sustainability

MDPI

Article

Toward Sustainable Heritage Documentation: A Deep Transfer Learning Approach to Classifying Saudi Architectural Façade Styles

Mayda Alrige [1,*], Abdullah Alattas [2], Ahmad Fallatah [2,3] and Riyan Sahahiri [2]

[1] Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[2] Department of Geomatics, Faculty of Architecture and Planning, King Abdulaziz University, Jeddah 21589, Saudi Arabia; afalattas@kau.edu.sa (A.A.); amfallatah@kau.edu.sa (A.F.); rmsahaheere@kau.edu.sa (R.S.)
[3] The Center of Excellence in Smart Environment Research, King Abdulaziz University, Jeddah 21589, Saudi Arabia
* Correspondence: malraegi@kau.edu.sa

Abstract

The Kingdom of Saudi Arabia launches the Saudi Architecture Map, which includes 19 architectural styles for the country's main regions. These architectural styles are shaped by distinct climatic, material, and cultural determinants across the country. A deep dive into the architectural elements that define these styles reveals many similarities and nuances, making manual classification of modern buildings challenging. Despite the kingdom's rich heritage, and since the launch of the Saudi Architecture Map and its contained styles and corresponding design guidelines, no operational digital pipeline currently exists for automated, quantitative, and reproducible classification of façade imagery against these styles. This limits heritage documentation and compliance monitoring at scale. This study attempts to fill this gap. We developed a deep-learning classification framework for building façade imagery considering the Saudi architectural styles. We curated a dataset of 3750 façade images, comprising 2320 full-building façade images and 1430 architectural-element images. We collected the data from online heritage repositories. We benchmarked five convolutional backbones (ResNet50, EfficientNetB0, EfficientNetV2M, ConvNeXtBase, and InceptionV3) under a unified two-stage transfer learning strategy. The first stage trained classifier heads on architectural-element images with frozen backbone weights, and the second stage fine-tuned the top 30% of backbone layers on full-building façade images. Among the five CNN backbones, ConvNeXtBase achieved the strongest performance across all four metrics, achieving 90.75% in overall accuracy, 88.06% in F1-Score, 92.05% in producer accuracy, and 86.65% in user accuracy. Ten of the eighteen styles exceeded 90% F1, and two styles—Bisha Desert and Tabuk Coast, both with very small corrected test counts (n = 6 and n = 1, respectively)—fell below 70%. These results indicate that ConvNeXtBase with two-stage transfer learning is a promising step toward automated documentation of Saudi architectural styles. Yet, further validation through ablation testing or calibration is needed before deployment for compliance monitoring.

Keywords: convolutional neural networks (CNNs); transfer learning; ConvNeXtBase; GIS; Saudi architectural map; architectural style identification; heritage documentation

Check for updates

Academic Editors: Najam Ul Hasan and Yasir Javed

Received: 9 August 2026
Revised: 7 September 2026
Accepted: 8 September 2026
Published: 10 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Cultural heritage helps unite communities and sustain their values. The United Nations's SDG 11.4 calls for protecting cultural and natural heritage. Architectural heritage is a tangible expression of regional identity that attracts local and external investment. It is found to be a catalyst for economic development and social inclusion [1].

On 16 March 2025, the Crown Prince of Saudi Arabia announced the Saudi Architecture Map [2]. This map presents 19 distinct regional architectural styles inspired by the Kingdom's geographic, climatic, and cultural diversity. These styles, along with their region-specific design guidelines, aim to direct Vision 2030's urban growth while maintaining heritage identity. Vision 2030 giga-projects seek a delicate balance between modernization and heritage conservation [3]. Reconciling modernization with conservation has already been an active concern in the literature, primarily in the kingdom of Saudi Arabia. Alnaim and Alnaim urge urban planners to integrate traditional forms, such as Najdi and Hijazi architecture, into new development projects [4], aiming to conserve national identity while keeping pace with modernization. A case study from Diriyah and Alula confirms this finding [5]. More concretely, Aldossary and their colleagues built a regeneration framework that pairs conservation policy with sustainable tourism and community participation [3]. In their study, they identified the need for systematic classification of architectural styles. To date, no automated method exists to classify building façades against Saudi's 19 official styles, leaving heritage documentation and compliance monitoring dependent on slow, subjective manual review. Manual classification, heritage documentation, and compliance monitoring are especially challenging because geographically adjacent styles are similar.

Convolutional neural networks with transfer learning give strong results for façade classification [6]. Despite growing interest in applying convolutional neural networks to architectural heritage classification, applying them to classify Saudi architectural styles remains limited. This study closes that gap by designing, training, and evaluating a deep-learning classifier for Saudi architectural styles and embedding it in a Gradio decision-support tool for heritage documentation and compliance monitoring. This research paper makes three contributions:

(1) A curated façade dataset from official Saudi Architecture guidelines and the Saudi Architecture Map.
(2) A two-stage transfer-learning pipeline trained on architectural elements, then fine-tuned on full façades using five CNN backbones under identical hyperparameters.
(3) A Gradio decision-support tool that outputs style predictions with confidence scores.

2. Background and Related Works

2.1. Saudi Architectural Styles

The recently published Saudi Architecture Map comprises nineteen styles, as outlined in Figure 1. Each style is shaped by a distinct combination of environmental and cultural factors such as façade elements, colors, palettes, and textures. Coastal Hejazi façades, for example, feature tall openings and connected wooden Rawasheen, whereas Najdi traditions emphasize solid mud walls with small triangular openings for thermal control and privacy. These regional contrasts are reflected in the official design guidelines as part of the Saudi Architecture Map. The visual overlaps between some geographically adjacent styles, such as connected Rawasheen in Jeddah versus separate Rawasheen in Taif, making manual classification subjective and challenging.

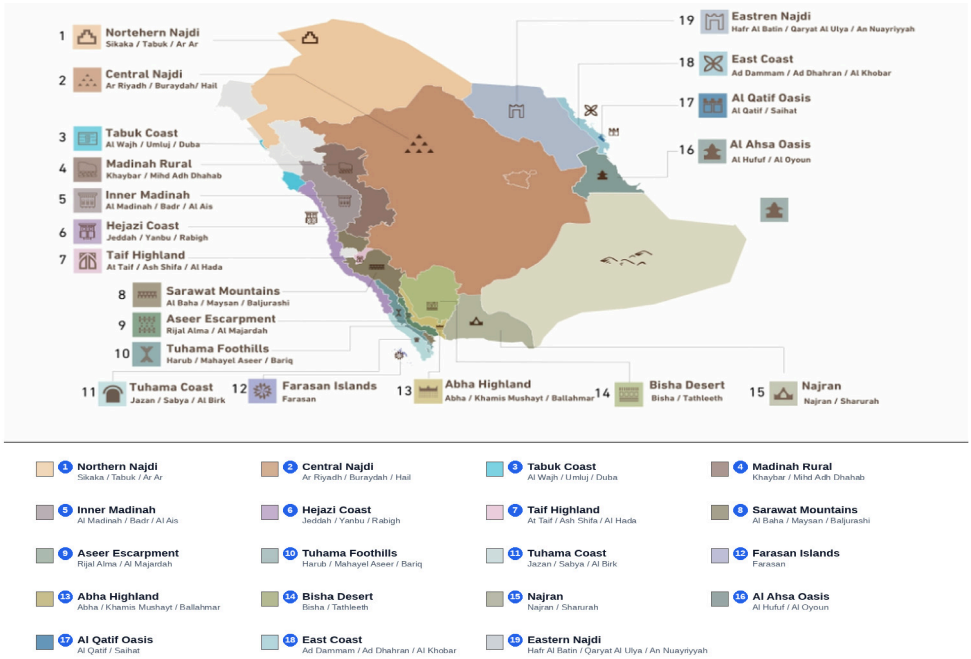


Figure 1. The geographic distribution of the 19 recognized Saudi architectural styles across the Kingdom of Saudi Arabia.

