COMPDYN 2025
10th ECCOMAS Thematic Conference on
Computational Methods in Structural Dynamics and Earthquake Engineering
M. Papadrakakis, M. Fragiadakis (eds.)
Rhodes Island, Greece, 15-–18 June 2025

# AN INTEGRATED METHODOLOGY COMBINING CNN WITH AN INTERVIEW-BASED APPROACH TO ENHANCE RAPID BUILDING INVENTORY AT THE MUNICIPALITY SCALE

**Onur Ulku[1], Maria Polese[1], Jelena Pejovic[2], and Nina Serdar[2]**

[1]Department of Structures for Engineering and Architecture, University of Naples "Federico II", Italy
via Claudio 21, 80125 Napoli
e-mail: {onur.ulku,maria.polese}@unina.it

[2] Faculty of Civil Engineering, University of Montenegro
Džordža Vašingtona bb., 81000 Podgorica
e-mail: {jelenapej,ninas}@ucg.ac.me

**Keywords:** Convolutional Neural Network, Image Recognition, Construction Material, Earthquake Engineering, Interview-based Survey.

**Abstract.** *Effective regional risk assessment depends on accurate evaluations of hazard, exposure, and vulnerability. Among these, exposure modelling is as crucial as robust hazard assessment and a reliable understanding of the vulnerability of exposed elements. Traditionally, exposure models have relied on census data, building-by-building surveys, and post-earthquake damage assessments. More recently, interview-based surveys have supplemented these sources by providing insights into building typologies. However, such methods face challenges including limited data access, high costs, time-consuming processes, and dependency on expert judgment, which may affect reliability. Advances in computational power and image processing have introduced new methods for exposure modelling. Convolutional Neural Networks (CNNs), widely used in machine learning, can efficiently classify building typologies based on visual patterns from images, offering geo-referenced insights that enhance risk assessments. However, these image-based methods are constrained by their reliance on visible features, mainly façades and roofs, which may not capture the full structural diversity due to regional construction practices. Tools like BRAILS [1] have achieved notable results using CNNs with street and satellite imagery, but they remain limited by their data sources. This paper proposes a novel, integrated methodology that combines CNN-based image processing with data from field surveys, satellite imagery, and interview-based surveys. By incorporating expert insights from interviews, the method enhances CNN performance, providing a more reliable and explainable machine learning model for exposure modelling. This fusion captures context-specific features that are not visually apparent in images, improving classification accuracy and better reflecting regional construction practices. A preliminary application of this methodology in Bar Municipality, Montenegro, demonstrates its potential.*

## 1 INTRODUCTION

The task of building inventory creation is critical, as the evaluation of reliable exposure is an essential step for effective risk assessment. Identifying and characterizing building characteristics significantly impacts the correct attribution of vulnerability class to buildings, which is as important as generating a high-resolution geo-distributed building inventory. Generating reliable exposure models is, therefore, a crucial aspect of risk assessment. Various sources have been utilized so far to gather building inventory data to support seismic risk assessments, including cadastral datasets [2, 3], census data on population and housing [4, 5, 6, 7], post-earthquake vulnerability assessments [8, 9, 10, 11], on-the-ground surveys [12, 13], and interview-based surveys [14, 15, 16]. These methods each contribute valuable information but come with significant limitations. Census and cadastral data often lack sufficient detail to accurately determine building typologies. Furthermore, post-earthquake and on-the-ground surveys can be time-consuming and costly, requiring human resources with technical expertise, which can introduce subjectivity and bias in assigning the correct typology to buildings. Interview-based surveys, though useful for gaining insights from local knowledge and expert opinions, may not be suitable for assembling geo-distributed building inventory unless referring to typologies in suitably defined urban sectors [16]. Given these limitations, traditional data collection methods alone are insufficient for creating the detailed, high-resolution exposure models necessary for reliable risk assessments. As such, these traditional methods should be considered supplementary rather than definitive sources of information.

