Advances in Consumer Research
https://acr-journal.com/

Volume-2 | Issue-5 | Nov 2025

Original Researcher Article

Comparing ResNet18 and Swin Transformer for Classifying Indian Temple Architecture

Vanishree Kavi Mahesh[1], Dr. Udai Shankar[2], Dr. Lubna Ansari[3] and Dr.Abhilash C B[4]
[1]Research Scholar (Corresponding Author) Department of Computer Engineering and Applications, Mangalayatan University, Aligarh, India.
Email-id: 20221377_vanishree@managalayatan.edu.in
[2]Professor, Department of Computer Engineering and Applications, Department of Computer Engineering and Applications, Mangalayatan University, Aligarh, India.
Email-id: udai.shankar@managalayatan.edu.in
[3]Assistant Professor, Department of Computer Engineering and Applications, Department of Computer Engineering and Applications, Mangalayatan University, Aligarh, India.
Email-id: lubna.ansari@managalayatan.edu.in
[4]Associate Professor, Department of Computer Science and Engineering, JSSATE Bengaluru.
Email-id: abhilashcb@jssateb.ac.in

Received: 02/10/2025
Revised: 31/10/2025
Accepted: 08/11/2025
Published: 13/11/2025

ABSTRACT
Hoysalas were a dynasty in Karnataka who ruled between 10th and 14th centuries. Rulers of this dynasty have built more than 150 temples which exhibit unique architectural features. This study uses a pretrained ResNet18 model to classify Indian temples as belonging to either Hoysala or non-Hoysala architecture. The study focuses on how dataset diversity and size affect model performance. The experiments show that model accuracy greatly increases with bigger, uncurated datasets than smaller, curated ones. This emphasizes how crucial big and varied datasets are for efficiently classifying heritage monuments. Despite having a rich architectural history, India lacks the data necessary for applying AI technologies. This research also proposes a model to expand heritage datasets using a gamified crowdsourcing model. The proposed crowdsourcing mechanism not only facilitates the collection of images but also improves metadata accuracy with the help of a large community. The proposed system ensures that the AI model improves with the enriched dataset.

Keywords: Digital archive; Iconography; India; Temple Sculpture; Iconography.

INTRODUCTION:

India's cultural heritage (PAL and KUMAR, 1986) is the most expansive compared to the rest of the world. One novel way to preserve this heritage is to document them digitally so that even when there is age-induced wear and tear, the monuments are accessible digitally. There are images and videos of every monument from at least a hundred years, which initially sat in people's albums and now on phones or scattered all through the net. The problem is how to unify them to create a big dataset that can be used for heritage preservation efforts. Once a large and diverse dataset is available, AI and ML can be integrated into systems to help classify and recognise monuments.

One way to build large datasets is through crowdsourcing coupled with machine learning (Mozafari et al., 2014). The need of the hour is to build a system that can help crowdsourced data and then classify and recognise them using AI models such as CNN and YOLO. This way, when heritage enthusiasts search for a monument, they get a large corpus of information. When photographers and collectors want to contribute their resources, the system can classify them with the help of AI and, of course, a human in the loop to perfect it. Once a corpus of this kind is built, the possibilities are endless. With the right metadata attachment (Salo et al., 2016) one can build AR VR systems, create geospatial mapping for locating monuments, and create engaging 3D models. The possibilities are endless. India's vast and diverse heritage presents both opportunities and challenges. One essential step in preserving this heritage is through digital documentation (Stanco et al., 2011). Since the past century, many monument images and videos have been captured but stored in personal archives or scattered across the internet. Hence, this resource remains largely unstructured and inaccessible.

One of India's noteworthy monuments is the temples built by the Hoysala rulers of Karnataka. The Hoysala temples (Kumaran and Barandhaman, 2022) are well-known for distinct features such as star-shaped podiums, Madanikas (Bignami, 2015) and the Hoysala symbol of a man slaying a lion. Figure 1 gives a glimpse into these architectural marvels. There are at least a dozen major temples belonging to this style in the districts of Hassan and Mysore and they are a major tourist attraction. Again, the problem is finding sizable datasets to apply AI/ML to analyse architectural features. This study is conducted to determine the characteristics and requirements of the dataset needed for classifying

temples into Hoysala and non-Hoysala categories using a pretrained ResNet18 model. Datasets of varying sizes and compositions are used in the experiments. Results of this study show that large, diverse datasets improve model performance when compared to smaller, curated ones. This finding highlights the necessity of building extensive datasets for heritage classification tasks.