2.2. Deep Learning for Architectural Heritage Emerged as the Leading Method for Analyzing Heritage Images

Over the last decade, deep convolutional neural networks (CNNs) have emerged as the leading method for analyzing heritage images [7], and transfer learning from vast generic image datasets supports this approach [6]. Llamas et al. demonstrated that pretrained CNNs outperform handcrafted pipelines [8]. Abed et al. also reported clear gains from transfer learning [9]. Similarly, Cosovic and Janković used CNNs to classify cultural-heritage images of religious and civic buildings [10], while Shehata and Alaboud specifically focused on identifying features of architectural heritage with deep learning and underlined the importance of element-level supervision for façade-level recognition [7].

More recent contributions have moved beyond style classification toward city-scale applications. For example, Siountri & Anagnostopoulos applied a deep learning algorithm to classify historic buildings in Athens [11] to help experts identify building types based on shared morphological traits across different historical periods. Similarly, Zou and his colleagues used the same deep learning method to measure regional architectural style in Hubei Province, China [12].

Another line of research shifts the focus from building-level architectural features to village-level heritage elements. Tan, Zhu & Chen proposed an automated deep learning method to identify broader heritage value elements (TVHVEs) that would normally require costly expert field surveys [13]. Also, Ma and their colleagues extended this line of work to develop an image-based recommendation system for local architectural heritage in Jiangxi, Chinan [14]. They used a deep hashing retrieval method on a newly constructed dataset of traditional Chinese architecture. The authors mapped the retrieval results onto 2D geographic space to reveal relationships among heritage categories.

Ottoni and Ottoni undertook multiple projects between 2024 and 2025, beginning with facade-based heritage dataset creation and later optimizing models in Brazil [6,15]. In 2024, they introduced ImageOP and 1613 labeled facade images from 32 religious monuments in Ouro Preto, and benchmarked MobileNet V2 alongside EfficientNet B0 for real-time mobile classification [15]. They later expanded ImageOP to ImageMG, adding 6449 images from

94 buildings in Minas Gerais [6]. Together, these studies show that CNN transfer learning with region-specific datasets can identify architectural styles.

3. Materials and Methods

3.1. Study Domain

The study examines geographic areas of Saudi Arabia using the Saudi Architecture Map, as seen in Figure 1, which links each region to one of the nineteen architectural styles and their environmental and cultural influences. Table 1 lists 19 architectural styles and 54 representative cities. The map provides geospatial context for interpreting classification results; it is not used as a model input.

Table 1. Number of representative cities associated with each Saudi architectural style.

No.	Architectural style	Cities	No.	Architectural style	Cities
1	Northern Najdi	4	11	Tuhama Coast	3
2	Central Najdi	3	12	Farasan Islands	1
3	Tabuk Coast	3	13	Abha Highland	3
4	Madinah Rural	2	14	Bisha Desert	2
5	Inner Madinah	3	15	Najran	2
6	Hejazi Coast	4	16	Al Ahsa Oasis	2
7	Taif Highland	3	17	Al Qatif Oasis	2
8	Sarawat Mountains	3	18	East Coast	4
9	Aseer Escarpment	2	19	Eastern Najdi	3
10	Tuhama Foothills	3		**Total**	**54**

3.2. Research Framework

This research was structured into five phases, summarized in Figure 2. Phase 1 covers the creation of the multi-source dataset. We physically separated full-building façade images from architectural-element images to prevent data leakage. Phase 2 highlights the data preprocessing stage, including augmentation for under-represented styles and class-weight computation. Phases 3 and 4 cover the two basic steps of the transfer learning strategy. Phase 3 aims to pretrain the model using Saudi Architecture Map elements, while Phase 4 aims to benchmark five different backbone DL models using façade images. During this phase, we benchmarked five CNN backbones with identical hyperparameters and quantitatively evaluated performance against four accuracy metrics: producer accuracy, user accuracy, overall accuracy and F1-Score. In Phase 5, we deployed the best-performing model through the Gradio-based decision-support interface. The interface accepts a building façade image and returns the underlying architectural style. It classifies the image using the top three predictions, with a color-coded confidence indicator.

3.3. Data Collection

We collected building images from publicly available repositories, including Getty Images, iStock, Alamy, TripAdvisor, and Flickr. We used the architectural style names defined by the Saudi Architecture Map as search queries. We excluded Inner Madinah because no images of that style were available. Hence, the classifiers cover 18 of these 19 officially recognized styles. The official regional design guidelines published alongside the Saudi Architecture Map, which describe characteristic materials, façade elements, colors, and textures for each style, served as a reference for selecting and verifying image quality. An image was included in the dataset if it was returned by a style-name search query on one of the five source platforms and passed manual visual verification against the corresponding style's official design-guideline description. We excluded Inner Madinah from the study because we could not source usable imagery for it under this procedure. We compiled

images by searching each platform (Getty Images, iStock, Alamy, TripAdvisor, and Flickr) by architectural style name and downloading the results. We did not purchase or otherwise secure individual per-image licenses beyond each platform's default terms of access. Usage terms vary substantially by source and by individual image: a subset of the Flickr imagery carries a Creative Commons license (the specific license type was not tracked per image during collection), while imagery from Getty Images, iStock, and Alamy is governed by those platforms' standard commercial/editorial terms, and TripAdvisor imagery is user-uploaded content under that platform's terms of service. A three-person review team labeled images jointly. Rather than labeling independently, the team reviewed and assigned style labels together for each image, cross-referencing the official regional design guidelines. When judgments diverged, the team's domain expert in Saudi architectural heritage served as the final determination.

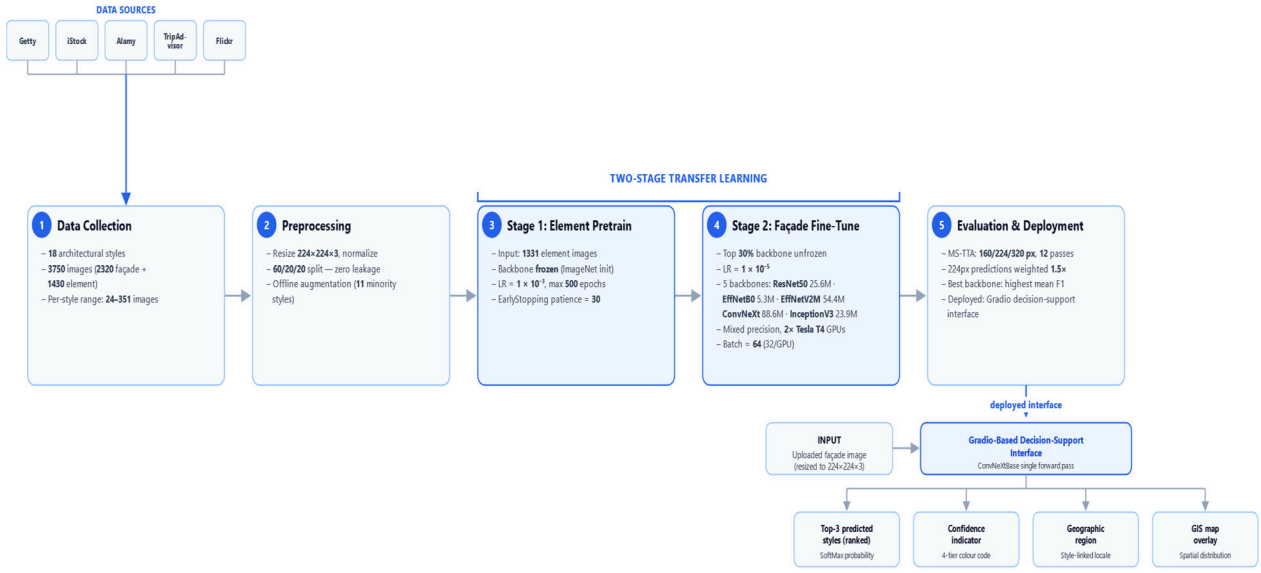


Figure 2. Methodological flowchart of the five-phase research framework.

During data collection, we obtained two types of images for each style:

(1) The first type contains architectural elements/parts that capture localized features such as doors, ornaments, openings, decorative patterns, and material textures. We obtained these images from the official Saudi Architecture Map website.

(2) The second type includes full-building façade images, used for style-level classification in Stage 2 of the training protocol.

