Artificial Intelligence in Travel and Tourism

Introduction

The most current development in artificial intelligence has revolutionised and had a big impact on numerous businesses in areas like facial recognition, medical diagnosis, autonomous driving, and more. This study investigates how deep learning technologies are used in the travel and tourism sector to classify images. According to (Statista, 2021), the travel and tourism sector has contributed more than $5.8 billion to the global GDP. the creation of a sophisticated methodology for landmark recognition in the travel and tourist industry. This topic is significant as it bridges cutting-edge AI techniques with the practical demands of enhancing the travel experience through automated landmark identification this explored in depth in the paper published about the impact of AI in travel and tourism (Bulchand-Gidumal, 2022).

However, landmark recognition is also a challenging problem, as landmarks can vary in appearance due to different viewpoints, lighting conditions, occlusions, or seasonal changes. Moreover, landmarks can be confused with other similar-looking objects or scenes or have multiple names or aliases. Therefore, I aimed to construct a landmark recognition model that is robust and accurate image classification that can handle these difficulties and recognize landmarks.

Objective

The research is aimed to answer the question of, how can an AI driven system effectively enhance the travel experience by recognizing different landmarks.

The project could have a multitude of applications in the realm of landmark identification. Although, the primary objective is to craft a sophisticated system that would performing the task of an automated tour guide. The system would be analyzing the images captured by a user and a match is obtained using the AI model, it would then be providing them with historical background information and cultural significance of the landmark. Moreover, the suggested system could also offer recommendations about nearby sites that might be of interest providing an overall immersive journey.

Background and Related Work

Landmark recognition has many applications in the field of travel and tourism, such as providing information and recommendations to tourists, creating virtual tours and guides, and enhancing the cultural and historical awareness of travellers. However, landmark recognition is also a challenging task, due to the large number of possible landmarks, the variations in their appearance, the occlusions and clutter in the images, and the lack of labelled data.

In the area of image recognition Convolutional Neural Network (CNN) is considered as cornerstone by many field experts. CNN is a powerful deep learning technique that has been widely applied to various image recognition tasks, such as object detection, face recognition, and scene classification as described by the work (Sharma, 2022). CNNs are especially suitable for these tasks because they can learn hierarchical features from raw pixel values, capturing complex patterns and structures in images. One of the image recognition tasks that can benefit from CNNs is landmark recognition,

which aims to identify and classify famous human-made or natural structures within an image, such as monuments, buildings, or mountains.

Numerous scholarly endeavors have been dedicated to enhancing the precision and resilience of landmark recognition systems, employing diverse methodologies and strategies. A prominent and influential endeavour in this domain was the Google Landmark Recognition Challenge (GLRC), a sequential set of competitions hosted by Kaggle during 2018 and 2019 (Weyand et al., 2020). The GLRC platform invited researchers and developers to design models capable of accurately identifying landmarks within an expansive dataset comprising over 5 million images spanning across 15 thousand categories (Weyand et al., 2020). This initiative effectively underscored the potential efficacy of deep learning models, specifically Convolutional Neural Networks (CNNs), in achieving remarkable performance levels within the context of landmark recognition (Dutreix et al., 2018). However, the challenges and limitations inherent to this task were also highlighted, including data imbalance, noise, the intricate diversity of landmarks, and the computational demands of intricate models (Dutreix et al., 2018).

Furthermore, the realm of mobile applications has witnessed the emergence of innovative solutions leveraging landmark recognition to enhance users' travel experiences. These applications employ diverse techniques to detect landmarks within images captured by users' devices, incorporating elements such as GPS data, geolocation information, pre-trained models, and cloud-based services. A notable instance is Google Lens, an application adept at landmark recognition that furnishes information and recommendations based on recognized landmarks. Google Lens synergistically harnesses GPS data and the Cloud Vision API, a pre-trained service proficient in landmark detection, subsequently providing data encompassing landmark names, locations, and associated confidence scores (Bilyk, 2020). Another noteworthy example involves the Landmark Recognition by Firebase ML Kit, an essential mobile Software Development Kit (SDK) capable of detecting landmarks using either the Cloud Vision API or on-device models.