Recent advances in computational capabilities and machine learning, particularly in image processing, offer promising solutions to exposure modelling. Image-based techniques now enable the identification of building features and categorization into typologies based on visual patterns. This approach provides a rapid and efficient method for gathering geo-referenced information on building stock. By integrating street view imagery, such as Google Street View (GSV), and satellite data with artificial intelligence, more reliable and spatially detailed exposure models can be developed. GSV imagery, combined with machine learning techniques, has gained significant attention due to its wide availability and extensive coverage. Numerous studies have employed this combination to estimate building age [17], screen for soft-story buildings [18], detect building façade elements [19], classify roof types [20], and identify key building attributes such as construction material, use, and condition [21]. Additionally, some studies focus on predicting and classifying three essential pieces of information related to the vulnerability of building typologies, namely construction material, number of stories, and construction epoch [22]. These studies highlight the potential for automating workflows with machine learning, achieving high accuracy and scalability in predicting building characteristics. However, image-based approaches are inherently limited to visible features such as façades and roofs, which may not fully represent regional construction practices.

To address these limitations, a hybrid model combining image-based processing with interview -based surveys can be developed. Machine learning algorithms, particularly Convolutional Neural Networks (CNNs), are well-suited for image classification tasks, yielding impressive results in predicting building typologies based on visual patterns. This hybrid model can leverage visual patterns associated with similar typologies while also incorporating expert opinions on characteristics that may not be discernible from satellite or street-view images. Interview-based techniques are especially useful for capturing fundamental building information, such as the construction year and construction practices, which are relevant to characterize a building's vulnerability and assist in identifying the typology of geo-referenced buildings. In this paper,

we propose a CNN-based training process to predict the construction material of the lateral load-resisting system, using visual similarities from images, along with transfer learning [23] from the ImageNet dataset [24] with feature extraction and supplemental data, to enhance both the accuracy and explainability, as well as the interpretability of the algorithm's predictions. The use of transfer learning allows for the improvement of efficiency and performance in CNNs by utilizing pre-trained models, which is especially beneficial when limited training data is available. Additionally, it is essential to ensure that the model's predictions are clear and easy to explain. This is important for decision-makers, as understanding how predictions are made is necessary to effectively use the results in risk assessments. Additionally, a preliminary application of this approach has been conducted in the municipality of Bar, Montenegro, where a database consisting of both image and supplemental data was gathered.

## 2 EXPLORA-T SURVEY AND IMAGE DATA ACQUISITION

The proposed methodology aims to develop a representative model that predicts the building's construction material of the lateral load-resisting system by classifying it as reinforced concrete (RC), unreinforced masonry (URM), or confined masonry (CFM); those material types are selected given their prevalence in Montenegro, as documented in [25] One of the crucial steps in this methodology is to analyze the statistical distributions of relevant building characteristics associated with different typologies in a specific area. This step ensures that all typologies are properly represented by their respective building characteristics and helps determine the data collection process.

Since on-the-ground data collection is a time-consuming and costly procedure, it must be optimized to develop a model that reflects general construction practices while ensuring there is enough data to train the CNN model effectively. To achieve this, supplemental data sources, such as cadastral data, census data, and interview-based surveys, can be utilized to understand the construction practices and statistical distributions of building characteristics. For this purpose, the EXPLORA-T survey, an interview-based survey, was conducted to determine the building characteristics and their statistical distributions in Bar Municipality, Montenegro.

The EXPLORA-T survey, like the CARTIS form [16], aims to identify building typologies and structural characteristics with their distribution within specific urban areas called "Sectors." These sectors are defined as homogeneous zones where buildings share similar structural characteristics and construction periods. The delineation of sectors is based on historical, bibliographic, and documentary research, which provides insights into construction practice. Cartographic and cadastral data, aerial imagery, and satellite data also help refine sector boundaries.