The study suggests a novel gaming-based crowdsourcing strategy to address the data scarcity problem. This approach proposes a method to allow users to add photographs and metadata through an interactive platform. By using these metadata-enriched images to further train the model, dataset expansion happens continuously. This framework provides a scalable way to preserve and promote India's cultural heritage by fusing the strength of AI with the combined efforts of contributors.

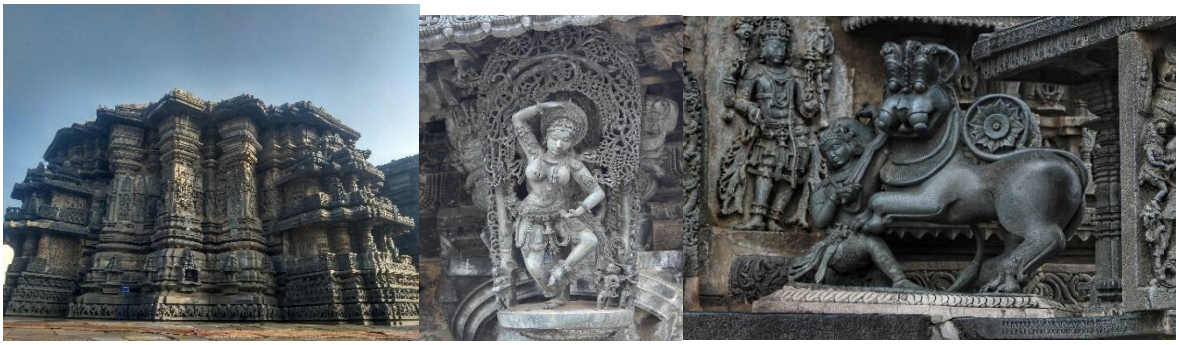

Figure 1: Collage of Hoysala temple with distinct features.

LITERATURE REVIEW

There is research going on worldwide to digitally preserve the built heritage. Now AI is being employed to enhance restoration procedures as this literature highlights. Below is a review of previous works on applying AI for heritage preservation.

Pretrained models such as ResNet and YOLO have been employed for architectural classification, providing good results with fine-tuning. However, their performance depends on the quality and size of the dataset (Abed et al., 2020).

Heritage resources such as images, videos, and texts are plentiful but not unified, creating a shortage of comprehensive datasets. This severely restricts the efficacy of machine learning models to analyse heritage data. Crowdsourcing has been used in some instances for building heritage datasets by encouraging the public to contribute images and, whenever possible metadata such as location, builders, and era (Vincent, 2017).

Since its founding in 1999, the AI & Cultural Heritage (AI & CH) working group has promoted cooperation between artificial intelligence and cultural heritage through seminars, educational programs, and cutting-edge documentation and preservation tools. Italian academics' contributions to cultural heritage projects over the years are highlighted in the article by Bordoni et al., 2013, which also highlights important AI techniques created in the field.

The ethical issues of incorporating AI into cultural heritage are examined by Tiribelli et al., 2024, who also offer recommendations for creating reliable AI and point out areas that require more study and regulation.

With a focus on Wuyuan County in Jiangxi Province, this study by Wang, 2022 investigates the use of AI technology in preserving and passing down the cultural landscape heritage of traditional villages. The research shows how AI techniques, such as image restoration and RF technology for real-time monitoring, can address issues including inadequate protection, supervision, and cultural loss. The research provides insights for preserving historical and cultural heritage.

Zhang et al., 2021 show how using social media images and the collaborative capabilities of AI and crowdsourcing can tackle the Cultural Heritage Damage Assessment (CHDA) challenge. The proposed CollabLearn system shows that to overcome AI's limitations in modelling damage is by combining AI with human expertise. This results in higher accuracy in estimating damage to cultural assets during disasters.