Methodology

At the outset, a comprehensive image dataset was compiled. The dataset was acquired by conducting targeted Google searches for a diverse range of 10 distinct landmark types. These images were subsequently organized and categorized into dedicated folders corresponding to their specific labels. Following meticulous data cleaning processes, the total count of viable images settled at 1557. The downloaded images that were not fitting a certain condition, such as JPEG, JPG, BMP, and PNG, were removed from the dataset.

During the execution of project, I relied heavily in the TensorFlow library, this due to its capabilities in data management, preprocessing, model training, and overall optimisation. TensorFlow's capabilities were used during the preprocessing stage to divide the dataset into training and testing subsets (Dillon, 2017). This made it easier to establish a controlled setting for model evaluation. The data were divided into a validation set of 311 and a training set of 1246. TensorFlow's built-in data augmentation techniques emerged as invaluable tools for enhancing the model's generalization capabilities. By applying transformations like rotations, flips, and shifts to the training data, the model was exposed to a more diverse range of images, minimizing overfitting and enhancing its adaptability to real-world variations.

Initially, a baseline model was created using basic machine learning algorithm Support Machine Vector (SVM), this step was aimed to lay foundation for comparison and evaluation of subsequent deep learning approaches. The choice of SVM was driven by its established efficacy in classification tasks and its capability to separate different categories in situations where there are many different characteristics being considered. The data was split into training and validation, as well as, standardized using the standard scalar.

There were several experimentations with different image recognition model, one of the chosen approaches for the project involved constructing a custom CNN, this model is structured in layers that process the input image in a way that allows it to recognize and understand features. The convolutional layers extract various patterns, edges, and textures from the image. These are combined in subsequent layers to build a representation of what's in the image. The pooling layers help to condense the information and make the processing more manageable.

After these steps, the model flattens the information and passes it through fully connected layers, like traditional neural networks. These layers learn to combine the extracted features to make accurate predictions about what the image represents. In this case, the model aims to classify images into one of ten different categories using a SoftMax function.

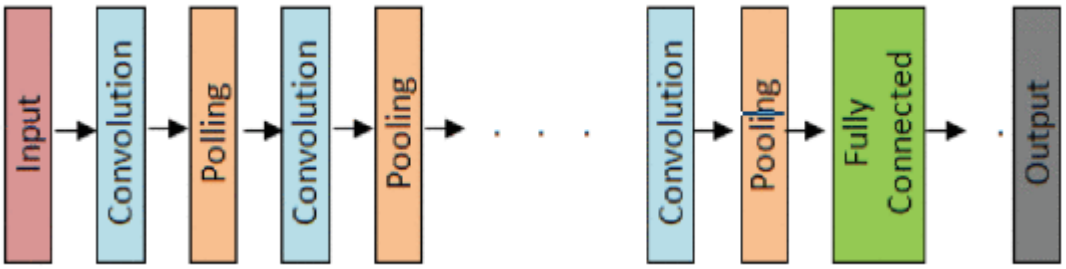

Figure 1 Design of a basic Convolutional Neural Network, adopted from (Sultana, 2018)

Overall, this model's structure helps it learn and identify important characteristics in images, making it useful for tasks like recognizing landmarks.

In quest to reach the optimal results for the project, I employed three pre-trained models in and compared them against a custom-designed Convolutional Neural Network (CNN) model. The selected pre-trained architectures encompass VGG16, ResNet 50, MobileNet, and EfficientNet.

The Visual Geometry Group 16 (VGG16) was utilized, this is a well-established pre-trained CNN architecture known for its straightforward yet effective design. Its architecture consists of multiple convolutional and pooling layers, resulting in a deep network capable of capturing intricate features (Tammina, 2019). Utilizing VGG16 as one of the benchmark models enables a comparative assessment of its performance in landmark recognition against the custom-designed CNN model.

In the course of the project experimentation with different transfer learning models Residual Network with 50 layers (ResNet 50) was also deployed. The ResNet 50 represents a groundbreaking advancement in pre-trained models, its architecture incorporates residual connections that tackle the vanishing gradient problem (Mukti, 2019), allowing for the successful training of extremely deep

networks. By employing ResNet 50 as a benchmark, this research seeks to evaluate the advantages offered by its residual connections in landmark recognition tasks.

