Unveiling India's Heritage Through AI

1st Dr. R. Reena Roy*
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, Tamil Nadu, India
reenaroy.r@vit.ac.in

2nd Charvi Upreti
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, Tamil Nadu, India
charvi.upreti2021@vitstudent.ac.in

3rd Adithya S Nair
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, Tamil Nadu, India
adithya.snair2021@vitstudent.ac.in

4th Teja Pasonri
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, Tamil Nadu, India
teja.pasonri2021@vitstudent.ac.in

***Abstract*—This paper presents a web based monument recognition application that is based on MobileNetV2 and InceptionV3 models. Both MobileNetV2 and InceptionV3 base models were fine-tuned with transfer learning techniques on a dataset of labelled Indian monument images. The MobileNetV2 model was chosen for the final implementation. This model was integrated with Flask to create a web interface for simplified access. The application uses the Wikipedia API to collect the summaries of the detected monument and stores them locally. The libretranslate API is used to provide users with a summary in their preferred language. The main goal of this project is to aid users in understanding the historical importance of Indian monuments.**

***Index Terms*—Monument Recognition, Cultural Heritage, Convolution Neural Networks (CNN's), Transfer Learning, Image Classification, Cultural EducatioMonument Recognition, Cultural Heritage, Convolution Neural Networks (CNN's), Transfer Learning, Image Classification, Cultural Education**

I. Introduction

India is a vast country with great monuments, each of which has a deep rooted cultural significance. Despite this, these monuments are often overlooked by a large part of the population in the modern day. Due to the lack of easy access to information and a passive approach to acquiring knowledge, people are missing out on he rich cultural heritage of India.

Individuals who lack a good understanding of the heritage of their country find it difficult to connect with their roots. This lack of knowledge affects national pride and the unity of the nation. Moreover this affects the tourist experience.

This paper proposes a monument detection method to rectify this issue. The system was created by performing transfer learning on a MobileNetV2 base model with a dataset of labeled Indian monuments.

The proposed system seeks to provide users with an easy to use web user interface, where users can upload their images to get information about the monument. The web interface will provide the users with a short summary of the history, cultural significance and interesting details on the architecture of the monument. This approach allows the users to appreciate the rich heritage of the monument.

As India is a tourist hot spot, it is important to cater to foreign users. In order to address this, the web application allows users to convert the details about the monument into the language of their choice. By overcoming language barriers, this application ensures that people from around the world can gain vital knowledge about Indian history.

This study proposes an idea to use technology to overcome the knowledge gap on India's cultural heritage. The simplification of knowledge seeking has the potential to change the basis of understanding culture in India.

II. Literary Review

The study conducted by Hassan et al. (2023) [1] examined the need for integrating artificial intelligence with computer vision and natural language processing, with a focus on the relationship between monument recognition and cultural heritage conservation. While highlighting the use of automated systems in cultural artifact preservation, the research also takes into account the need for scalability and adaptability.

Similarly, the study by Mishra et al. (2024) [2] examined the applications of AI-based visual inspection systems to track the structural integrity of cultural heritage sites. The study emphasized how crucial artificial intelligence is for precisely recognizing building materials and pinpointing monument damage. While the research does touch upon the complexity involved in implementing such a setup, it also serves as a valuable reminder of the need for a practical solution.

Meanwhile, the focus of Rane et al. (2023) [3] study was on the development of environment friendly tourism made possible by AI technologies such as augmented reality (AR) and virtual reality (VR). The research using artificial intelligence based virtual tours demonstrated the effectiveness of immersive experiences in giving visitors a historical narrative. The study does require a thorough understanding of user intention to increase the efficiency of these technologies for cultural exploration.

Regarding the classification of urban buildings, the research by Taoufiq et al. (2020) [4] showcased HierarchyNet, a hierarchical CNN-based model intended for urban structure classification. The study shows that this model can effectively increase classification accuracy while lowering the computational complexity of conventional CNN architectures. With potential applications in cultural heritage preservation, HierarchyNet presents a promising method for being able to classify urban buildings by integrating knowledge from various levels of abstraction.

