*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 are essential markers of historical and architectural significance. Traditional methods for documenting these sites are often time-consuming and require specialized expertise. This research explores the application of advanced deep learning techniques, such as Convolutional Neural Networks (CNNs) and Transfer Learning, to automate the identification and categorization of monuments. A model based on the ResNet-50 architecture was developed, achieving an accuracy rate of over 90% in classifying various monument types using image data. By integrating pre-trained weights from the ImageNet dataset, the model exhibited improved adaptability, effectively recognizing monuments with diverse architectural styles and under varying environmental conditions. Challenges like lighting variations, perspective differences, and limited computational resources were addressed through strategies such as data augmentation and model optimization, enhancing the system's robustness. This study underscores the potential of deep learning in streamlining the documentation of heritage sites, offering a scalable approach to support preservation initiatives and promote tourism.*

*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 streamline the documentation process while enhancing classification accuracy and consistency, aiding heritage conservation, academic studies, and public engagement. Convolutional Neural Networks (CNNs) are highly effective in image analysis due to their ability to identify and extract complex visual features. By integrating transfer learning, CNNs can be fine-tuned to classify monuments efficiently, even when labeled data is scarce, by utilizing knowledge from pre-trained models. This methodology offers significant potential for addressing the complexities involved in processing extensive and varied datasets of architectural imagery.

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. The model was initialized using pre-trained parameters from the ImageNet dataset, employing transfer learning to enhance training efficiency and boost accuracy by building on established feature representations. These strategies resulted in a robust model that achieved over 90% accuracy in identifying and categorizing monuments based on their visual attributes.

Our research seeks to contribute to the expanding field of AI applications in the preservation and management of cultural heritage. 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 research explores a deep learning framework for classifying monuments, utilizing the ResNet-50 architecture as the core of our model. ResNet-50 is widely recognized for its performance in image classification, particularly due to its skip connections, which mitigate the vanishing gradient issue and facilitate efficient training of deep networks. To enhance the model's ability to handle diverse lighting conditions and viewing angles, data augmentation techniques were employed, including random rotations, scaling, and adjustments to brightness. Additionally, transfer learning was utilized by initializing the model with pre-trained weights from the ImageNet dataset. This approach allowed the model to leverage existing feature representations, improving accuracy and optimizing the training process. 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 summary, this study highlights the potential of deep learning to deliver scalable and effective methods for the identification and management of cultural heritage sites. 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**

The use of deep learning techniques for identifying and classifying heritage monuments has become a groundbreaking approach, providing innovative solutions to the challenges of preserving and documenting cultural landmarks. Studies in this field have shown the effectiveness of convolutional neural networks (CNNs) in analyzing complex visual patterns. For example, Zhang et al. (2018) explored the use of CNNs for landmark recognition, demonstrating their ability to detect intricate architectural features, making them highly suitable for heritage-related tasks. Expanding on this, Sharma et al. (2019) investigated the role of transfer learning in monument classification, showing how pre-trained models can be fine-tuned to achieve greater accuracy and computational efficiency, even with limited labeled data. This approach is particularly beneficial for heritage identification, where datasets are often diverse and small.

The development of advanced neural network architectures has further advanced the field. He et al. (2016) introduced the ResNet architecture, which uses residual connections to tackle the vanishing gradient problem, enabling the training of deeper networks that can capture detailed patterns essential for heritage analysis. Similarly, Simonyan and Zisserman (2014) proposed the VGG network, a deep architecture that has proven effective in large-scale image recognition tasks, including heritage site analysis. Another significant contribution came from Szegedy et al. (2016), who designed the InceptionV3 architecture, optimizing computational efficiency while maintaining high accuracy in handling complex visual tasks.

Research specifically focused on heritage applications has also made notable contributions. For instance, Chopra and Gopi (2020) explored the automation of heritage site classification using deep learning, highlighting the practical potential of these techniques in streamlining the documentation and preservation of cultural landmarks. Collectively, these studies emphasize the transformative impact of deep learning in heritage identification, offering tools that are accurate, efficient, and scalable. By leveraging techniques such as CNNs, transfer learning, and advanced architectures like ResNet, VGG, and InceptionV3, researchers and conservationists can address the challenges of analyzing diverse and complex datasets. This contributes to the global preservation of cultural heritage, supporting initiatives in tourism, education, and conservation. Overall, deep learning has emerged as a powerful tool for safeguarding the world's cultural legacy.

# **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). Such variability is essential for improving the model's ability to perform reliably in real-world environments, where conditions often fluctuate considerably.