Another transfer learning that was tested which is the MobileNet, the model is known for its efficiency and lightweight design. The model's architecture is tailored for scenarios where computational resources are constrained, making it suitable for the projects real-time applications on mobile devices (Sandler, 2018). A comparison was carried to explore the trade-offs between model efficiency and recognition accuracy in the context of landmark identification.

EfficientNet had been proved to be very robust model for image classification therefore it had to be explored as an option for the project. The EfficientNet model introduces a paradigm shift in balancing model complexity and performance. Its design is characterized by compound scaling, encompassing depth, width, and resolution dimensions, thereby optimizing the architecture across multiple scales (Koonce, 2021). Employing EfficientNet as a benchmark model for this study enables a meticulous examination of its ability to attain high recognition accuracy while efficiently managing computational resources. This architecture's innovative approach holds the promise of achieving remarkable performance levels, with a reduced computational footprint.

All the transfer learning models underwent a consistent training process. Their pre-trained weights were directed through a flattening layer, followed by two designated dense layers, each comprising 512 units. The initial dense layer employed a Rectified Linear Unit (ReLU) activation function, while the subsequent layer employed SoftMax activation to choose from the available 10 classes. Additionally, all models were optimized using the Adam optimizer, and the sparse categorical crossentropy was employed as the loss function, with accuracy serving as the chosen metric for evaluation.

The effectiveness of the developed models was assessed through a range of evaluation metrics, with a focus on providing a comprehensive understanding of their performance. Primarily, two widely used evaluation tools were employed: the Confusion Matrix and the Classification Report.

The Confusion Matrix is a fundamental tool for visualizing the performance of classification models. It provides a tabular representation of predictions versus actual labels, allowing for the determination of True Positives, True Negatives, False Positives, and False Negatives. By evaluating the distribution of these outcomes, the Confusion Matrix offers insights into the model's strengths and weaknesses in identifying different classes. In the context of this project, the Confusion Matrix aids in understanding the model's tendency to misclassify landmarks and facilitates the identification of specific challenges.

The Classification Report is a concise summary of key classification metrics, including Precision, Recall, F1-score, and Support. Precision measures the model's accuracy in predicting positive instances, while Recall gauges its ability to correctly identify positive instances. The F1-score harmonizes these two metrics, providing a balanced assessment of the model's overall performance. The Support metric quantifies the number of instances in each class, enabling the identification of class imbalances. Utilizing the Classification Report offers a detailed overview of the model's precision-recall trade-off and helps in making informed decisions regarding model fine-tuning.

Discussion and Results

The initial model constructed, serving as the baseline of the experiment was the SVM, it demonstrated a poor performance in classifying the diverse images, yielding an aggregate accuracy

of 71%. Notably, its most proficient classification pertained to the Sudan pyramids, boasting 35 True Positive (TP) predictions. On the contrary, its weakest predictive capability manifested in identifying the Egypt pyramids, registering merely 6 instances of accurate positive predictions. Intriguingly, the model demonstrated instances of misclassification, notably between the Colosseum and the London Bridge. Specifically, it misclassified 11 instances of the Colosseum as the London Bridge and 7 images of the London Bridge as the Colosseum.