In order to identify the cultural heritage, Jindam et al. (2023) [5] used deep learning methods and satellite imagery from websites such as google earth to gather data. The method in this paper followed the extraction of features through the application of methods like Mean-Shift Detection (MSD) and Local Binary Patterns (LBP). Following their extraction, these features were fed into a Convolutional Neural Network (CNN) model. CNNs can capture spatial hierarchies in data,

which makes them a popular choice for image classification tasks. The dataset here was divided into 30% for testing and 70% for training. In addition, the streamlit framework was used in the development of the user interface, which streamlined the user interaction and deployment of the model for real world heritage identification.

The study of Alami Merrouni et al (2023) [6] proposed a new text summarization tool called EXABSUM that works in both extractive summary as well as abstractive summary. The process needed the use of quite many preprocessing techniques applied in the extractive summarization method such as tokenization, lowercase text transformation, part-of-speech tagging, lemmatization and the removal of non-meaningful words or stop words. However, the summarizer was compiled by assigning proper scores to sentences that stood out of redundant ones alteration. Actually, the summarization technique by EXABSUM is quite different from the traditional method, for it began with compression and merging of sentences and finished the process by means of ranking based on the ratio and number of keywords. This double feature, the ability to create extractive and abstractive summaries, has been put into action for different purposes such as text summarization for diverse content.

The study by Liu et al. (2024) [7] has shown that MT (machine translation) has a limitation by comparing the translation of the Chinese and English versions of the one given sentence for both human and machine systems. During the research, the trouble MT systems face, particularly in understanding a word-for-word exchange of information, was noted. But on the other hand, the research also highlighted the role of human supported AI interpretation as it renders a clearer understanding of language, which has an effect on machine translation.

In the study that Locher (2020) [8], conducted focused on the analysis of English subtitles of the Korean TV drama scenes, it was found that even though the subtitles of the original Korean context may be lost in translation, viewers can still get understanding and appreciation of the content from the material, which shows the audience ability to overcome cultural and language barrier and have their own meaning.

From summary to a text summarization article, Mohd et al.(2023) [9] paper provided study on semantic sum, which was a simple python based tool for summarizing text in english. With the help of libraries such as scikit-learn, PyTorch, and gensim, semantic sum has enabled individuals to do the analysis and summarization of the PDF documents by uploading it to the platform and present the summaries through a web-based interface. The summarizer utilizes Natural Language Processing (NLP) strategies by removing inconsistencies such as stopwords, punctuation, and URLs. Further, context understanding was improved using lemmatization and Word2Vec as means to text understanding. Following this, clustering approach and novel ranking method for summary creation was used.

In the paper of Razali et al. (2023) [10] a landmark recognition model was developed for smart tourism applications. The UMS landmark dataset which was publicly available was taken and a transfer learning approach was used for feature extraction using pre-trained CNN models. The features extracted from the data were then given as an input to several machine learning algorithms which contain Linear Support Vector Machine (LSVM), Gradient-Boosting Decision Tree (GBDT), and Multilayer Perceptron (MLP) among other algorithms. The evaluation indicated a 100% accuracy in UMS landmark dataset while the accuracy in Scene-15 dataset was 94.26%. This thus confirms the fact that smart tourism can be greatly influenced through the use of light deep learning models to perform landmark recognition efficiently.

Automated monument detection technology has received increasing attention, mostly in heterogeneous cultural circumstances such as India. In study of Aniket et al. (2021) [11] used Convolutional Neural Networks (CNNs) on TensorFlow environment. The study considered issues related to the model interpretability and data diversity, an imperative aspect of getting insight into the country's diverse past. Through focusing on such aspects the study will contribute to designing the new monument systems, targeting the peculiarities of the Indian cultural environment.