### 2.1 General Introduction and Description of Bar Municipality

Bar Municipality, characterized by its diverse urban landscape and strategic importance along Montenegro's Adriatic coast, was selected for the pilot study due to its varied building types influenced by historical, coastal, and rural contexts. The area spans 598 square kilometres, hosting a population of over 40,000, and serves as a critical hub for tourism, trade, and cultural heritage. The diversity in building constructions, ranging from modern reinforced concrete structures in the coastal areas to traditional masonry in rural regions, presents unique challenges for exposure modelling. The Spatial Urban Plan of Bar outlines land use and characteristics of the built environment and serves as an input for effectively dividing Bar municipality within the sector necessary for EXPLORA-T survey realization.

## 2.2 EXPLORA-T Survey of Bar

The EXPLORA-T survey in Bar was conducted to identify and document building typologies across the municipality. This survey was particularly focused on assessing residential buildings. Municipality was divided into three sectors to capture the variability across coastal, historic, and rural areas (see Figure 1). Each sector presented distinct building characteristics that were essential for understanding the regional construction practices:

- Coastal Area (Sector 01): Dominated by modern constructions related to tourism and urban development. Buildings are primarily reinforced concrete with multi-story designs.
- Historic Area (Sector 02): Features a mix of medieval masonry and contemporary structures, reflecting the historical evolution of the area.
- Rural Area (Sector 03): Characterized by traditional masonry techniques, with buildings often constructed for functionality over form, emphasizing low-rise, adaptable designs suitable for agricultural lifestyles.

Two technicians, with expertise in local construction practices and structural engineering, provided the data on residential building typologies for each sector. The gathered data ensured that the classifications reflected actual building practices rather than theoretical models.

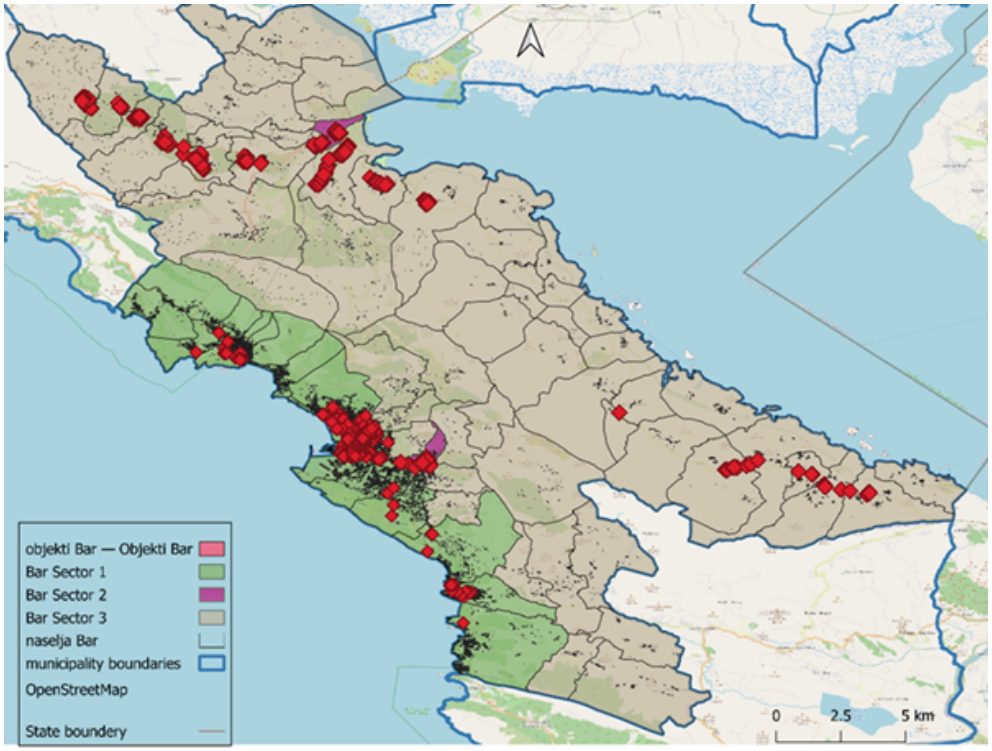


Figure 1: Sectors in Bar municipality Bar and location gathered image data.