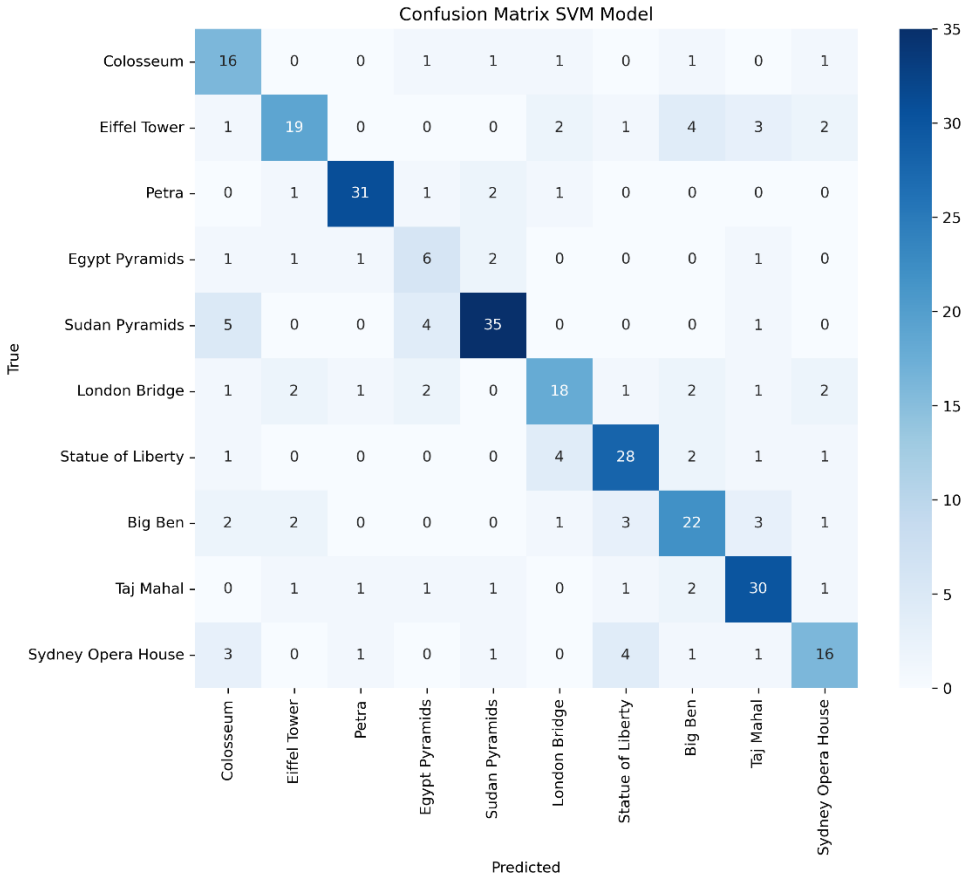

Figure 2 Confusion matrix representation for the SVM model performance, this displayed higher true positive rate for Sudan pyramid.

To ensure an optimal experimental environment and mitigate the risk of overfitting, robust data augmentation techniques were implemented across various models. Notably, data augmentation is critical for bolstering the models' capacity to generalize effectively.

The initial customized CNN model, devoid of supplementary regularizers, showcased a commendable overall accuracy of 95% and a validation accuracy of 80.70%, the classification report of this model is displayed in figure (3). Subsequently, the introduction of an L2 regularizer aimed at potential performance enhancement. However, this yielded no discernible improvement, with the model attaining an overall accuracy of 87.16% and a validation accuracy of 77.48%.

	precision	recall	f1-score	support
Big Ben	0.65	0.94	0.77	16
Colosseum	0.56	0.91	0.69	11
Egypt Pyramids	0.74	0.88	0.80	16
Eiffel Tower	0.92	0.80	0.86	15
London Bridge	0.85	0.58	0.69	19
Petra	1.00	0.90	0.95	20
Statue of Liberty	0.70	1.00	0.82	7
Sudan Pyramids	1.00	0.75	0.86	24
Sydney Opera House	1.00	0.74	0.85	19
Taj Mahal	0.71	0.77	0.74	13
accuracy			0.81	160
macro avg	0.81	0.83	0.80	160
weighted avg	0.85	0.81	0.81	160

Figure 3 Classification report of customised CNN, it had a validation accuracy of 81%

Furthermore, the integration of early stopping regularization, along with dropout layers, culminated in a diminished performance. Despite an intended training duration of 100 epochs, early stopping was activated at the 50th epoch due to a patience setting of 10. The resultant accuracy achieved was 81.54%, with a validation accuracy of 79%. This incremental improvement was only marginal when juxtaposed with the L2 regularizer-enhanced model. Moreover, a model configuration employing Batch Normalization exhibited even lower accuracies, at 82.10% overall and 75.50% in validation. Consequently, comprehensive hyperparameter fine-tuning revealed that a foundational CNN model excelled in terms of accuracy, proficiently discerning the various landmark categories. It's worth noting that the experimentation extended to different optimizers as well. In addition to the Adam optimizer, variations involving RMSprop and SGD optimizers were investigated and there was no improvement to the models performance.

Conversely, the application of transfer learning models for landmark identification presented notably superior outcomes compared to the CNN model. The ResNet model exhibited remarkable accuracy, achieving 99.52% overall accuracy and an impressive validation accuracy of 93.78%. This performance pinnacle was reached at the 96th epoch, with a training duration exceeding one hour for 100 epochs. Upon testing the model, the prediction as displayed in figure (4), it appeared that it mainly confused the prediction of the Eiffel Tower.