On the other hand, in the study which is focusing on monument conservation, El Antably et al (2020) [12] used semantic segmentation techniques to historical buildings in Cairo. Their research underscored the primary role of AI at pinpointing areas of degeneration on these objects. Nevertheless, the research showed the possible scalability obstacles and the necessity of practical application in real conservation settings, which could further be developed in future research on artificial intelligence in the field of monument protection.

The study conducted by Khandelwal et al. (2023) [13] therefore examines the accuracy of selected architectures which are ResNet50, InceptionResNetV2, and MobileNetV2 models. The sample was trained on a data set with 4000 pictures of 24 monument types that performed well when accuracy was considered. Nevertheless, the MobileNetV2 model scored the best result with the accuracy that reached 99.7% for training set and validation accuracy of 95.58%, thus proving that convolution neural networks can be useful for the tasks of monument classification.

In the study of Kukreja et al. (2023) [14], the paper used hybrid deep learning approach, which includes CNN and LSTM model, which was used for various classifications of heritage monuments. The study demonstrated promising results by exceeding 92.37% of accuracy in binary classification of heritage and the non-heritage as well as 95.89% accuracy in normal classifications, It was demonstrated by the mixed approach that large neural network architecture can be of great significance in enhancing the quality and accuracy of monument detection systems.

In a study on aerial photographic analysis for relic identification, Jaiswal et al. (2023) [15] used deep learning techniques to identify monuments from aerial images. They used checkboxes to tabulate a collection of almost 2,000 images for their research. The data set was improved by applying a variety of data enhancement techniques, including resizing, rescaling, rotating, adjusting brightness, and applying a mosaic. An analysis of the model's performance revealed a notable improvement in accuracy: while the VGG-16 architecture achieved 63% accuracy, the YOLOv5 PyTorch model demonstrated an astounding 90% accuracy.

In the paper of Yang (2023) [16], a smart English translation model was designed based on Improved GLR (Grammar lexicon Rule) algorithm. The model was able to improve the translation accuracy which was 75.1% to 99.1% after using the smart text processing. This research work shows how effective model was for dealing with challenges related to english language translation.

Similarly, the paper work of Tian et al. (2022) [17] resemble this model by using the transformer network to create a french to english machine translation model. So, when the french was translated to english, the model show an accuracy translation of 80%, which means that this model was capable of handling large language database.

Concerning the cultural heritage the study conducted by Das et al. (2020) [18] examined the potential of artificial intelligence (AI) to preserve the cultural heritage in a wider context with a focus on Indian archaeology. This paper showed that AI can collaborate with conventional techniques to help point out sites that were little known to archaeologists, thus demonstrating that AI has the potential to showcase sites that are less known. The paper also presents the research gaps present in the methods, indicating the need for further investigation to study the multi-disciplinary relationship between AI and conventional preservation techniques.

III. Background

A. Convolution Neural Network (CNN)

Convolution Neural Network (CNN) [19] is a deep learning model well-suited to analyze grid-like visual data. By employing layers such as convolution, pooling, and fully-connected, CNNs efficiently

extract key information from images, enhancing their understanding of complex visual scenes. This capability makes CNNs particularly effective in tasks requiring image analysis.

B. Transfer Learning

Transfer learning [20] is a technique that uses a model previously trained on a large dataset as a starting point for further learning tasks.

C. MobileNetV2

MobileNetV2 [21] is a unique lightweight yet powerful convolution neural network architecture designed for mobile and embedded devices, emphasizing performance without compromising accuracy. Its use of depth-wise separable convolutions, inverted residuals, bottleneck design, linear bottlenecks, and squeeze-and-excitation (SE) blocks minimizes computational demands, making it ideal for resource-constrained environments like smartphones and low-power devices. While offering good accuracy, it excels in tasks such as mobile image recognition and real-time classification, making it a valuable tool for various applications where efficiency is important.

D. InceptionV3

Inception v3 [22] is an advanced image recognition model employing a convolutional neural network for image analysis and object detection. It comprises both symmetric and asymmetric building blocks, featuring convolutions, average pooling, max pooling, concatenations, dropouts, and fully connected layers. Batch normalization is heavily utilized across the model, enhancing training stability by normalizing activation inputs. The Softmax function computes the loss, contributing to effective model optimization and accurate predictions.