The EXPLORA-T survey conducted in Bar Municipality identified four main building typologies, highlighting the diversity of construction practices within the area. These typologies include three masonry types—confined masonry (M1), unconfined masonry with regular stone (M2), and unconfined masonry with irregular stone units (M3)—and one reinforced concrete type (RC1).

Both interviewed experts reported similar findings regarding the percentage of isolated / adjoining buildings within each typology, storey height, and the presence of structural irregularities. The main differences between the interviewed experts were related to the percentage distribution of typologies within the sectors and the presence of certain typologies in specific sectors. Survey findings in terms of the distribution of building typologies (M1, M2, M3, RC1) are given in Figure 2. For instance, in Sector S01, Technician 1 reported a dominant presence of M1 (89%), whereas Technician 2 observed a more varied distribution, noting substantial amounts of RC1 (60%). The distribution of typologies varied across sectors, with the technicians agreeing more consistently on the prevalence of M1 and M2 in Sector S02 but differing significantly in their assessments of Sectors S01 and S03. These variations were attributed to different focuses in their analysis or potential biases based on their individual experiences and perspectives.

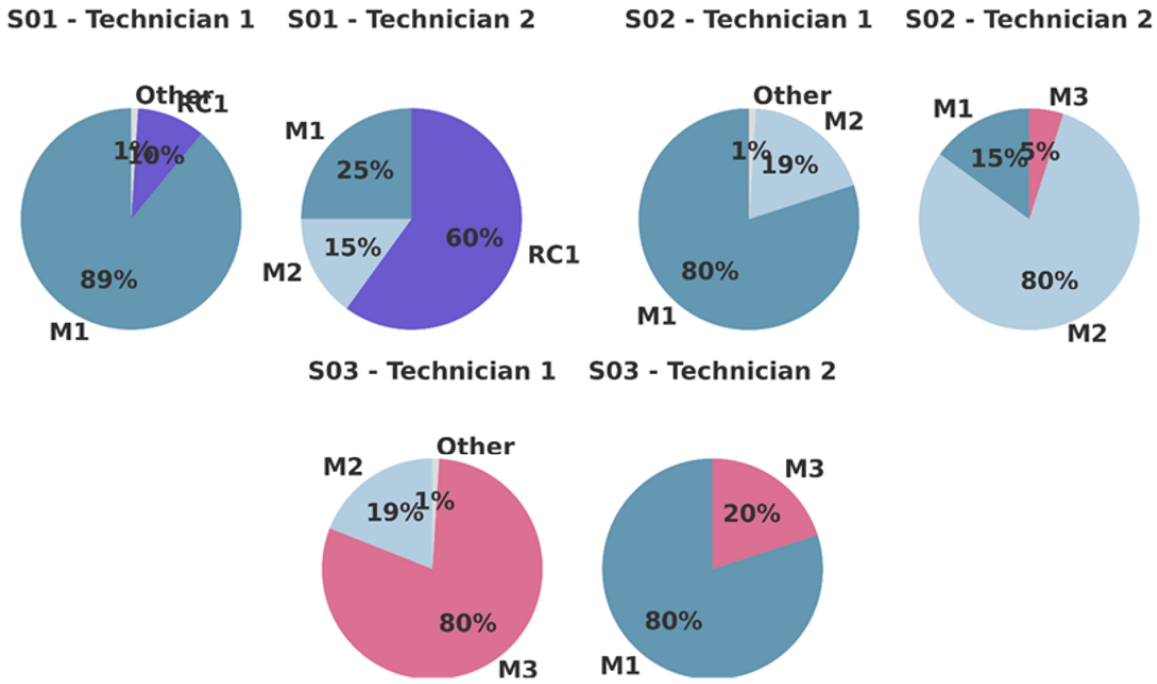


Figure 2: Survey observations of Technician 1 and Technician 2 on the distribution of building typologies within the Sectors.