The final curated dataset comprises 3750 images: 2320 full-building façade images and 1430 architectural-element images, as shown in Table 2. All eighteen styles are represented in the building façade dataset, enabling a complete multi-class supervised classification experiment. The most represented styles in the building façade dataset are Coastal Hejazi (351 images), Central Najdi (233 images), and Abha Highland (229 images), while the least represented are Eastern Najdi (24 images), East Coast (34 images), and Tabuk Coast (54 images). All images were collected at a resolution sufficient for 224 × 224-pixel preprocessing and were manually verified against the style descriptions in the official design guidelines.

Table 2. Number of buildings and parts for each Saudi architectural style.

No.	Architectural Style	Parts	Buildings	No.	Architectural Style	Parts	Buildings
1	Abha Highland	87	229	10	Farasan Island	87	144
2	Al Ahsa Oasis	75	82	11	Madinah Rural	61	113
3	Al Qatif Oasis	65	80	12	Najran	66	124
4	Aseer Escarpment	91	179	13	Northern Najdi	69	177
5	Bisha Desert	73	45	14	Sarawat Mountains	95	164
6	Central Najdi	91	233	15	Tabuk Coast	76	54
7	Coastal Hejazi	87	351	16	Taif Highland	88	127
8	East Coast	79	34	17	Tuhama Coast	89	54
9	Eastern Najdi	65	24	18	Tuhama Foothills	86	106
	Total	**1430**	**2320**		**Grand Total:**	**3750 images**	

3.4. Data Preparation

We organized all images into folders by architectural style. Some architectural classes contain many building images, such as Coastal Hejazi (351 images), whereas others, such as Eastern Najdi, contain only 24 images. The class-imbalance ratio between the largest and smallest building styles was approximately 15:1.

The architectural data suffers from class imbalance. So, we addressed it at two complementary levels:

1- We applied augmentation for the eleven under-represented training styles.
2- We computed separate class weights for Stage 1 and Stage 2 and passed them to the loss because the class distributions differ between the two stages.

To prevent data leakage between the two training stages, we used architectural-element images exclusively in Stage 1 and excluded them from any validation or test set. The façade images were split into 60% training, 20% validation, and 20% test. After splitting, Stage 1 contained 1331 element images; Stage 2 comprised 1622 training images from the 60/20/20 split itself, 302 validations, and 396 held-out test images. After offline augmentation of the eleven under-represented styles (298 additional images), the effective Stage 2 training set used for model fitting totaled 1920 images. Of the 396 intended held-out test images, 389 were successfully scored during evaluation (7 images across five styles—Abha Highland, Bisha Desert, Central Najdi, Coastal Hejazi, and Tuhama Foothills—could not be loaded at evaluation time); reported test-set metrics are computed on these 389 images.

All images were resized to 224 × 224 pixels to match the input size expected by ImageNet-pretrained backbones. Eleven training styles have fewer than 80 images. Therefore, we augmented the training set with horizontal flips, rotations of $\pm 7^{\circ}$, brightness of $\pm 20\%$, contrast of +30%, and a 0.85× zoom, and saved the results to disk before training. This augmentation step was applied strictly after the building/site-group-aware train/validation/test split, as clarified in Table 3, and only to images already assigned to the training partition, so no augmented or duplicated image could appear in the validation or test sets. In total, we generated 298 augmented images across these eleven styles, bringing each to the 80-image training target shown in Table 3. This offline augmentation targeted only the eleven under-represented training styles. Independently, a second, real-time augmentation layer was applied in-graph to every training image in both stages, on every epoch: random horizontal flip, rotation of up to $\pm 18^{\circ}$, zoom of up to 15%, contrast adjustment of up to $\pm 20\%$, and brightness adjustment of up to $\pm 15\%$. This real-time layer was not previously disclosed and is reported here for completeness.

Table 3. The prepared data used during training.

Architectural Style	Parts	Train (Original)	Train (Augmented)	Train (Total)	Validation	Test
Abha Highland	87	140	0	140	47	42
Al Ahsa Oasis	75	49	31	80	20	13
Al Qatif Oasis	65	69	11	80	0	11
Aseer Escarpment	91	107	0	107	35	37
Bisha Desert	73	37	36	73	0	8
Central Najdi	91	189	0	189	0	44
Coastal Hejazi	87	264	0	264	32	55
East Coast	79	24	56	80	2	8
Eastern Najdi	65	14	66	80	4	6
Farasan Island	87	116	0	116	0	28
Madinah Rural	61	89	0	89	0	24
Najran	66	74	6	80	24	26
Northern Najdi	69	106	0	106	46	25
Sarawat Mountains	95	98	0	98	32	34
Tabuk Coast	76	53	27	80	0	1
Taif Highland	88	98	0	98	3	26
Tuhama Coast	89	32	48	80	18	4
Tuhama Foothills	86	63	17	80	39	4
Total	1430	1622	298	1920	302	396

3.5. Deep Learning Model Training, Validation and Testing

We used a two-stage transfer-learning protocol for each CNN backbone.

- Phase 3 (Stage 1):
  We trained the classifier head on architectural-element images (materials, doors, decorative patterns, openings), while all backbone weights were frozen at their ImageNet initialization. We used the following hyperparameters:
  1. A learning rate of 0.001;
  2. A maximum ceiling of 500 epochs;
  3. A patience of 30 on validation loss.
- Phase 4 (Stage 2):
  To ensure a fair comparison, we benchmarked five CNN backbones using the same hyperparameters:
  4. ResNet50 (25.6 M parameters);
  5. EfficientNetB0 (5.3 M parameters);
  6. EfficientNetV2M (54.4 M parameters);
  7. ConvNeXtBase (88.6 M parameters);
  8. InceptionV3 (23.9 M parameters).

  For each backbone model, the architecture followed the same pipeline:
  9. An input layer of $224 \times 224 \times 3$;
  10. The backbone's preprocessing layer;
  11. The pretrained backbone (ImageNet weights);
  12. A Global Average Pooling layer;
  13. A Dropout layer with a rate of 0.3;
  14. A final Dense layer outputs eighteen logits in float32, stabilizing mixed-precision training.

During this stage, we froze the bottom 70% of the backbone layers to preserve low-level features (edges, textures) and fine-tuned only the top layers. We reduced the learning rate to 1×10^{-5} to avoid forgetting ImageNet features. This two-stage design mirrors the learning

progression of a heritage specialist: first recognize architectural elements, then identify the regional style based on the full façade. Both stages used the Adam optimizer and minimized a sparse categorical cross-entropy loss, with a global batch size of 64 (32 images per GPU across 2 NVIDIA Tesla T4 GPUs under a mirrored-strategy distribution policy) and mixed-precision (float16) training in a Kaggle notebook environment. Class weights were computed independently for Stage 1 and Stage 2 using scikit-learn's balanced formula (weight per class is inversely proportional to its sample count) and passed to the loss function; weights ranged from 0.84 to 1.22 in Stage 1 and 0.46 to 1.43 in Stage 2, reflecting the styles' differing representation. A fixed random seed (42) was used for the Stage 1 element-image validation split. Results reflect a single training run per backbone rather than an average across repeated runs.

We evaluated the model on 396 held-out images (389 successfully scored) using standard geomatics metrics.

15. Overall Accuracy: Fraction of correctly classified images.
16. Producer Accuracy: Probability of an image of a given style being classified into that style (the complement of omission error).
17. User Accuracy: Probability of an image predicted to be of a given style actually belonging to that style (the complement of the commission error).
18. F1-Score: The harmonic mean of producer accuracy and user accuracy.

3.6. DL Model Deployment: Gradio-Based Decision-Support Interface

To translate the trained model into an operational decision-support tool, we built a Gradio-based decision-support interface that embeds the best-performing model. The pipeline accepts an image and applies multi-scale test-time augmentation (MS-TTA), resizing and cropping the image at three scales:

19. 160×160—Original and horizontal flip (2 views).
20. 224×224—Original and horizontal flip, and two small rotations ($\pm 5^{\circ}$) (4 views).
21. 320×320, center-cropped back to 224×224—Original and horizontal flip (2 views).