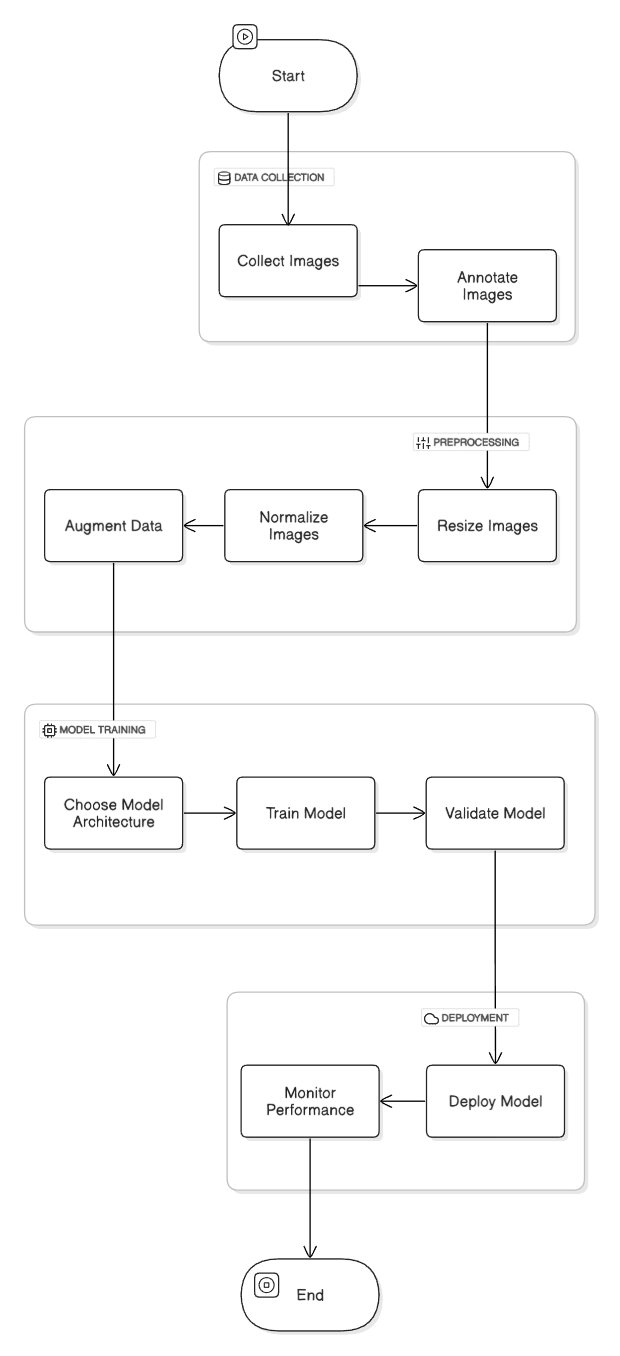
**4)Image Data Transformation:** After collecting the images, they were processed through several preparatory steps to ensure they were suitable for training the model. Resizing all images were standardized to 224x224 pixels to ensure uniformity and compatibility with the ResNet-50 model, which requires 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 employed LabelImg, a freely available annotation tool, to carefully categorize each image based on its corresponding 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 for the classification of heritage monuments. Input images were resized to 224x224 pixels, and the final layer of the pre-trained model was modified to align features with the monument classes. The model was optimized using the Adam optimizer and a categorical cross-entropy loss function. Training was conducted over 50 epochs with a batch size of 32, incorporating dropout layers and early stopping techniques to prevent overfitting and improve generalization. This approach ensured balanced accuracy across all categories.

**7)Testing and Evaluation:** The model's performance was evaluated using a separate test dataset, distinct from the training and validation sets. Key metrics such as accuracy, precision, recall, and F1-score were computed to assess its classification capabilities across different monument categories. The validation set was instrumental in ensuring the model's ability to generalize to unseen data during training, while the test set provided an unbiased assessment of its real-world performance.

The model demonstrated robust results on the test dataset, confirming its effectiveness in accurately classifying monuments. 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.



