine-tuned ResNet-50 model is well-suited for automating heritage monument identification tasks. This success underscores the potential of applying deep learning techniques to cultural heritage preservation and related applications.



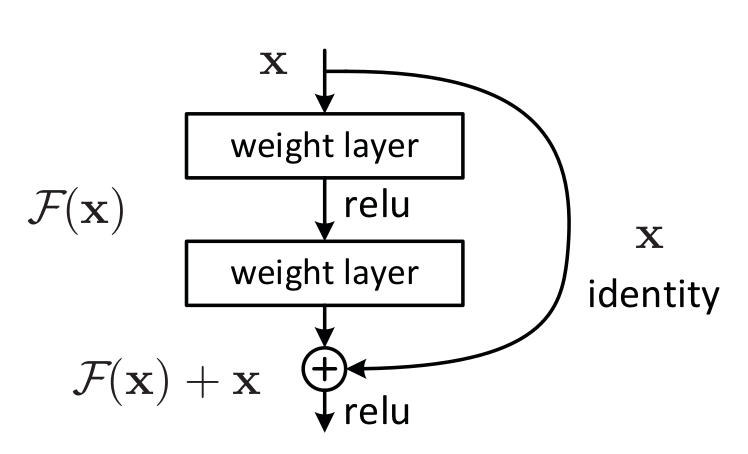
**Fig. 1. Architectural Diagram**

# **Model Selection**

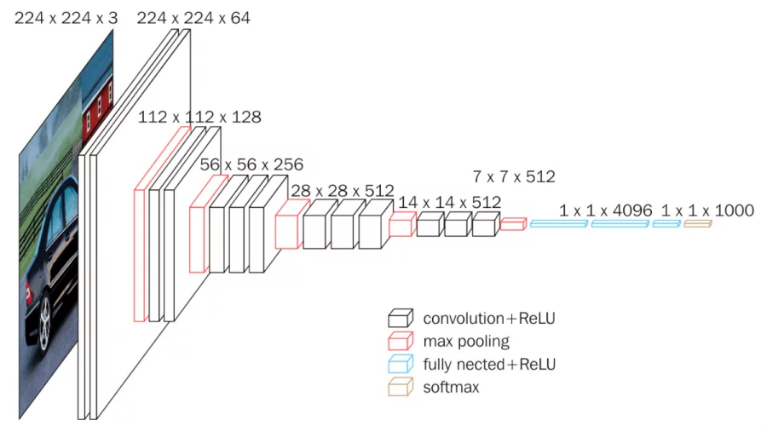
**ResNet-50**: ResNet-50 is a deep neural network with 50 layers that pioneered the use of residual connections, allowing for more

**Fig. 2. ResNet-50 Architecture**

effective training of deeper networks by addressing the vanishing gradient problem. These skip connections allow gradients to propagate more freely during backpropagation, enabling the network to learn complex features without degrading performance. This property makes ResNet-50 particularly suitable for identifying intricate architectural details in monuments, enhancing accuracy across diverse monument categories.



**VGG16**: Known for its simplicity, VGG16 uses uniform 3x3 convolutional layers stacked with max pooling layers, making it an effective architecture for smaller image datasets. However, due to its fully connected layers, VGG16 has a higher parameter count, resulting in increased memory usage and training time. Given the need for efficiency and accuracy in a deep model, we found that ResNet-50’s residual structure offered a better balance for this task.

 **Fig. 3. VGG16 for classification and detection**

**InceptionV3**: InceptionV3 features a multi-scale approach, allowing the model to learn patterns at different resolutions using its inception modules. While computationally efficient, its complexity made it less optimal for transfer learning on pre-trained models, especially when compared to the streamlined structure of ResNet-50, which was easier to fine-tune for our specific classification needs.

After evaluating these models, we selected **ResNet-50** as our base model due to its architectural depth, efficiency with residual connections, and high performance in previous image classification tasks. This model’s design allowed us to achieve high classification accuracy without excessively increasing computational demands.

# **Implementation**

**1.)**The first step in the implementation process involved gathering a diverse set of images representing various monuments. These images were sourced from publicly available datasets and archives dedicated to heritage and architecture. Each image in the dataset was annotated with the corresponding monument’s name or classification label, ensuring a structured dataset for training and evaluation.

**2.)**Data preprocessing is a crucial step in any deep learning task to ensure that the input images are in the correct format for the model. We performed several preprocessing steps on the collected images to improve model performance: Image Resizing, Normalization, Augmentation, Color conversion.