This yields eight augmented views in total, each passed through ConvNeXtBase independently and combined via a weighted average into the final prediction.

We run the system in a notebook and launch it manually after training because Gradio needs an active kernel. The interface returns the top three predictions, ranked by probability, along with a color-coded confidence indicator: dark green for very-high confidence ($\geq$95%), green for high confidence (85–94%), yellow for moderate confidence (70–84%), and red for below 70%. These SoftMax outputs are presented to the user as a color-coded confidence indicator, but they are raw model probabilities, not calibrated confidence values. The reported value, such as 95%, should not be interpreted as a 95% empirical likelihood of correctness.

This tool is intentionally designed as a decision-support layer rather than a fully automated classifier. The four-tier color scheme makes model uncertainty explicit and aligns with heritage-management practice, in which automated outputs must be cross-checked by a domain expert before being added to an official inventory. The system is deployed in a notebook environment and launched manually after model training because the Gradio server requires a live computational kernel.

4. Results

4.1. Dataset Creation and Integration

The dataset contains 2320 full-building façades and 1430 architectural-element images. This is 3750 images in total, covering eighteen architectural styles (Inner Madinah excluded).

After the 60/20/20 split of the building façade dataset, we had 1920 images for training, 302 for validation, and 396 for the held-out test set. After data augmentation, we targeted approximately 80 training images per style, augmenting styles whose original building images fell below 80. The final prepared data is outlined in Table 3.

4.2. Deep Learning Model Training and Validation

Table 4 summarizes the performance of the five CNN backbones evaluated under identical hyperparameters using a single-pass forward evaluation on the held-out test set (multi-scale test-time augmentation was used only in the deployed Gradio interface described in Section 3.6, not in this benchmark evaluation) (Figure 3). ConvNeXtBase performed the best across all four metrics, with an overall accuracy of 90.75%, an F1-Score of 88.06%, a producer accuracy of 92.05%, and a user accuracy of 86.65%. Producer and user accuracy were closely matched across five CNN backbones. This indicates a balanced trade-off between omission and commission errors across the eighteen classes. Several classes in Table 5 report 100% producer accuracy on very small held-out test counts (e.g., East Coast, n = 8; Eastern Najdi, n = 6; Al Ahsa Oasis, n = 13; Najran, n = 26); 95% Wilson confidence intervals for these classes are wide (e.g., East Coast [67.6%, 100.0%]; Eastern Najdi [61.0%, 100.0%]; Tabuk Coast, n = 1, [20.7%, 100.0%]), and these point estimates should be interpreted with that uncertainty in mind rather than as precise performance figures.

Table 4. The CNN backbones' performance across 5 models.

Model	Overall Accuracy	Producer Accuracy	User Accuracy	F1-Score
ResNet50	83.80	81.75	73.87	75.24
EfficientNetB0	84.06	83.84	77.07	77.56
EfficientNetV2M	85.35	87.48	81.16	81.26
ConvNeXtBase	90.75	92.05	86.65	88.06
InceptionV3	78.66	76.93	77.27	75.90

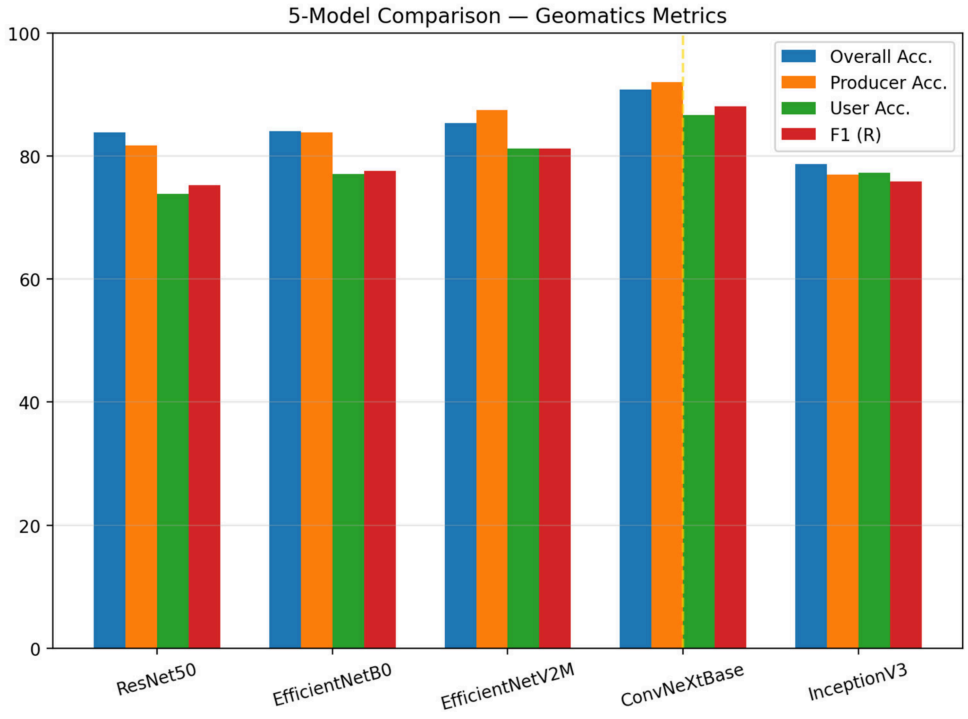


Figure 3. The performance of the five CNN backbones.

Early stopping terminated Stage 1 training at 103 epochs for ResNet50, 158 epochs for EfficientNetB0, 152 epochs for EfficientNetV2M, 129 epochs for ConvNeXtBase, and 72 epochs for InceptionV3. Stage 2 converged after 60, 182, 81, 87, and 62 epochs for the five models, respectively. The convergence is illustrated in Figure 4.

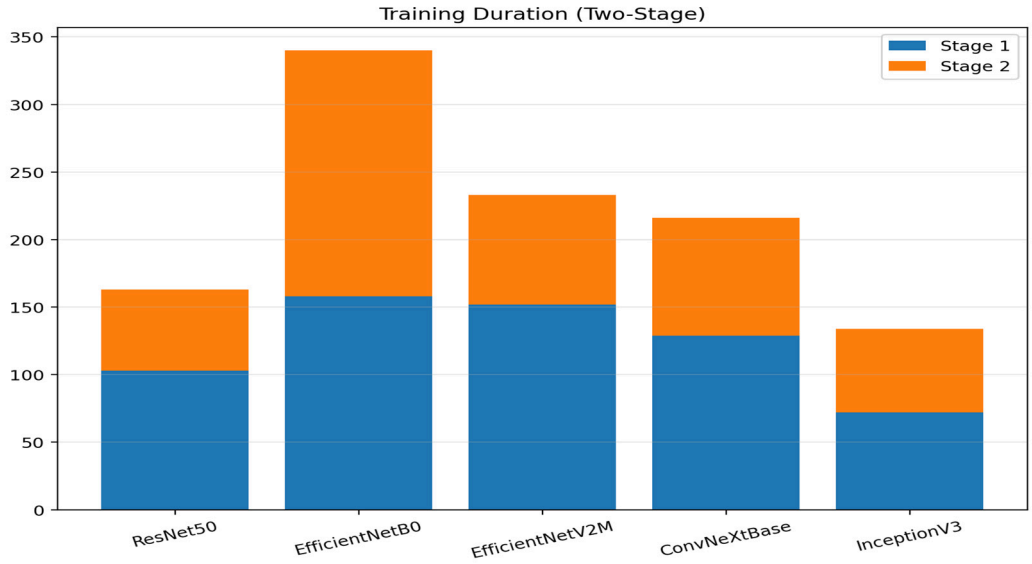


Figure 4. Number of epochs during data training across the five CNN backbones.