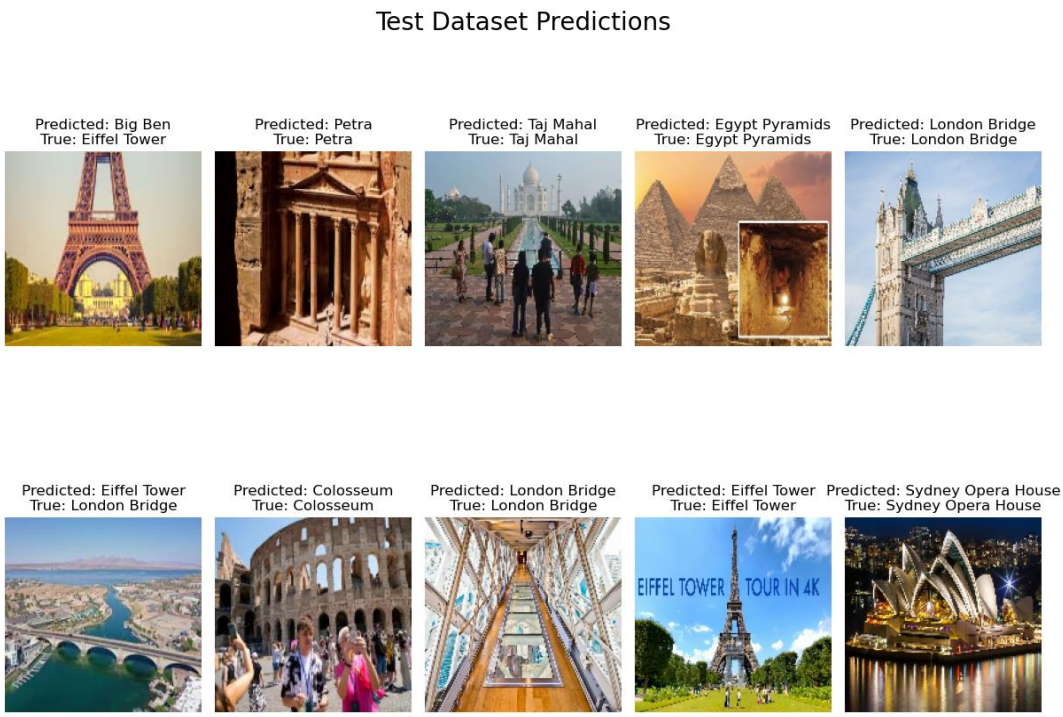

Figure 4 Testing the ResNet model, this showed confusion mainly with Eiffel Tower landmark

Likewise, the EfficientNet model yielded comparable results to ResNet, with an accuracy of 99.36% and the highest validation accuracy of 93.38%. Notably, the training of EfficientNet was notably expedited, completing 100 epochs in approximately 37 minutes. The model had a total of 147 True Positive predictions and this is displayed in the confusion matrix below.