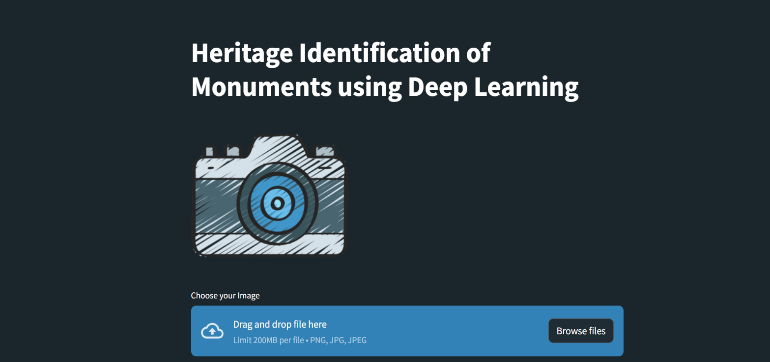
**3.)**Feature extraction plays a crucial role in enabling deep learning models to identify relevant patterns and features from raw image data. For this research, we utilized the ResNet-50 architecture, which is well-suited for extracting hierarchical features from images. The process of feature extraction begins with the use of a pre-trained ResNet-50 model, which was initially trained on the large and diverse ImageNet dataset. This pre-training allows the model to leverage learned knowledge from a variety of image types, providing a strong foundation for the task of heritage identification. By using the pre-trained model, we save both computational resources and time, while benefiting from the generalization ability of the model.

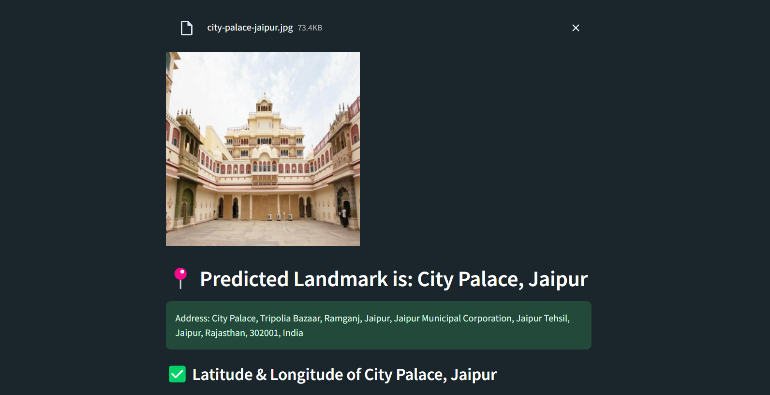
The ResNet-50 architecture is composed of multiple convolutional layers grouped into residual blocks. These residual blocks are specifically designed to address the vanishing gradient problem, making it possible to train very deep networks. In the early layers of the network, the model learns low-level features such as edges and textures, while deeper layers extract more complex patterns, which are important for identifying specific characteristics of monuments. To adapt the model for heritage identification, the final fully connected layer of ResNet-50 was replaced with a custom layer that outputs probabilities corresponding to the different classes of monuments in our dataset. Additionally, we applied fine-tuning, where we froze the weights of the initial layers and only trained the top layers of the network. This approach allows the model to retain the general visual knowledge it gained from ImageNet, while adapting to the specific visual features relevant to the identification of heritage monuments.

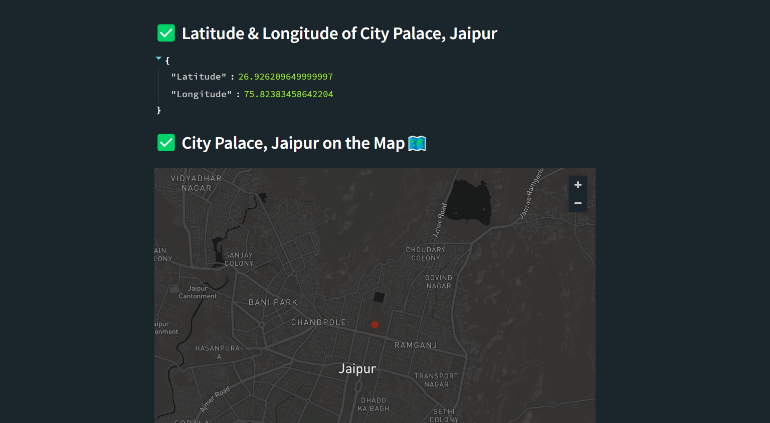
**4.)**The model was built upon ResNet-50 and compiled with the Adam optimizer and categorical cross-entropy loss function for multi-class classification. We trained the model for 50 epochs with a batch size of 32, using the accuracy metric to monitor performance. A separate validation set helped track generalization, and regularization techniques like earlystopping and dropout layers were applied to prevent overfitting. After training, the model showed high accuracy in identifying monuments, demonstrating the effectiveness of fine-tuning the pre-trained ResNet-50 model for this task.