4.3. Ablation Study

To empirically validate the two-stage training design, we ran a controlled single-stage ablation: ConvNeXtBase fine-tuned directly on façade images using ImageNet weights, skipping Stage 1 entirely. We froze all other hyperparameters (backbone unfreeze ratio, learning rate, epochs/patience, augmentation, class weights, the corrected group-aware split, and random seed) and held them identical to the two-stage run. As seen from Table 5, the two-stage design outperformed the single-stage baseline across all four metrics: overall accuracy (90.75% vs. 88.95%), mean F1 (88.06% vs. 84.55%), mean producer accuracy (92.05% vs. 89.56%), and mean user accuracy (86.65% vs. 83.83%). This provides direct empirical evidence that the element-level Stage-1 pretraining step meaningfully contributes to final performance.

Table 5. Ablation comparison—two-stage vs. single-stage training (ConvNeXtBase).

Training Variant	Overall Accuracy	Producer Accuracy	User Accuracy	F1-Score
Two-Stage (ConvNeXtBase)	90.75	92.05	86.65	88.06
Single-Stage (ConvNeXtBase)	88.95	89.56	83.83	84.55

4.4. Model Deployment Results

ConvNeXtBase was chosen to be deployed into the Gradio-based decision-support interface because it had the highest mean F1-Score. Its per-class performance on the held-out test set—precision (user accuracy), recall (producer accuracy), and F1-Score for each of the 18 architectural styles—is reported in Table 6.

The workflow of the Gradio-based decision-support interface is illustrated in Figure 5. It starts by asking the user to either drag-and-drop or upload the façade-based building image. Then, the interface will display the classified architectural style along with the confidence score and geographic region. The interface integrates a GIS-based interactive map, directly linking the ConvNeXtBase predictions with the precise spatial distribution of the regional styles.

Table 6. Per-class performance of the deployed ConvNeXtBase model on the held-out test set.

Architectural Style	Producer Accuracy (%)	User Accuracy (%)	F1-Score (%)	Test Images (n)
East Coast	100.0	100.0	100.0	8
Tuhama Coast	100.0	100.0	100.0	4
Tuhama Foothills	100.0	100.0	100.0	3
Coastal Hejazi	100.0	98.2	99.1	54
Farasan Island	100.0	93.3	96.5	28
Aseer Escarpment	94.6	94.6	94.6	37
Abha Highland	87.5	100.0	93.3	40
Sarawat Mountains	97.1	89.2	93.0	34
Najran	100.0	86.7	92.9	26
Taif Highland	88.5	95.8	92.0	26
Al Ahsa Oasis	100.0	81.2	89.7	13
Madinah Rural	91.7	84.6	88.0	24
Al Qatif Oasis	81.8	90.0	85.7	11
Eastern Najdi	100.0	75.0	85.7	6
Northern Najdi	84.0	80.8	82.3	25
Central Najdi	65.1	90.3	75.7	43
Bisha Desert	66.7	66.7	66.7	6
Tabuk Coast	100.0	33.3	50.0	1

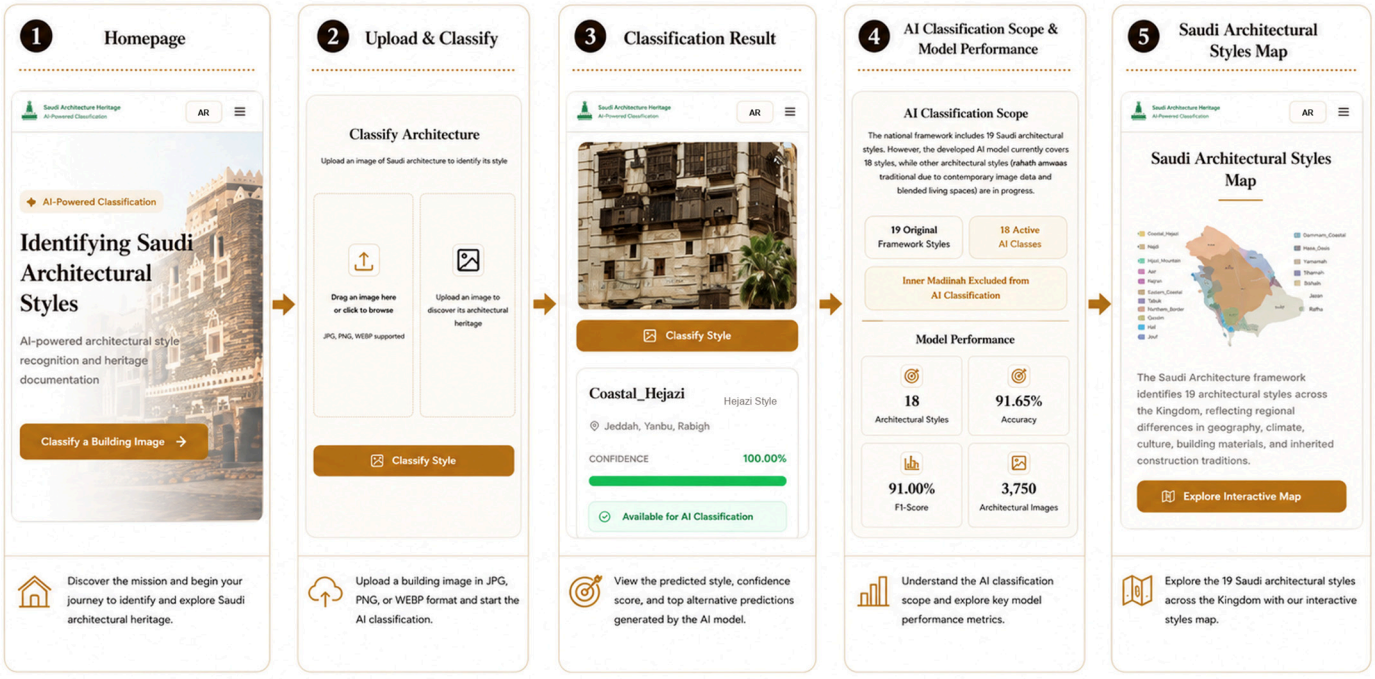


Figure 5. Gradio-based decision-support interface workflow.

Most misclassifications cluster among visually and geographically adjacent styles as outlined in Figure 6. Central Najdi is the most frequently confused class, with only 28 of its 43 test images (65.1%) correctly classified; the remaining 15 errors spread across six other styles, most often Northern Najdi (five images, 11.6%) and Najran (four images, 9.3%), with smaller numbers going to Bisha Desert, Eastern Najdi (two each), and Al Ahsa Oasis and Sarawat Mountains (one each). This suggests overlapping visual characteristics with several neighboring interior/Najdi-region styles rather than a single dominant rival. A similar pattern appears between the two oasis styles, where two of 11 (18.2%) Al Qatif Oasis test images were misclassified as Al Ahsa Oasis. Bisha Desert's low producer accuracy in Table 6 is explained by two of its six (33.3%) test images being predicted as Madinah

Rural. These region-coherent confusions suggest the model is picking up genuine but overlapping regional architectural cues rather than making arbitrary errors. The per-class precision/recall/F1 for all five backbones is provided in Appendix B.