IV. Dataset

The dataset was obtained from Kaggle. It consists of 24 classes of monuments, organized into separate 'test' and 'train' folders. Due to overlap between the two folders, we only utilized the 3118 images in the training folder.

V. Methodology

In this paper the dataset was divided into an 80/10/10 split for training, validation, and testing. Then, TensorFlow's ImageDataGenerator was utilized to efficiently load the images. During training and validation, various data augmentation techniques such as rotation, shifting, shearing, flipping, and zooming to expand the dataset were applied to enhance its variability and robustness.

Further, transfer learning techniques were employed to utilize the pre-trained base models MobileNetV2 and InceptionV3 for monument recognition . These base models were extended with a custom head tailored for monument classification. The model configuration included compilation with the Adam optimizer and utilization of the categorical cross-entropy loss function.

The model was trained for 10 epochs and Its performance was evaluated using the validation dataset. Then it tested on the test dataset. Performance Evaluation was conducted using the scikit-learn library. The performance has been discussed in detail in later sections. The attached Figure 2 depicts the flowchart of our method along with the user interface.

Following the model training, a Flask application was created to offer a user-friendly platform for predicting monuments from uploaded images. Users can upload images through the application, and the trained model is utilized to predict the depicted monument.

Additionally, concise summaries and translation functionalities were integrated using Wikipedia and the LibreTranslate API.

The attached images depict screenshots of the created web page, labeled as Figure 2, Figure 3, and Figure 4.