Lu, 2024 examines patterns in publications from the Web of Science database for the application of AI in preserving documentary material. The study summarises AI's advantages and disadvantages in document protection. It also examines future research possibilities by looking at keywords, author contributions, and citation trends using programs like VOSviewer.

Despite substantial research on using AI for cultural heritage, many unanswered questions remain (Münster et al., 2021). Current research covers various topics, from crowdsourcing models to damage assessment and dataset augmentation. This

research examines how current efforts are fragmented with dispersed datasets, inconsistent methodologies, and a lack of fusion between ethical and practical considerations.

METHODOLOGY

3.1 Model and Dataset

This study classifies Hoysala and non-Hoysala temples using two different models - ResNet18 model and Swin Transformer (Swin-Tiny) for classifying Hoysala and non-Hoysala temples. This was done in order to evaluate what works better with smaller datasets. While both models are pretrained on ImageNet, Swin-Tiny uses self-attention mechanisms with which it captures architectural patterns better. Here, the images are resized to 224×224 pixels and increased in number using cropping, flipping, and inducing colors. There are three datasets of different sizes and nature: first is a small dataset which has 50 images per class, second a dataset with 250 Hoysala, 50 non-Hoysala which creates an imbalance in the two classes which can be a normal occurrence in real-life data, and third is a large dataset with about 450 per class but it is just a collation of all available images without any filtering by humans.

Both models are fine-tuned using a linear classification using cross-entropy. The Swin Transformer captures more nuanced temple features but ResNet18 will only focus on borders and edges. Both the models were evaluated by using confusion matrices, accuracy, precision, recall, and F1-scores.

This study also proposes a gamified crowdsourcing mechanism for obtaining images. Since many have photos they have taken at the monuments, a fun way to make them share will grow the corpus and hence the models will train better to recognize iconographic feature. In the gamification users are asked to upload images and also key in details such as the place, temple name, the history they have learned about it and so on. These details are stored after vetting with an expert along with the images and will be used to retrain the models. With time, it will be helpful in determining in what way dataset size, diversity, and expert feedback impact classification especially when the datasets are small or imbalanced.

3.2 Proposed Gaming Model for Crowdsourcing

The concept shows how creating a gaming (Yen et al., 2015) platform where users can upload temple images and check whether they belong to the Hoysala architectural style. In this process, when users upload an image, model that is and the fine-tuned will classify it as Hoysala or non-Hoysala. Users will then provide feedback about the model’s prediction and the accuracy of the classification. The platform then checks with the user for consent and then stores images and any detail provided by them for future training. This results in a user-driven dataset expansion, increasing the model's accuracy over time.

This gaming platform has many benefits. It allows for dataset increase by collecting images and metadata from users. The crowdsourcing will increase the dataset and with bigger datasets the AI models will learn to identify nuanced architectural features. For example, if there are many images of a particular sculpture, the model will recognize that sculpture with better accuracy and precision. When people take part in building datasets with their own images, they will be more aware of heritage preservation efforts and this will create awareness and develops pride and responsibility among users for heritage. Additionally, with diverse user inputs, the platform will help test the classification model. Figure 2 below shows the gamifying features diagrammatically.