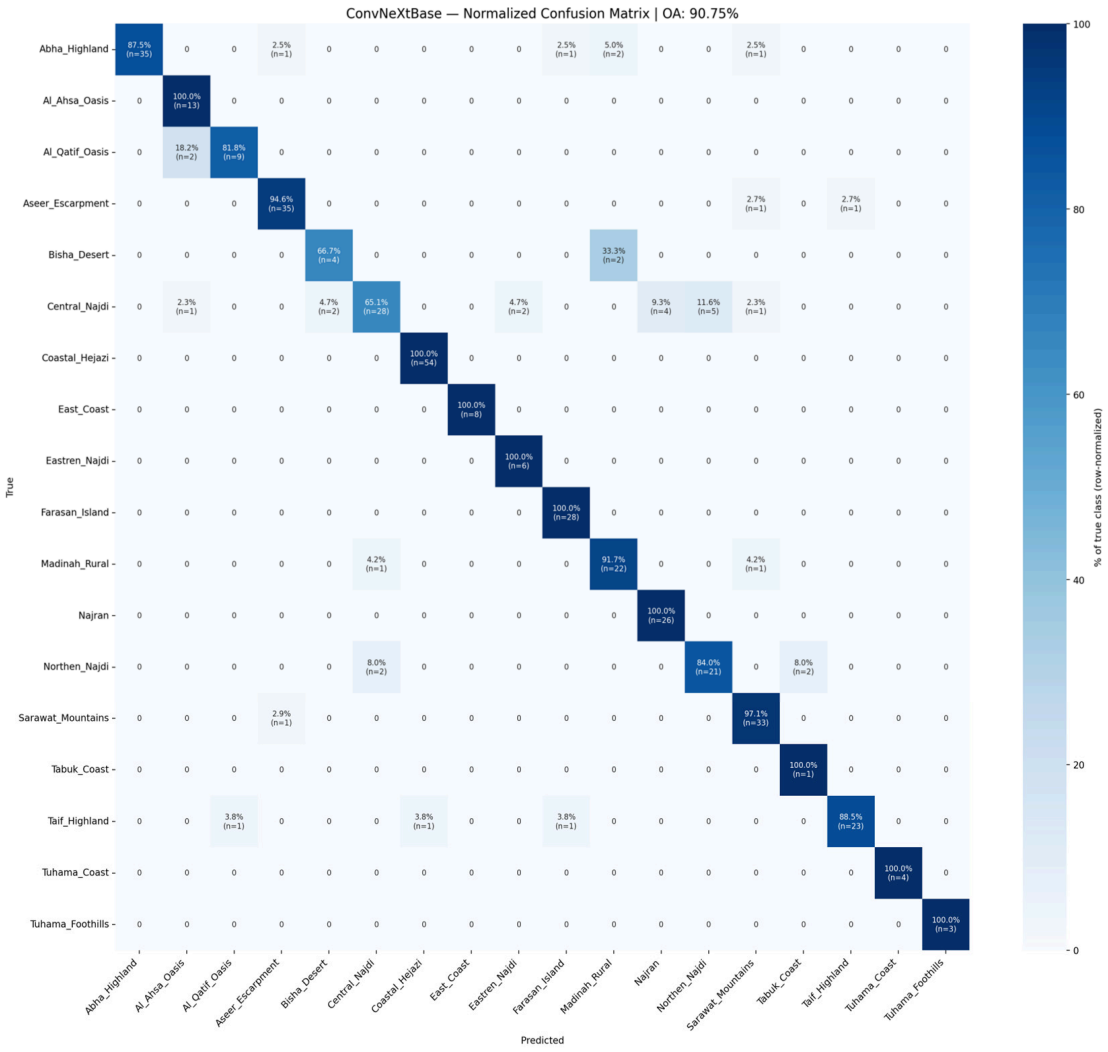


Figure 6. The row-normalized confusion matrix of the deployed ConvNeXtBase model on the held-out test set.

5. Discussion

We conducted a deep-learning-based study that automatically classifies historical buildings into 18 of the 19 officially recognized Saudi architectural styles (Inner Madinah was excluded because no usable imagery could be sourced for it). Our results indicate that a two-step transfer-learning framework captures the visual characteristics of Saudi regional architectural styles. On an eighteen-class problem with strong intra-cluster visual similarity between classes, ConvNeXtBase achieves an overall accuracy of 90.75% and an F1-Score of 88.06%. The similar producer (92.05%) and user (86.65%) accuracies show the model's predictions are balanced across styles. The confidence-tiered decision-support design is intended to mitigate, not eliminate, the risks introduced by the dataset's size, class imbalance, and the absence of formal confidence calibration; the system should be regarded as a support tool for expert review rather than a deployment-ready classifier for heritage documentation or compliance monitoring.

These findings extend the line of work initiated by Llamas et al. [8], Abed et al. [9], and Cosovic and Jankovic [10]. They are consistent with the broader trend reported in the

deep-learning heritage literature. These studies demonstrate that combining pretrained CNNs with transfer learning outperforms handcrafted-feature classifiers for heritage element recognition. This element-level supervision literature study motivated the two-stage training design: first on architectural elements and then on full facade-based imagery. Shehata and Alaboud demonstrated the value of element-level supervision for facade analysis [7], which mirrors the present study's design choice to dedicate a separate Stage 1 to architectural elements (materials, doors, patterns) before fine-tuning on full facades.

On a broader scale, the present study compares with city- or province-scale pipelines, such as the studies from Siountri et al. [11] for Athens, Zou et al. [12] and Tan [13] for Hubei Province, and Ottoni and Ottoni [15] for the heritage town of Ouro Preto. The present study extends the spatial coverage to the Kingdom of Saudi Arabia using the launched Saudi Architecture Map. To the best of our knowledge, this study represents the first attempt to use a deep learning approach to map Saudi architecture spatially.

Additionally, the Gradio-based decision-support interface adds a practical layer to facilitate the work of practitioners and decision-makers. The color-coded confidence ranges translate the SoftMax probability into a decision-support signal. Because these confidence values are uncalibrated SoftMax outputs rather than verified probabilities of correctness, they should inform, not replace, expert review, reinforcing why the system is positioned as a decision-support aid rather than an autonomous classifier.

22. Green $\geq$ 85%, for high-confidence predictions.
23. Yellow 70–84%, for moderate-confidence predictions, which can be flagged for expert review.
24. Red < 70, for low-confidence predictions, which also can be flagged for expert review.

Beyond this heritage-preservation Gradio-based decision-support tool, the automated classification pipeline proposed here contributes to sustainability in several concrete ways. For sustainable urban development, it enables compliance monitoring against the Saudi Architecture Map's regional design guidelines to scale with the pace of construction, rather than being bottlenecked by manual expert review as urban growth accelerates. For heritage conservation, automated, repeatable classification helps identify stylistically non-compliant new construction earlier in the planning process before it is built, rather than relying on after-the-fact remediation. For long-term heritage management, the classifier's outputs, combined with the GIS layer already integrated into the deployed interface, can populate a structured, updatable digital record of architectural style distribution that supports longitudinal monitoring of heritage preservation over time, in place of one-off manual surveys that go stale as cities change.

Limitations are inevitable in any study. The first limitation concerns the dataset size and source: our dataset contains 2320 full-building façade images across eighteen styles, and we excluded the Inner Madinah style because no images were available. As a result, we could not evaluate model performance on this style, and classification is currently limited to the eighteen styles for which data exists, pending collection of additional images. The second limitation is the dataset's bias toward canonical, idealized facades rather than the full range of real-world variation. Because most images were drawn from official design guidelines and Supplementary Materials—sources that, by nature, depict representative, idealized exemplars—the dataset likely under-represents non-canonical or atypical facades within each style, which may limit the model's ability to generalize to real-world buildings that deviate from guideline imagery. The third limitation concerns the experimental design itself and comprises three related issues. (a) Results reflect a single training run per backbone rather than an average across repeated runs with different random seeds, so confidence intervals are not available for the reported metrics. (b) Data labeling was done via joint consensus by a three-person review team rather than independently by

multiple raters, so we could not compute a formal inter-annotator agreement statistic, such as Cohen's or Fleiss' κ. (c) Finally, the deployed interface's GIS component is currently limited to a visualization/context layer rather than a model input, meaning climate and geographic factors were treated as backgrounds rather than fed into the model. As a result, this information does not yet inform the classification itself. Integrating these factors as auxiliary GIS features might reduce confusion between visually similar styles.

These limitations point to five directions for future work. First, expand the dataset to include under-represented styles and, preferably, add field-collected and street-level photos. Second, integrate the visual stream with geospatial data (latitude, longitude, climatic zone, topography). In other words, use GIS as a spatial filter rather than only for recommendations. Third, deploy the trained model on an interactive map. Fourth, evaluate additional architectures beyond the five CNN backbones already benchmarked here, using the same two-stage transfer-learning steps. Fifth, extend this ablation to the remaining untested hyperparameters, such as training without offline augmentation, training without class weights, and alternative backbone-freezing ratios, with repeated training runs across multiple random seeds to establish confidence intervals for the reported metrics.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/su18189313/s1, the two-stage training notebook (Saudi_Architecture_CORRECTED.ipynb), implementing the full data split, augmentation, class-weighting, and two-stage transfer-learning pipeline for all five CNN backbones under the corrected building/site-group-aware split; and the single-stage ablation notebook (Saudi_Architecture_Ablation_SingleStage.ipynb).