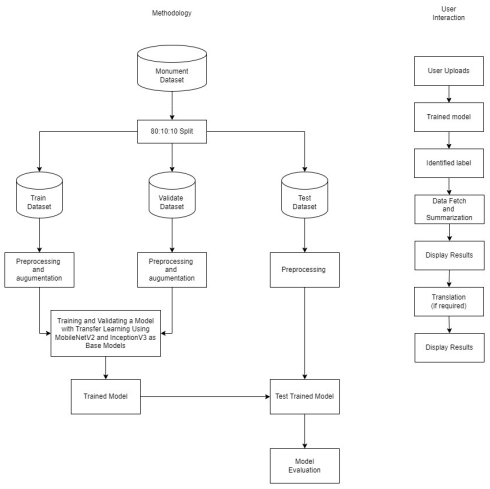

Fig. 1. Methodology flowchart and User Interface

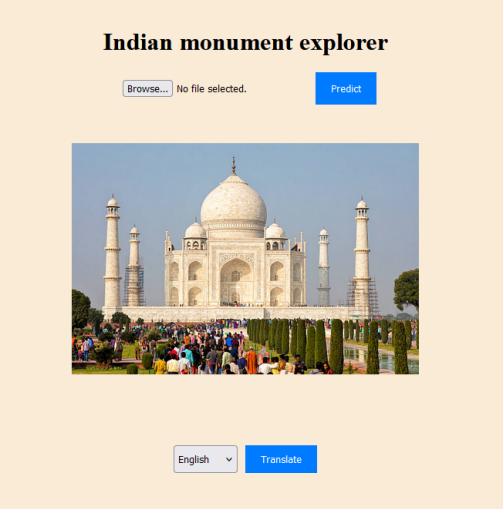

Fig. 2. Web Application: On Upload

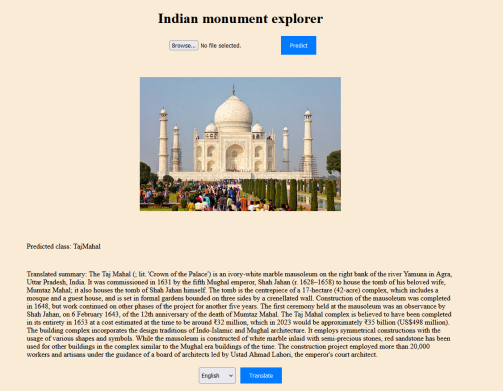

Fig. 3. Web Application: Result after prediction

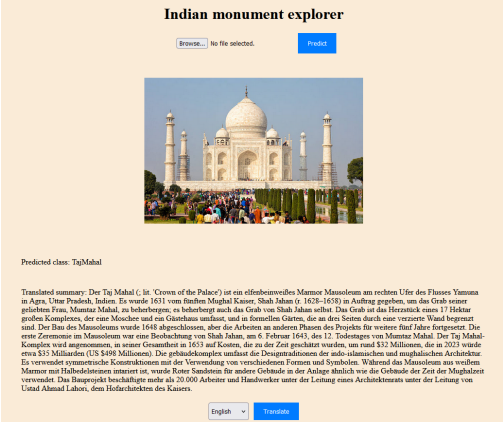

Fig. 4. Web Application: After Translation

VI. Performance Measures

The confusion matrix, along with the classification report, is a table that summarizes the results of a classification algorithm. Classification report helps evaluate performance using key metrics like recall, accuracy, precision, and F1-score. Accuracy measures the proportion of correct predictions out of all predictions made. Precision indicates the accuracy of positive identifications. Recall calculates the ratio of correctly predicted positive samples to all actual positive samples. F1-score is the harmonic mean of precision and recall. In the equations below:

TP (true positives) are correct positive predictions. TN (true negatives) are correct negative predictions. FP (false positives) are incorrect positive predictions. FN (false negatives) are incorrect negative predictions.

$$\textbf{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\textbf{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\textbf{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\textbf{F1-score} = 2 \cdot \frac{\textbf{Precision} \cdot \textbf{Recall}}{\textbf{Precision} + \textbf{Recall}} \tag{4}$$

VII. Result and Discussion

MobileNetV2 gave overall accuracy of 95.03%. With f1 Score of 94.98% ,recall of 95.03% and precision of 95.38%. The accuracy and loss graphs of this model have been attached as Figure 6 and Figure 5 respectively.

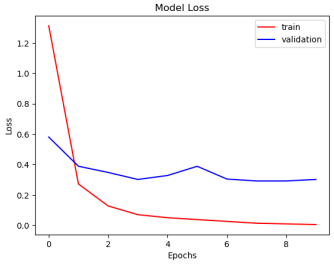

Fig. 5. MobileNetV2 Loss-Epochs graph

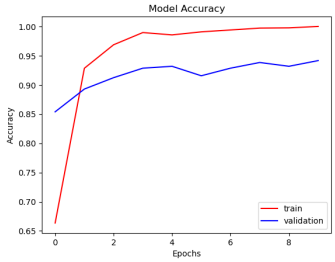

Fig. 6. MobileNetV2 Accuracy-Epochs graph

Inception v3 gave overall accuracy of 90.99%. With f1 Score of 90.86%, recall of 90.99% and precision of 92.17%.The accuracy and loss graphs of this model have been attached as Figure 8 and Figure 7 respectively.

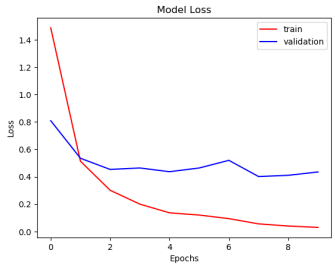

Fig. 7. InceptionV3 Loss

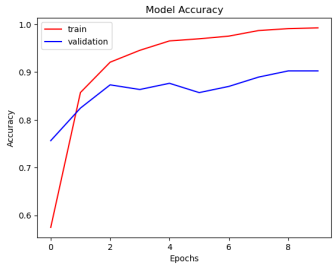

Fig. 8. InceptionV3

The sklearns summary() function outlines the details of the models. Total parameters represent the total number of parameters in a model, while trainable parameters are the weights that are updated during training to optimize performance. Non-trainable parameters, on the other hand, are pre-existing weights that remain unchanged during training.