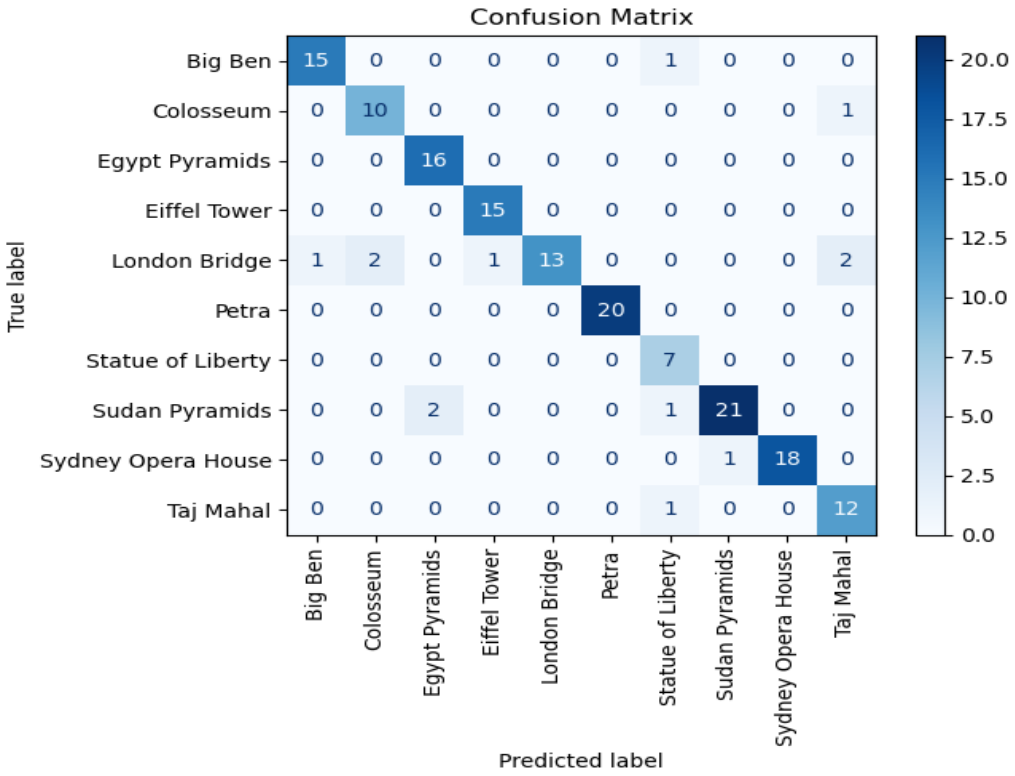

Figure 5 confusion matrix of the EfficientNet model

While the VGG16 pretrained model demonstrated moderate performance relative to other pretrained counterparts, the model produced a 98.15% accuracy and validation accuracy of 92.05%. However, this model consumed the most time to train with over 3 hours and 12 minutes. The model accuracy kept gradually improving, although by the end of the training cycle the validation accuracy seemed to drop slight as shown in figure (6)