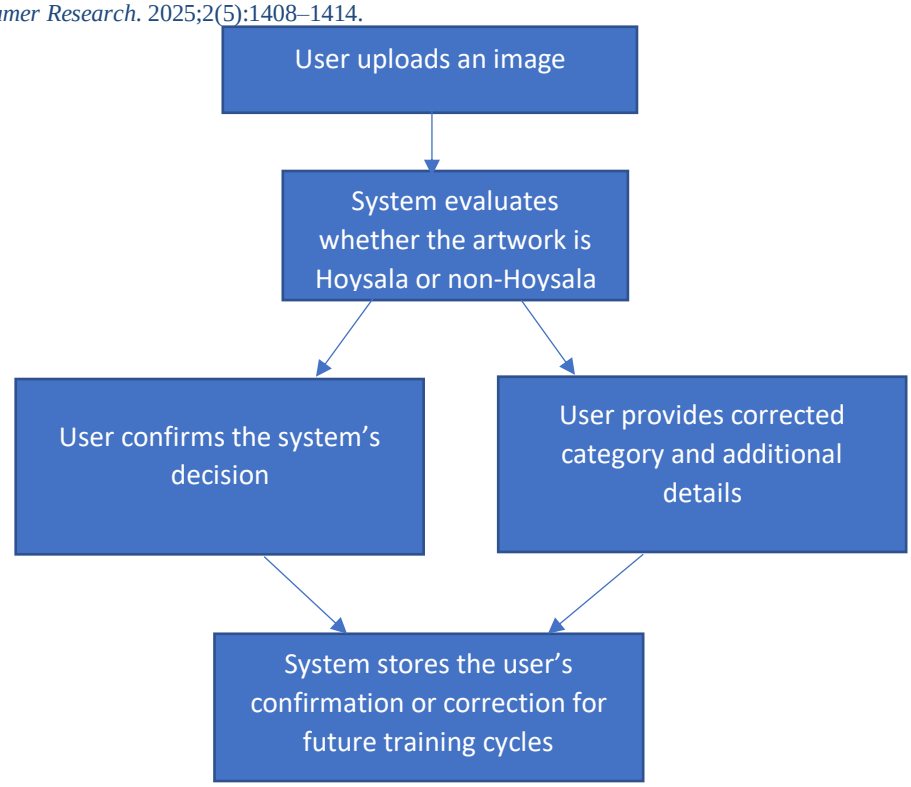

Figure 2: Methodology for image crowdsourcing

RESULTS AND DISCUSSION

Confusion matrices for ResNet18 and Swin Transformer (Swin-Tiny) for three datasets are shown in Table 1. Swin-Tiny shows 82% accuracy bettering ResNet18 which shows 76% on the small but balanced dataset. Swin-Tiny's transformer architecture helps it to capture better global architectural patterns such as shapes of podiums and roofs and temple layouts. But ResNet18 focuses more on local textures like patterns, ornaments, borders, leading to higher misclassifications.

Swin-Tiny again has high precision for the majority Hoysala class and improves recall for non-Hoysala images compared to ResNet18 for the imbalanced dataset. This suggests that attention-based models such as Swin-T can handle imbalance better since they learn discriminative features from examples.

Swin-Tiny has 92% accuracy compared to 88% by ResNet18 on the large uncurated dataset. The Transformer recognizes the large and noisy images better because it has a good global context and long-range spatial dependencies. The inclusion of crowdsourced metadata and human-in-the-loop validation further enhances model performance, particularly for visually ambiguous or partially occluded images.

In summary the results show that Swin Transformer is better for architectural classification of heritage monuments because of its global feature extraction, context sensitivity, and tolerance to imbalanced data. Alos, crowdsourced inputs can help with dataset growth and improves model accuracy and offers explainability for heritage researchers.

Table 1: Confusion Matrices (rows = true class, columns = predicted)

Dataset	Model	True Class	Pred: Hoysala	Pred: Others	Total
Small (100)	ResNet18	Hoysala (50)	38	12	50
		Others (50)	12	38	50
		Total	50	50	100
	Swin-T	Hoysala (50)	42	8	50
		Others (50)	10	40	50
		Total	52	48	100
Imbalanced (300)	ResNet18	Hoysala (250)	210	40	250
		Others (50)	12	38	50
		Total	222	78	300
	Swin-T	Hoysala (250)	220	30	250
		Others (50)	5	45	50
		Total	225	75	300
Large (900)	ResNet18	Hoysala (450)	400	50	450

		Others (450)	55	395	450
		Total	455	445	900
	Swin-T	Hoysala (450)	420	30	450
		Others (450)	40	410	450
		Total	460	440	900

Precision and recall exhibited by differently sized and composed image data

Dataset	Model	Accuracy	Precision (Hoysala)	Precision (Others)	Recall (Hoysala)	Recall (Others)	F1 (Hoysala)	F1 (Others)
Small (100)	ResNet18	0.76	0.76	0.76	0.76	0.76	0.76	0.76
	Swin-T	0.82	0.84	0.83	0.84	0.80	0.84	0.81
Imbalanced (300)	ResNet18	0.74	0.95	0.49	0.84	0.76	0.89	0.59
	Swin-T	0.82	0.98	0.60	0.88	0.90	0.93	0.72
Large (900)	ResNet18	0.88	0.88	0.89	0.89	0.87	0.88	0.88
	Swin-T	0.92	0.91	0.93	0.93	0.91	0.92	0.92