MobileNetV2 distribution of the parameters it has been attached as Figure 9.

```
Total params: 2,808,714 (10.71 MB)

Trainable params: 183,576 (717.09 KB)

Non-trainable params: 2,257,984 (8.61 MB)

Optimizer params: 367,154 (1.40 MB)
```

Fig. 9. MobilNetv2 Model Summary

InspectionV3 distribution of the parameters it has been attached as Figure 10.

```
Total params: 22,648,426 (86.40 MB)

Trainable params: 281,880 (1.08 MB)

Non-trainable params: 21,802,784 (83.17 MB)

Optimizer params: 563,762 (2.15 MB)
```

Fig. 10. InceptionV3 Model Summary

Finally Figure 11 and Figure 12 illustrate the confusion matrix for both the models.

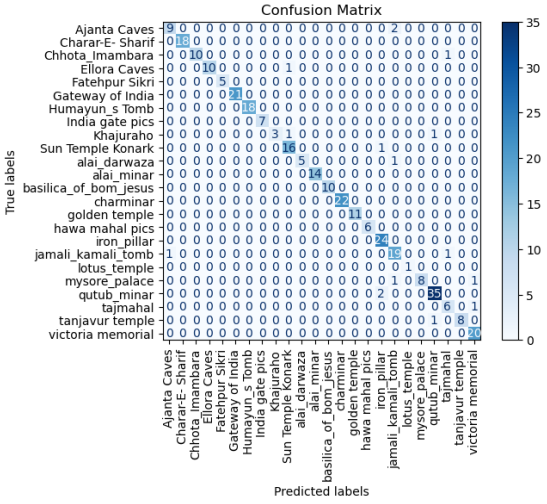

Fig. 11. MobileNetV2 Confusion Matrix

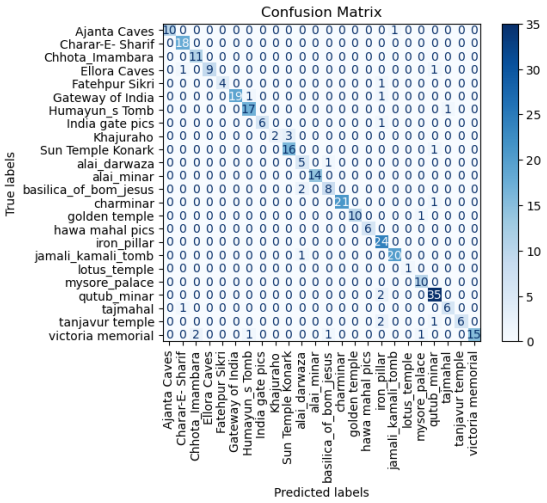

Fig. 12. InceptionV3 Confusion Matrix

VIII. Conclusion

This paper presents an approach to utilize transfer learning for the recognition of Indian monuments. High accuracies of 95.03% and 90.99%, respectively, were achieved by using MobileNetV2 and InceptionV3 as base models. The best model with 95.03% was utilized for predictions in the web application. The summarization and translation functionalities using Wikipedia and LibreTranslate APIs enable overcoming language barriers and sharing of information about the monuments.

References

[1] M. A. Hassan, A. Hamdy, and M. Nasr, "Survey Study: Monument Recognition using Artificial Intelligence," FCI-H Informatics Bulletin, vol. 5, no. 2, pp. 1-6, 2023.

[2] M. Mishra, T. Barman, and G. V. Ramana, "Artificial intelligence-based visual inspection system for structural health monitoring of cultural heritage," Journal of Civil Structural Health Monitoring, vol. 14, no. 1, pp. 103-120, 2024.

[3] N. Rane, S. Choudhary, and J. Rane, "Sustainable tourism development using leading-edge Artificial Intelligence (AI), Blockchain, Internet of Things (IoT), Augmented Reality (AR) and Virtual Reality (VR) technologies," Blockchain, Internet of Things (IoT), Augmented Reality (AR) and Virtual Reality (VR) technologies, pp. 1-20, Oct. 31, 2023.