Author Contributions: Conceptualization, R.S.; Methodology, R.S.; Software, A.F.; Validation, A.F.; Formal analysis, A.F.; Investigation, R.S.; Resources, R.S.; Data curation, A.F.; Writing—original draft, M.A.; Writing—review & editing, M.A.; Visualization, A.A.; Supervision, A.A.; Project administration, A.A.; Funding acquisition, M.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, Saudi Arabia, for funding this project under grant No. (IPP: 472-612-2026) and for providing technical support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

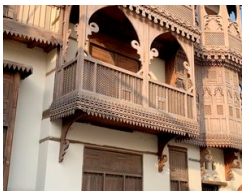
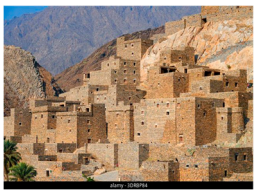
Data Availability Statement: The training and evaluation code is provided as Supplementary Materials. The raw façade and architectural-element images cannot be publicly redistributed, as they were sourced from third-party platforms (Getty Images, iStock, Alamy, TripAdvisor, Flickr) under those platforms' own terms of access rather than under a license permitting redistribution. Derived data—including class labels, the building/site-group mapping used for the corrected split, and the trained model weights for all five backbones and the ablation baseline—are available from the corresponding author upon reasonable request.

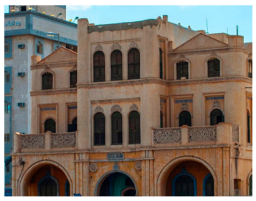
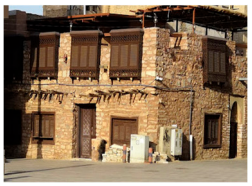
Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. The Confidence Scores Returned by the Gradio-Based Decision-Support Interface for Each Image (Sample)

We tested the deployed interface on a total of 177 sample cases, using façade images from each architectural style; each sample represents the same building from a different angle or zoom level. The per-image confidence scores returned by the deployed ConvNeXtBase model for each sample are reported below. These 177 samples are a subset of the formal held-out test set (approximately ten per style, fewer for styles with limited available imagery, e.g., seven for East Coast), selected to demonstrate the deployed interface's

per-image confidence-score behavior across styles qualitatively. They are not a separate or independent sample, and are not used to re-evaluate accuracy.

Hejazi Coast				
				
Confidence Score: 89.07	**Confidence Score: 99.96**	**Confidence Score: 99.99**	**Confidence Score: 100**	**Confidence Score: 99.19**
Northern Najdi				
				
Confidence Score: 99.79	**Confidence Score: 95.67**	**Confidence Score: 85.28**	**Confidence Score: 77.02**	**Confidence Score: 96.46**
Sarawat Mountain				
				
Confidence Score: 98.85	**Confidence Score: 99.99**	**Confidence Score: 98.34**	**Confidence Score: 82.71**	**Confidence Score: 82.1**
Najran				
				
Confidence Score: 98.74	**Confidence Score: 99.6**	**Confidence Score: 100**	**Confidence Score: 97.55**	**Confidence Score: 99.65**

Abha Highlands				
				
Confidence Score: 99.53	**Confidence Score: 99.76**	**Confidence Score: 97.24**	**Confidence Score: 84.09**	**Confidence Score: 95.51**
Al Ahsa Oasis				
				
Confidence Score: 99.75	**Confidence Score: 99.3**	**Confidence Score: 99.05**	**Confidence Score: 98.7**	**Confidence Score: 95.95**
				
Confidence Score: 99.99	**Confidence Score: 97.43**	**Confidence Score: 98.94**	**Confidence Score: 96.61**	**Confidence Score: 99.99**
				
Confidence Score: 87.01	**Confidence Score: 100**	**Confidence Score: 99.99**	**Confidence Score: 99.81**	**Confidence Score: 99.87**

https://doi.org/10.3390/su18189313









Confidence Score: 98.93	Confidence Score: 99.05	Confidence Score: 99.34	Confidence Score: 75.82	Confidence Score: 92.21









Confidence Score: 79.10	Confidence Score: 99.99	Confidence Score: 99.55	Confidence Score: 78.18	Confidence Score: 82.10

Architectural Style	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Northern Najdi	99.79	95.67	85.28	77.02	96.46	96.79	98.5	97.45	95.33	88.38
Central Najdi	100	81.1	92.41	88.11	96.21	98.61	99.75	98.44	99.72	99.82
Tabuk Coast	99.42	40.95	19.78	80.46	85.41	99.68	82.15	84.44	88.04	40.92
Hejazi Coast	89.07	99.96	99.99	100	99.19	100	99.94	99.34	99.89	99.34
Sarawat Mountains	98.85	99.99	98.34	82.71	82.1	98.86	99.81	98.59	97.52	96.87
Abha Highland	99.53	99.76	97.24	84.09	95.51	62.18	78.73	97.2	92.18	99.99
Najran	98.74	99.6	100	97.55	99.65	96.6	99.99	99.81	92.96	99.99
AL Ahsa Oasis	99.75	99.3	99.05	98.7	95.95	99.96	82.44	96.55	99.8	92.34
Aseer Escarpment	99.97	99.99	99.73	93.47	100	99.98	95.41	99.98	99.93	99.76
Tuhama Coast	78.8	93.99	96.71	52.66	91.55	86.65	99.2	99.79	99.79	98.57
Tuhama Foothills	0	96.65	92.73	0	25.77	98.67	72.1	94.42	75.88	95.93
Farasan Islands	87.01	100	99.99	99.81	99.99	99.87	99.97	98.53	100	99.97
Al Qatif Oasis	79.1	99.99	99.55	78.18	99.84	85.82	98.02	82.1	18.36	25.75
East Coast	99.96	99.94	99.6	99.87	98.58	99.56	98.8			
Taif Highland	99.99	97.43	98.94	96.61	99.99	99.78	99.96	99.93	88.79	99.92
Madinah Rural	98.93	99.05	99.34	75.82	92.21	99.76	93.27	61.47	98.92	82.99
Bisha Desert	99.99	99.94	99.77	83	99.96	71.28	0	95.29	99.94	99.87
Eastern Najdi	89.73	56.26	94.71	99.46	91.29	98.53	64.67	99.05	98.24	84.9

Appendix B. Per-Class Precision (User Accuracy), Recall (Producer Accuracy), and F1-Score for All Five CNN Backbones on the Held-Out Test Set (Corrected Building/Site-Group-Level Split, Seed = 42)

Architectural Style	Backbone	Producer Accuracy (%)	User Accuracy (%)	F1-Score (%)	Test Images (n)
Abha Highland	ResNet50	77.5	100.0	87.3	40
	EfficientNetB0	90.0	100.0	94.7	40
	EfficientNetV2M	80.0	100.0	88.9	40
	ConvNeXtBase	87.5	100.0	93.3	40
	InceptionV3	77.5	86.1	81.6	40
Al Ahsa Oasis	ResNet50	61.5	66.7	64.0	13
	EfficientNetB0	53.9	87.5	66.7	13
	EfficientNetV2M	92.3	60.0	72.7	13
	ConvNeXtBase	100.0	81.2	89.7	13
	InceptionV3	69.2	75.0	72.0	13
Al Qatif Oasis	ResNet50	72.7	72.7	72.7	11
	EfficientNetB0	81.8	90.0	85.7	11
	EfficientNetV2M	90.9	76.9	83.3	11
	ConvNeXtBase	81.8	90.0	85.7	11
	InceptionV3	72.7	66.7	69.6	11
Aseer Escarpment	ResNet50	86.5	88.9	87.7	37
	EfficientNetB0	83.8	91.2	87.3	37
	EfficientNetV2M	86.5	91.4	88.9	37
	ConvNeXtBase	94.6	94.6	94.6	37
	InceptionV3	89.2	84.6	86.8	37
Bisha Desert	ResNet50	50.0	50.0	50.0	6
	EfficientNetB0	50.0	75.0	60.0	6
	EfficientNetV2M	50.0	100.0	66.7	6
	ConvNeXtBase	66.7	66.7	66.7	6
	InceptionV3	50.0	60.0	54.5	6
Central Najdi	ResNet50	74.4	94.1	83.1	43
	EfficientNetB0	72.1	93.9	81.6	43
	EfficientNetV2M	53.5	100.0	69.7	43
	ConvNeXtBase	65.1	90.3	75.7	43
	InceptionV3	55.8	77.4	64.9	43
Coastal Hejazi	ResNet50	96.3	94.5	95.4	54
	EfficientNetB0	90.7	94.2	92.5	54
	EfficientNetV2M	96.3	96.3	96.3	54
	ConvNeXtBase	100.0	98.2	99.1	54
	InceptionV3	94.4	86.4	90.3	54
East Coast	ResNet50	100.0	100.0	100.0	8
	EfficientNetB0	100.0	88.9	94.1	8
	EfficientNetV2M	100.0	100.0	100.0	8
	ConvNeXtBase	100.0	100.0	100.0	8
	InceptionV3	100.0	100.0	100.0	8