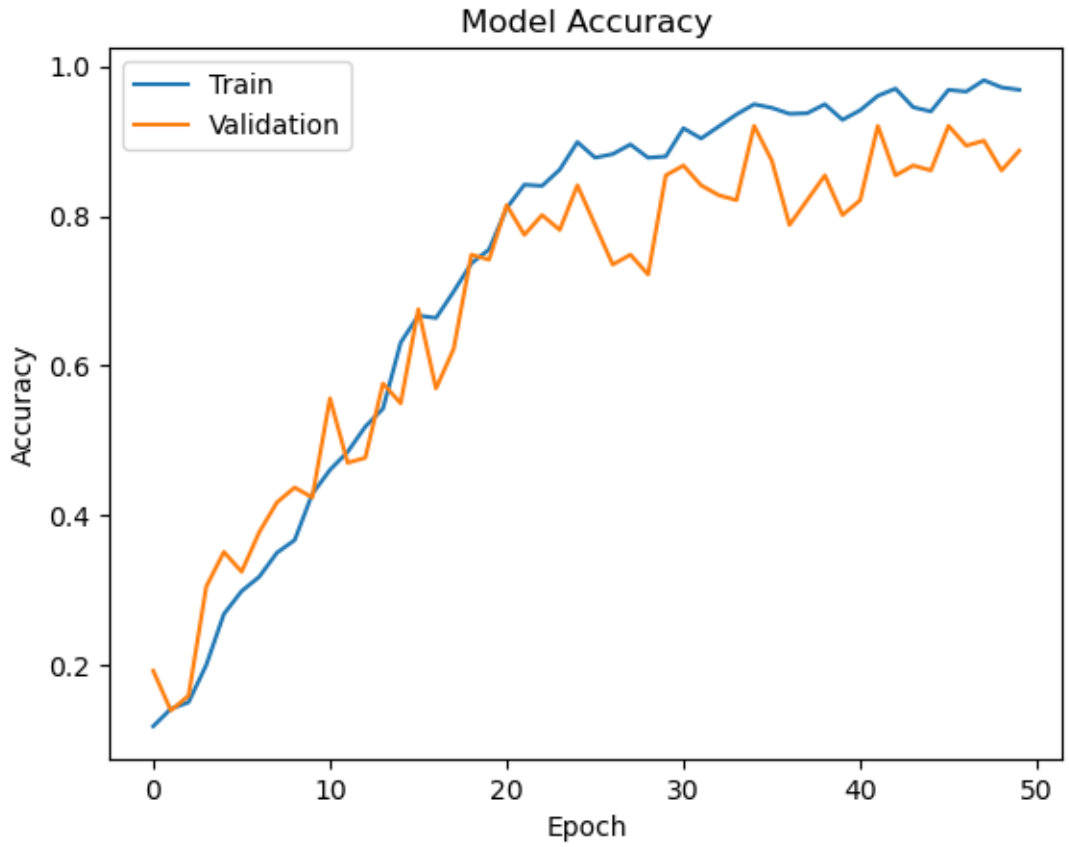

Figure 6 The model accuracy of the VGG16 model, it showed a slight drop in the model validation by the end of the training

The MobileNet model displayed the least favourable outcomes. This model achieved an accuracy of 94.70% and a validation accuracy of 77.41%. The comprehensive training of the MobileNet model spanned approximately 81 minutes. The model had 123 True positive predictions as displayed in figure 7 confusion matrix, with the most being 20 correct of the Sudan pyramids.