[4] S. Taoufiq, B. Nagy, and C. Benedek, "Hierarchynet: Hierarchical CNN-based urban building classification," Remote Sensing, vol. 12, no. 22, pp. 3794, 2020.

[5] S. Jindam, J. K. Mannem, M. Nenavath, and V. Munigala, "Heritage Identification of Monuments using Deep Learning Techniques," Indian Journal of Image Processing and Recognition (IJIPR), vol. 3, no. 4, pp. 1-7, 2023.

[6] Z. A. Merrouni, B. Frikh, and B. Ouhbi, "EXABSUM: a new text summarization approach for generating extractive and abstractive summaries," Journal of Big Data, vol. 10, no. 1, pp. 163, 2023.

[7] Y. Liu and J. Liang, "Multidimensional comparison of Chinese-English interpreting outputs from human and machine: Implications for interpreting education in the machine-translation age," Linguistics and Education, vol. 80, pp. 101273, 2024.

[8] M. A. Locher, "Moments of relational work in English fan translations of Korean TV drama," Journal of Pragmatics, vol. 170, pp. 139-155, 2020.

[9] M. Mohd et al., "Semantic-Summarizer: Semantics-based text summarizer for English language text," Software Impacts, vol. 18, pp. 100582, 2023.

[10] M. N. Razali et al., "Landmark recognition model for smart tourism using lightweight deep learning and linear discriminant analysis," International Journal of Advanced Computer Science and Applications, vol. 14, no. 2, 2023.

[11] A. Ninawe et al., "Cathedral and indian mughal monument recognition using tensorflow," in Soft Computing Applications: Proceedings of the 8th International Workshop Soft Computing Applications (SOFA 2018), Vol. I 8, pp. 186-196, Springer International Publishing, 2021.

[12] I. Zohier, A. El Antably, and A. S. Madani, "An AI lens on historic Cairo: A deep learning application for minarets classification," 2020.

[13] S. Khandelwal et al., "A Study on Efficient Image Classification of Historical Monuments Using CNN," in 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 220-225, IEEE, 2023.

[14] V. Kukreja, R. Sharma, and S. Vats, "A Hybrid Deep Learning Approach for Multi-Classification of Heritage Monuments Using a Real-Phase Image Dataset," in 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE, 2023.

[15] G. K. Jaiswal et al., "Identification of monuments from aerial images using deep learning techniques," International Journal of Engineering Applied Sciences and Technology, vol. 8, no. 02, pp. 123-129, 2023. ISSN No. 2455-2143.

[16] S. Yang, "Intelligent English Translation Model Based on Improved GLR Algorithm," Procedia Computer Science, vol. 228, pp. 533-542, 2023.

[17] T. Tian et al., "A French-to-English machine translation model using transformer network," Procedia Computer Science, vol. 199, pp. 1438-1443, 2022.

[18] B. R. Das, H. B. Maringanti, and N. S. Dash, "Role of Artificial Intelligence in Preservation of Culture and Heritage," in Digitalization Of Culture through Technology, Routledge, 2022, pp. 92-97.

[19] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," Insights Imaging, vol. 9, no. 4, Art. no. 4, Aug. 2018. doi: 10.1007/s13244-018-0639-9.

[20] N. Donges, "What Is Transfer Learning? A Guide for Deep Learning — Built In," https://builtin.com/data-science/transfer-learning. (accessed Apr. 17, 2024).

[21] N. Sharma, "What is MobileNetV2? Features, Architecture, Application and More," Analytics Vidhya. https://www.analyticsvidhya.com/blog/2023/12/what-is-mobilenetv2/. (accessed Apr. 17, 2024).

[22] "Advanced Guide to Inception v3 — Cloud TPU," Google Cloud. https://cloud.google.com/tpu/docs/inception-v3-advanced. (accessed: Apr. 17, 2024).

[23] D. Kumar, "Indian monuments image dataset.", https://www.kaggle.com/datasets/danushkumarv/indian-monuments-image-dataset (accessed: Mar. 10, 2024.)