Architectural Style	Backbone	Producer Accuracy (%)	User Accuracy (%)	F1-Score (%)	Test Images (n)
Eastern Najdi	ResNet50	83.3	62.5	71.4	6
	EfficientNetB0	100.0	66.7	80.0	6
	EfficientNetV2M	83.3	71.4	76.9	6
	ConvNeXtBase	100.0	75.0	85.7	6
	InceptionV3	50.0	100.0	66.7	6
Farasan Island	ResNet50	89.3	89.3	89.3	28
	EfficientNetB0	89.3	83.3	86.2	28
	EfficientNetV2M	96.4	90.0	93.1	28
	ConvNeXtBase	100.0	93.3	96.5	28
	InceptionV3	92.9	81.2	86.7	28
Madinah Rural	ResNet50	91.7	64.7	75.9	24
	EfficientNetB0	70.8	68.0	69.4	24
	EfficientNetV2M	91.7	68.8	78.6	24
	ConvNeXtBase	91.7	84.6	88.0	24
	InceptionV3	70.8	63.0	66.7	24
Najran	ResNet50	96.2	83.3	89.3	26
	EfficientNetB0	100.0	81.2	89.7	26
	EfficientNetV2M	100.0	78.8	88.1	26
	ConvNeXtBase	100.0	86.7	92.9	26
	InceptionV3	80.8	67.7	73.7	26
Northern Najdi	ResNet50	76.0	86.4	80.8	25
	EfficientNetB0	76.0	86.4	80.8	25
	EfficientNetV2M	80.0	83.3	81.6	25
	ConvNeXtBase	84.0	80.8	82.3	25
	InceptionV3	76.0	82.6	79.2	25
Sarawat Mountains	ResNet50	85.3	85.3	85.3	34
	EfficientNetB0	91.2	73.8	81.6	34
	EfficientNetV2M	85.3	87.9	86.6	34
	ConvNeXtBase	97.1	89.2	93.0	34
	InceptionV3	82.3	73.7	77.8	34
Tabuk Coast	ResNet50	100.0	25.0	40.0	1
	EfficientNetB0	100.0	25.0	40.0	1
	EfficientNetV2M	100.0	33.3	50.0	1
	ConvNeXtBase	100.0	33.3	50.0	1
	InceptionV3	100.0	100.0	100.0	1
Taif Highland	ResNet50	80.8	95.5	87.5	26
	EfficientNetB0	84.6	84.6	84.6	26
	EfficientNetV2M	88.5	85.2	86.8	26
	ConvNeXtBase	88.5	95.8	92.0	26
	InceptionV3	73.1	86.4	79.2	26
Tuhama Coast	ResNet50	50.0	33.3	40.0	4
	EfficientNetB0	75.0	60.0	66.7	4
	EfficientNetV2M	100.0	100.0	100.0	4
	ConvNeXtBase	100.0	100.0	100.0	4
	InceptionV3	50.0	50.0	50.0	4

Architectural Style	Backbone	Producer Accuracy (%)	User Accuracy (%)	F1-Score (%)	Test Images (n)
Tuhama Foothills	ResNet50	100.0	37.5	54.5	3
	EfficientNetB0	100.0	37.5	54.5	3
	EfficientNetV2M	100.0	37.5	54.5	3
	ConvNeXtBase	100.0	100.0	100.0	3
	InceptionV3	100.0	50.0	66.7	3

References

1. Niu, X.; Wang, X.; Bi, X. Mapping the Nexus of Sustainability and Cultural Heritage: A Systematic Review and Empirical Analysis. *npj Herit. Sci.* **2025**, *13*, 471. [CrossRef]
2. HRH the Crown Prince Launches the Saudi Architecture Characters Map to Celebrate the Kingdom's Architectural Heritage. Available online: https://spa.gov.sa/ (accessed on 22 July 2026).
3. Aldossary, M.J.; Alqahtany, A.M.; Alshammari, M.S. Cultural Heritage as a Catalyst for Sustainable Urban Regeneration: The Case of Tarout Island, Saudi Arabia. *Sustainability* **2025**, *17*, 4431. [CrossRef]
4. Alnaim, M.M.; Alnaim, M.A. Integrating Vernacular Architecture into Contemporary Urban Regeneration: Heritage, Identity and Modernization in Saudi Cities. *Land* **2026**, *15*, 494. [CrossRef]
5. Socio-Cultural Sustainability in Contemporary Saudi Architecture: Blending Tradition and Modernity to Shape the Saudi Identity. Available online: https://www.mdpi.com/2075-5309/15/24/4531 (accessed on 28 July 2026).
6. Ottoni, A.L.C.; Ottoni, L.T.C. A Deep Learning Approach for Cultural Heritage Building Classification Using Transfer Learning and Data Augmentation. *J. Cult. Herit.* **2025**, *74*, 214–224. [CrossRef]
7. Shehata, A.M.; Alaboud, N.S. Identification of Features of Architectural Heritage Using Deep Learning Techniques. *Civ. Eng. Archit.* **2025**, *13*, 3280–3298. [CrossRef]
8. Llamas, J.; M. Lerones, P.; Medina, R.; Zalama, E.; Gómez-García-Bermejo, J. Classification of Architectural Heritage Images Using Deep Learning Techniques. *Appl. Sci.* **2017**, *7*, 992. [CrossRef]
9. Abed, M.H.; Al-Asfoor, M.; Hussain, Z. Architectural Heritage Images Classification Using Deep Learning with CNN. In Proceedings of the 2nd International Workshop on Visual Pattern Extraction and Recognition for Cultural Heritage Understanding, Co-Located with the 16th Italian Research Conference on Digital Libraries (IRCDL 2020), Bari, Italy, 29 January 2020; Volume 2602, pp. 1–12.
10. Cosovic, M.; Jankovic Babic, R. CNN Classification of the Cultural Heritage Images. In Proceedings of the 2020 19th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, 18–20 March 2020.
11. Siountri, K.; Anagnostopoulos, C.-N. The Classification of Cultural Heritage Buildings in Athens Using Deep Learning Techniques. *Heritage* **2023**, *6*, 3673–3705. [CrossRef]
12. Zou, H.; Ge, J.; Liu, R.; He, L. Feature Recognition of Regional Architecture Forms Based on Machine Learning: A Case Study of Architecture Heritage in Hubei Province, China. *Sustainability* **2023**, *15*, 3504. [CrossRef]
13. Tan, G.; Zhu, J.; Chen, Z. Deep Learning Based Identification and Interpretability Research of Traditional Village Heritage Value Elements: A Case Study in Hubei Province. *Herit. Sci.* **2024**, *12*, 200. [CrossRef]
14. Ma, K.; Wang, B.; Li, Y.; Zhang, J. Image Retrieval for Local Architectural Heritage Recommendation Based on Deep Hashing. *Buildings* **2022**, *12*, 809. [CrossRef]
15. Ottoni, A.L.C.; Ottoni, L.T.C. ImageOP: The Image Dataset with Religious Buildings in the World Heritage Town of Ouro Preto for Deep Learning Classification. *Heritage* **2024**, *7*, 6499–6525. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.