Results and Discussion

The results show the differences between ResNet18 and the Swin Transformer on the three datasets. For the small balanced dataset (50 images per class), ResNet18 achieves of a 76% accuracy with precision and recall nearly equal but without good generalization because of the small dataset size. Swin-Tiny performs better on this dataset by reaching 82% accuracy because its attention mechanism can recognise structural patterns that ResNet18 cannot.

With the imbalanced dataset consisting of 250 Hoysala, 50 non-Hoysala temple images, accuracy of ResNet18 continues to be low at 74%. It shows high precision for Hoysala class but shows poor performance on non-Hoysala class, indicating it requires a larger dataset. Swin-Tiny improves its accuracy to 82% and also shows better recall for small non-Hoysala images. It indicates that attention-based models can perform better even when there is a dataset imbalance since they learn global features.

The large uncurated dataset consisting of 450 images per class, both models perform better. ResNet18 achieves 88% accuracy, but Swin-Tiny reaches 92%. ResNet18 depends on local features, such as borders and carvings, which differ with lighting and angle and this leads to misclassifications. However, since Swin-Tiny also incorporates global architectural features such as star-shaped podiums, tower profiles, and building symmetry, it has stronger generalization even with diverse and noisy images.

4.2 Analysis of Results

- With larger datasets accuracy improved for both ResNet18 and Swin-Tiny. This confirms the importance of dataset size and also diversity in temple classification.
- ResNet18 performed poorly with imbalanced datasets, indicating that it requires balanced datasets, while Swin-Tiny continued to show better performance since it learns from broader architectural features.
- Attention-based Swin-Tiny performed better than ResNet18 in all datasets, especially in with noisy or diverse images where global structural features hold more relevance than fine textures.
- ResNet18 captures local features such as borders, carvings, and textures effectively. But when there were lighting variations and altered perspective these were insufficient.
- Swin-Tiny had stronger generalization across diverse samples since it learns global features like star-shaped podiums, shikhara (tower) profiles, and symmetry.

4.3 Implications of the Crowdsourcing Model

Today almost everyone has a mobile phone with a decent camera, and as a result, thousands of photos of monuments are sitting in people's galleries or already floating around online. If there is a platform where these can be brought together, it could quickly grow into a very large heritage dataset. The value is not only in the photos themselves, but also in the small pieces of information people can add—like where the photo was taken, during what festival or season, or even stories they have heard connected to the site. The dataset becomes richer and allows machine learning models to perform far better with details incorporated and helps in improving both precision and recall when identifying different temple features.

As the collection grows, the models start to notice details that are not always easy for humans to track consistently. For example, the star-shaped platforms common in Hoysala temples, the borders around doorways, or sculptures like the Madanikas can be recognised more reliably. Over time, this could help scholars compare styles across regions and even track how designs changed over centuries. Things that seem like minor variations in carvings or tower shapes may actually point to cultural exchanges or evolving traditions.

The crowdsourcing approach is also important because it brings the wider community into heritage preservation. When people add their own photos or even small stories,

they feel part of the process, not just outside observers. That kind of personal involvement builds a quiet sense of pride. Also, the fact that pictures come from different times—old family albums as well as new phone cameras—means changes in the temples can actually be tracked over the years. Old family photographs and new digital ones together allow comparisons that show how temples are aging, where damage is happening, or how restoration efforts have changed their appearance.

In this way, crowdsourcing is not just about enlarging datasets for AI models. It is also about creating shared responsibility, encouraging community involvement, and giving researchers and conservationists long-term material for protecting monuments.

CONCLUSION

This research indicates that dataset size and diversity have a significant effect on the quality of classification for heritage monuments. If the dataset is small or restricted, the models lose details or mix similar styles. For larger and more diverse collections, accuracy improves and the models are better at identifying distinctive features. The contrast between Swin-T and ResNet18 also highlights that various models pick up on various types of information — one emphasizing detailed specifics, the other overall patterns. Having both methods operating provides a more complete picture.