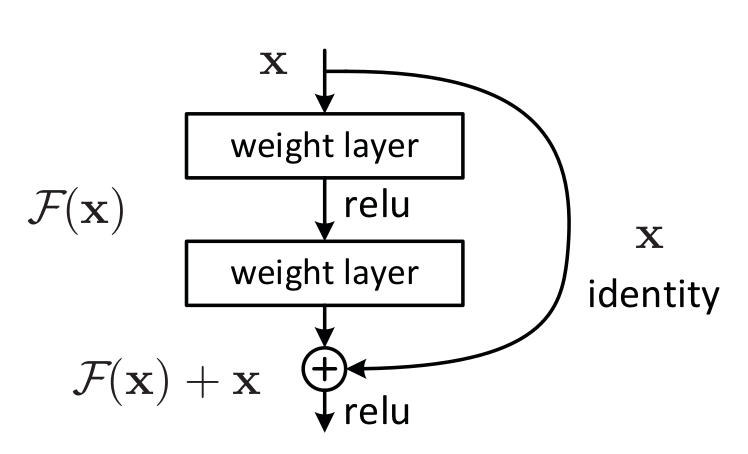
**Fig. 1. Architectural Diagram**

# **Model Selection**

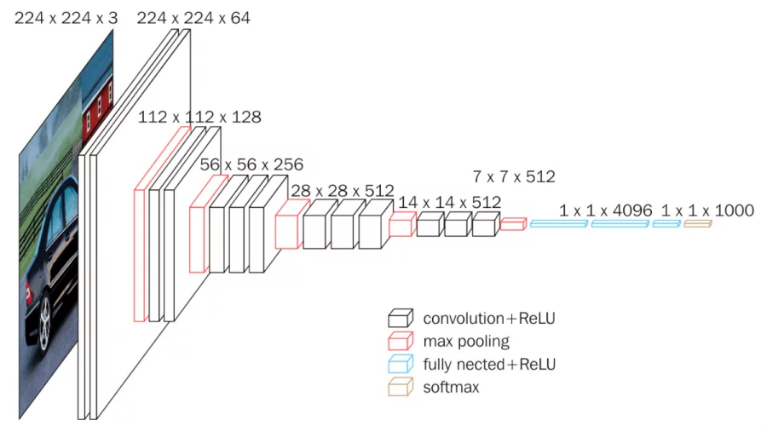
**ResNet-50**: ResNet-50 is a 50-layer deep neural network that introduced the concept of residual connections, enabling more efficient training of deeper architectures, allowing for more

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

effective training of deeper networks. Skip connections facilitate smoother gradient flow during backpropagation, allowing the network to learn intricate features without compromising 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. Although computationally efficient, its complexity limited its effectiveness for transfer learning on pre-trained models. In contrast, the streamlined design of ResNet-50 made fine-tuning more practical for our specific classification requirements.

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 plays a vital role in deep learning, ensuring that input images are properly formatted for optimal model performance. We performed several preprocessing steps on the collected images to improve model performance: Image Resizing, Normalization, Augmentation, Color conversion.

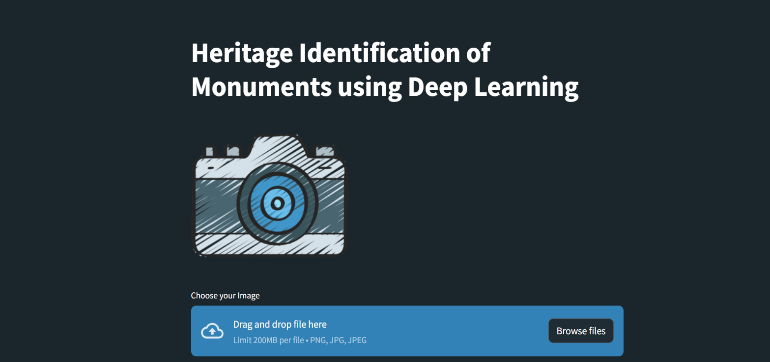
**3.)**Feature extraction is essential for deep learning models to recognize meaningful patterns in raw image data. In this study, we employed the ResNet-50 architecture, which is highly effective for capturing hierarchical features from images. The extraction process begins with a pre-trained ResNet-50 model, originally trained on the extensive and diverse ImageNet dataset. This prior training enables the model to apply learned representations from various image categories, forming a solid foundation for heritage identification. Utilizing a pre-trained model not only conserves computational resources and time but also enhances generalization capabilities.