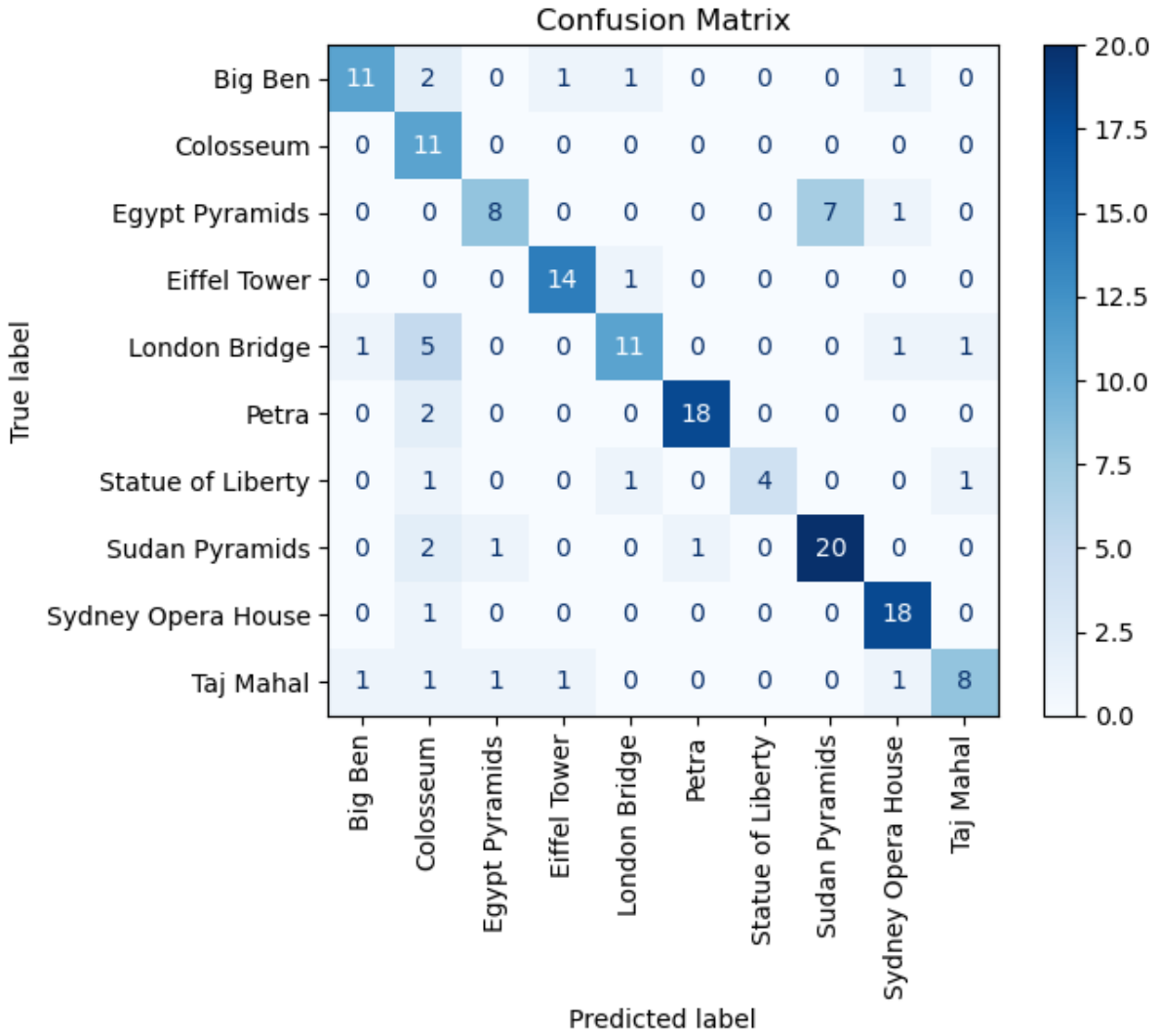

Figure 7 MobileNet Confusion matrix, it had the most correct prediction of the Sudan pyramid landmark.

In summation, meticulous experimentation with diverse hyperparameter configurations elucidated that the foundational CNN model emerged as the most adept at classifying different landmark categories. Transfer learning models, most notably ResNet and EfficientNet, excelled in yielding superior accuracy rates. The collective findings underscore the pivotal role of model architecture in driving the precision and efficacy of landmark recognition, delineating a path for further optimization and advancement in this domain.

Conclusion and Future work

In conclusion, this study investigated the use of deep learning, specifically Convolutional Neural Networks (CNNs), for landmark recognition in the travel and tourism sector. The aim was to create an AI-driven system that functions as an automated tour guide, identifying landmarks and providing historical insights. Results revealed diverse model performance, with transfer learning models like ResNet 50 and EfficientNet achieving accuracy rates of 99.52% and 99.36%, respectively.

For future work the system could be trained on a larger dataset of images. This would likely improve the accuracy of the system and could be tested on a wider range of landmark types. This would help

to evaluate the generalizability of the system. The system could also be integrated with a mobile application.

Overall, this study has demonstrated the potential of deep learning techniques for landmark recognition. The developed system achieved high accuracy levels, and the results suggest that the system could be further improved by training on a larger dataset and testing on a wider range of landmark types.

References

Statista (2021) Global Tourism. Available at: https://www.statista.com/topics/962/global-tourism/#topicOverview (Accessed: 14 August 2023).

Bulchand-Gidumal, J., 2022. Impact of artificial intelligence in travel, tourism, and hospitality. In *Handbook of e-Tourism* (pp. 1943-1962). Cham: Springer International Publishing.

Sharma, S. and Guleria, K., 2022, April. Deep learning models for image classification: comparison and applications. In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 1733-1738). IEEE.

Weyand, T., Araujo, A., Cao, B. and Sim, J., 2020. Google landmarks dataset v2-a large-scale benchmark for instance-level recognition and retrieval. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2575-2584).

Dutreix, M., Hatch, N., Kuppan, R., Pattanashetty, P. S. K., & Sundaresan, A. (2018, April 25). Google Landmark Recognition and Retrieval Challenges. Accesed from [https://nhatch.github.io/files/landmarks_report.pdf]

Bilyk, Z.I., Shapovalov, Y.B., Shapovalov, V.B., Megalinska, A.P., Zhadan, S.O., Andruszkiewicz, F., Dołhańczuk-Śródka, A. and Antonenko, P.D., 2020. Comparing Google Lens recognition accuracy with other plant recognition apps. In *Proceedings of the Symposium on Advances in Educational Technology, AET*.

Dillon, J.V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M. and Saurous, R.A., 2017. Tensorflow distributions. *arXiv preprint arXiv:1711.10604*.

Sultana, F., Sufian, A. and Dutta, P., 2018, November. Advancements in image classification using convolutional neural network. In *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)* (pp. 122-129). IEEE.

Tammina, S., 2019. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, *9*(10), pp.143-150.

Mukti, I.Z. and Biswas, D., 2019, December. Transfer learning based plant diseases detection using ResNet50. In *2019 4th International conference on electrical information and communication technology (EICT)* (pp. 1-6). IEEE.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4510-4520).

Koonce, B. and Koonce, B., 2021. EfficientNet. *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*, pp.109-123.