The second challenge is how exactly to construct such massive datasets. Conducting extensive fieldwork for years is not always feasible, but crowdsourcing provides a more practical approach. With so many individuals possessing high-quality camera phones, images and stories already exist, but unorganized. An easy, even gamelike method of sharing what one records, with some basic details such as where and when taken, could get people into the habit of sharing and giving back to the community. In this manner, the heritage preservation comes from within the community itself, and the dataset would grow significantly faster than could ever be handled by an individual researcher.

In the future, these datasets are able to drive more than classification. They can enable models to learn to identify repeated mythological narratives etched on temple walls, or monitor sculptures over time. And using devices like AR or VR, this data can be disseminated in common means. For instance, a person in a temple could take their phone and point it at a sculpture and receive a brief explanation, while a student could walk through a virtual version of the temple with carvings highlighted step by step. These are simple but potent means of keeping heritage relevant and accessible, not only for scholars but also for the general public.

REFERENCES

1. Abed, M. H., Al-Asfoor, M., & Hussain, Z. M. (2020). Architectural heritage image classification using deep learning with CNN.
2. Vincent, M. L. (2017). Crowdsourced data for cultural heritage. Heritage and archaeology in the digital age: Acquisition, Curation, and Dissemination of Spatial Cultural Heritage Data, 79-91.
3. Bordoni, L., Ardissono, L., Barceló, J. A., Chella, A., de Gemmis, M., Gena, C.,& Sorgente, A. (2013). The contribution of AI to enhance understanding of Cultural Heritage. Intelligenza Artificiale, 7(2), 101-112.
4. Tiribelli, S., Pansoni, S., Frontoni, E., & Giovanola, B. (2024). Ethics of Artificial Intelligence for Cultural Heritage: Opportunities and Challenges. IEEE Transactions on Technology and Society.
5. Wang, X. (2022). Artificial intelligence in the protection and inheritance of cultural landscape heritage in traditional villages. Scientific programming, 2022(1), 9117981.
6. Zhang, Y., Zong, R., Kou, Z., Shang, L., & Wang, D. (2021). Collablearn: An uncertainty-aware crowd-ai collaboration system for cultural heritage damage assessment. IEEE Transactions on Computational Social Systems, 9(5), 1515-1529.
7. Lu, Y. (2024). Application of AI in the Field of Documentary Heritage: A Review of the Literature. Journal of Artificial Intelligence Research, 1(2), 15-21.
8. Kumaran, R.N., & Barandhaman, V. 2022. Immortal Monuments and Sacred Temples. Journal of History, Archaeology and Architecture
9. PAL, R., & KUMAR, V. (1986). CULTURAL HERITAGE OF INDIA.
10. Mozafari, B., Sarkar, P., Franklin, M., Jordan, M., & Madden, S. (2014). Scaling up crowdsourcing to very large datasets: a case for active learning. Proceedings of the VLDB Endowment, 8(2), 125-136.
11. Salo, K., Giova, D., & Mikkonen, T. (2016). Backend infrastructure supporting audio augmented reality and storytelling. In Human Interface and the Management of Information: Applications and Services: 18th International Conference, HCI International 2016 Toronto, Canada, July 17-22, 2016. Proceedings, Part II 18 (pp. 325-335). Springer International Publishing.
12. Stanco, F., Battiato, S., & Gallo, G. (2011). Digital imaging for cultural heritage preservation. Analysis, Restoration, and Reconstruction of Ancient Artworks.
13. Yen, I. L., Zhou, G., Zhu, W., Bastani, F., & Hwang, S. Y. (2015, June). A smart physical world based on service technologies, big data, and game-based crowd sourcing. In 2015 IEEE International Conference on Web Services (pp. 765-772). IEEE.
14. Bignami, C. (2015). Re-use in the Art Field: the Iconography of Yakṣī. Journal of Indian Philosophy, 43, 625-648.
15. Münster, S., Utescher, R. and Ulutas Aydogan, S., 2021. Digital topics on cultural heritage

investigated: how can data-driven and data-guided methods support to identify current topics and trends in digital heritage?. Built Heritage, 5, pp.1-13.