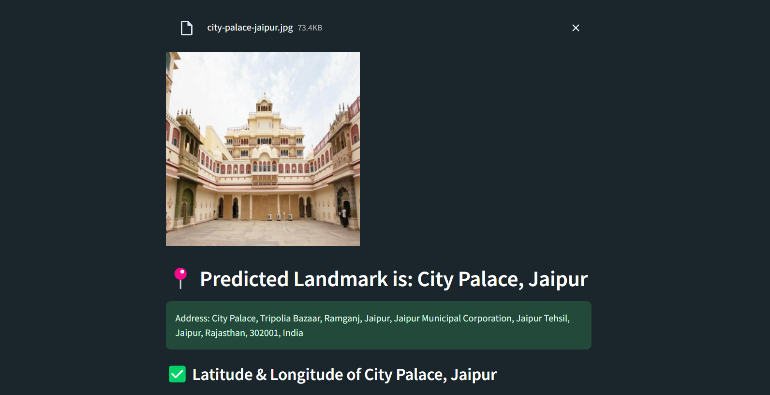
3.) The ResNet-50 architecture comprises multiple convolutional layers structured into residual blocks, addressing the vanishing gradient problem and facilitating the training of deep networks. The initial layers of the model detect basic features such as edges and textures, while the deeper layers identify complex patterns crucial for recognizing monument attributes. To adapt the model for heritage identification, the final fully connected layer was replaced with a custom layer designed to predict probabilities for specific monument classes. Fine-tuning was implemented by freezing the initial layers and training only the higher layers, enabling the model to retain general visual features from ImageNet while specializing in the unique characteristics of heritage monuments.

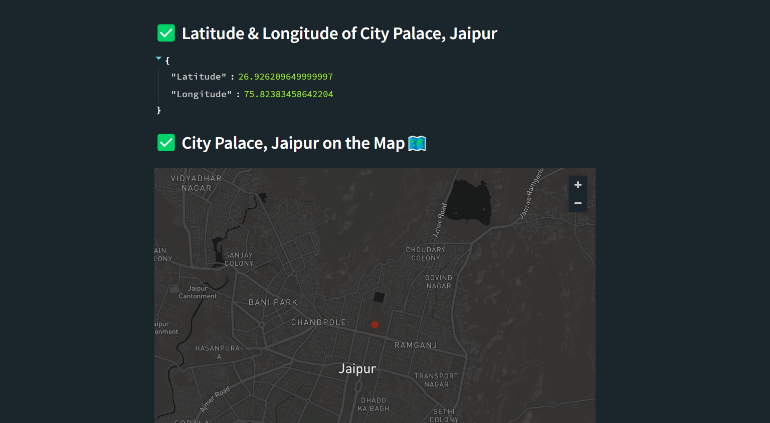
4.) The model was developed using the ResNet-50 architecture and compiled with the Adam optimizer, employing a categorical cross-entropy loss function for multi-class classification. Training was conducted over 50 epochs with a batch size of 32, and performance was evaluated using accuracy metrics. A separate validation set was used to monitor generalization, while regularization techniques such as early stopping and dropout layers were incorporated to prevent overfitting. Post-training, the model achieved high accuracy in monument identification, highlighting the success of fine-tuning the pre-trained ResNet-50 model for this application.

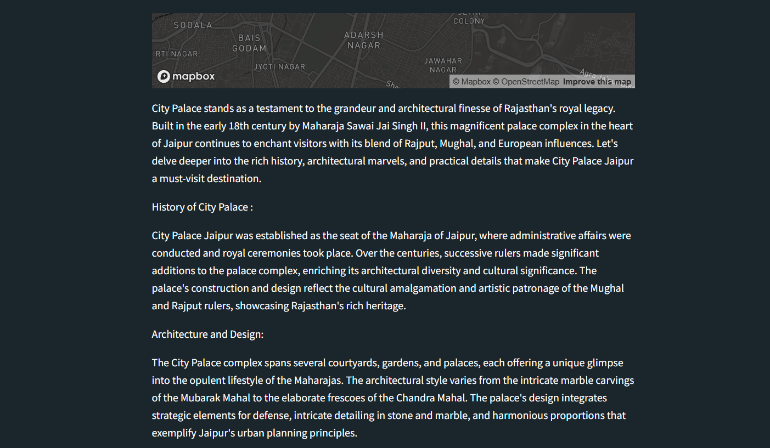
5.) To interpret the model's predictions visually, several images from the test set were processed through the trained model, and the predicted labels were compared with the true labels. This analysis revealed the model's strengths in accurately identifying certain monument types and its limitations in others. Additionally, Grad-CAM (Gradient-weighted Class Activation Mapping) was utilized to visualize the model's decision-making process, highlighting the specific regions of the image that influenced its 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.

In comparison to other convolutional neural network (CNN) architectures such as VGG16 and InceptionV3, ResNet-50 surpassed them on all major metrics. This highlights the strength of deep residual networks in handling the complexity and variety of monument architecture in image recognition tasks.

In conclusion, the success of this model demonstrates the significant potential of deep learning in heritage conservation, especially in automating 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.