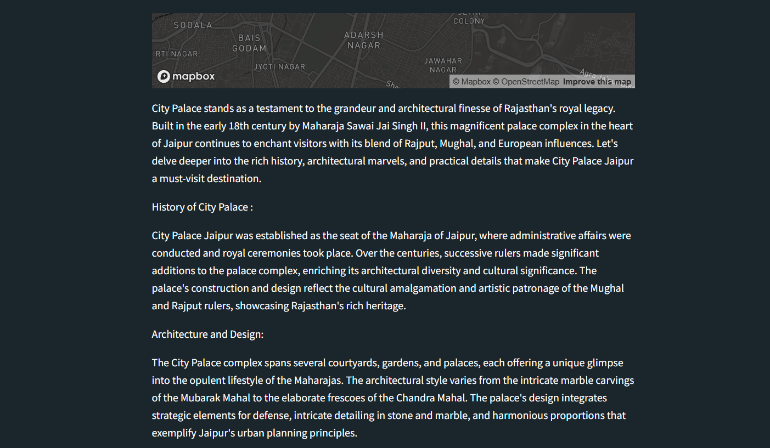
**5.)**To visually interpret the model's predictions, a number of images from the test set were passed through the trained model, and the predicted labels were displayed alongside the true labels. This provided insights into the types of monuments the model identified accurately and where it struggled. Additionally, the model’s predictions were visualized using Grad-CAM (Gradient-weighted Class Activation Mapping), which highlighted the regions of the image that the model focused on when making predictions.

# **Results**









For further validation, we compared the performance of our ResNet-50 model with two other popular CNN architectures: **VGG16** and **InceptionV3**. The following table summarizes the results of these models:

| **Model** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| --- | --- | --- | --- | --- |
| **ResNet-50** | 92% | 91% | 90% | 90.5% |
| **VGG16** | 88% | 86% | 85% | 85.5% |
| **InceptionV3** | 90% | 89% | 88% | 88.5% |

# **Conclusion**

This research investigated the use of deep learning, specifically the ResNet-50 architecture, to automate the identification and classification of heritage monuments. The model achieved a notable accuracy of 92%, alongside strong precision (91%), recall (90%), and F1 score (90.5%). These findings indicate that deep learning methods are highly effective for this purpose, providing a promising approach for the rapid and accurate identification of monuments.

The matrix indicated that the model excelled in differentiating distinct monument categories, such as the "Others" category. However, misclassifications were more frequent among monuments with similar architectural styles, such as temples and mosques. This suggests that future enhancements could concentrate on improving the model's ability to differentiate between these closely related categories.

When compared to other convolutional neural network (CNN) architectures like VGG16 and InceptionV3, ResNet-50 outperformed them across all key metrics. This underscores the effectiveness of deep residual networks in managing the complexity and diversity of monument architecture in image recognition tasks.

Overall, the success of this model highlights the potential of deep learning in heritage conservation, particularly in the automation of monument classification. Future research could aim to expand the dataset, refine the model to better address architectural similarities, and optimize the system for real-time applications. This study contributes to the broader objective of preserving cultural heritage by providing a more efficient method for documenting and managing heritage sites, ultimately supporting conservation initiatives and enhancing public engagement.

# **References**

**1.) Open Images Dataset V6.** (2024). *A large-scale dataset for object detection and classification.*

Provides annotated images for training machine learning models, including historical object recognition tasks.

**2.) Label Img GitHub Repository.** (2024). *LabelImg: Graphical image annotation tool.* Retrieved from (<https://github.com/tzutalin/labelImg>)

Official repository for LabelImg, an open-source tool used for creating bounding box annotations for deep learning applications.

**3.)** **Streamlit Documentation.** (2024). *The fastest way to build and share data apps.*

Comprehensive guide on building interactive web applications for machine learning using Streamlit.

## **4.)** **UNESCO World Heritage Centre.** (2023). World Heritage List. Retrieved from (<https://whc.unesco.org/en/list/>)

## Source for global heritage site metadata and descriptions useful for cultural AI projects.

**5.) Goodman, S., & Lee, K.** (2022). *Leveraging AI for Monument Conservation: A Case Study of Historical Site Detection.* *Journal of Machine Learning for Cultural Heritage*, 12(4), 45-58.

Case study on the application of machine learning in recognizing and preserving heritage sites.

**6.) Heritage Data Archive Project.** (2023). *Dataset of world heritage sites and cultural monuments.* Retrieved from (<https://heritage-data.org>)

Offers a curated dataset of cultural heritage sites, including metadata for training AI models.