The evaluation of the EXPLORA-T survey conducted in Bar Municipality highlights the importance of rigorous field validation when interpreting survey data, especially in cases where expert assessments may differ. The insights derived from the field visits and the photographic documentation process reveal that Technician 1 should be considered more reliable for Sectors S01 and S02, particularly for RC1 and M1 typologies. For Sector S03, Technician 2's insights regarding M1 appear to be more accurate and should be given precedence in further analysis.

### 2.3 Image Database for Bar

Parallel to the EXPLORA-T survey, an extensive image database was compiled to support the machine-learning component of the study. Over 1214 images were systematically collected, ensuring coverage of each identified building typology across all sectors. This extensive collection aimed to ensure that each identified building typology was adequately represented, with

at least 50 buildings per typology, with 3 different pictures taken from various angles, meaning 150 pictures for each building typology, should be collected in each sector. This choice was based on a preliminary assessment of the building database of Alvalade, Lisbon [26], where the training dataset was systematically reduced, and the model's performance was monitored. The decision to use 50 buildings per typology within a sector was determined as the minimum dataset size before a significant decline in model performance was observed. The overview of the locations where the photographs were taken is shown in Figure 1.

In the image database for Bar Municipality, each photograph is enriched with a comprehensive set of attributes essential for a detailed analysis. Geographical details such as the country, municipality, and precise coordinates ensure each building is accurately mapped. The database records various building characteristics, including the type of lateral load-resisting system present—ranging from none to complex systems like dual frame-wall constructions—and the materials used, such as reinforced concrete or various forms of masonry, with specific notes on masonry units when applicable. The construction year of each building is categorized by significant periods, aligning with developments in local building regulations (i.e. before 1964, between 1965 and 1981, after 1981). Physical attributes of the buildings, such as the presence of basements, floor area, and the height of individual floors, are also collected.

Each entry in the database is meticulously documented, including the source of the image and any supplementary notes that highlight unique features or conditions of the buildings are added to support further image data processing.

## 3 METHODOLOGY

The methodology consists of training a CNN model to predict the construction material of the lateral load-resisting system using given building characteristics. This hybrid approach overcomes the limitations of image-only recognition by incorporating contextual features that are not visually observable, such as the construction year or the material type, ensuring more accurate building classification. It enhances the reliability and interpretability of exposure models, accounting for diverse construction practices across regions. For this purpose, the CNN architecture is augmented with additional layers to integrate supplemental data into the model.

The CNN architecture (illustrated in Figure 3) is designed with two primary components. The first component focuses on feature extraction from images using existing CNN architectures, such as ResNet50 [27], VGGNet [28], DenseNet [29], and LeNet [30]. These models leverage transfer learning from large existing datasets like ImageNet [24] to identify and capture visual patterns in images, ensuring robust feature representation for the task. The second component integrates additional layers from supplemental data, such as construction year and the number of storeys, into the CNN model. Supplemental data was collected alongside the image database during the on-the-ground survey, ensuring that relevant building characteristics were gathered for each image. Integration of additional layers is achieved by concatenating the final layers of the deep neural network designed for supplemental data with the existing CNN architecture.

In the first phase, data preparation involves examining the dataset and addressing any relationships between features. If the dataset is imbalanced, data augmentation techniques should be applied to improve the model's accuracy. Image augmentation methods include random rotation, axis distortion, zooming, horizontal flipping, and random colour transformations to enhance prediction accuracy and prevent overfitting. Additionally, the supplemental data must be preprocessed to be effectively used for model training. Preprocessing methods for numerical and categorical data include feature scaling techniques, such as standardization, normalization,

and non-linear transformations (e.g., Power Transform and Quantile Transform). In particular, the methodology uses two types of data retrieved from the image database of Bar presented in Section 2.3, namely the construction year and the number of storeys; these features are both categorical data. For categorical features, encoding methods can be used to convert them into binary vectors.

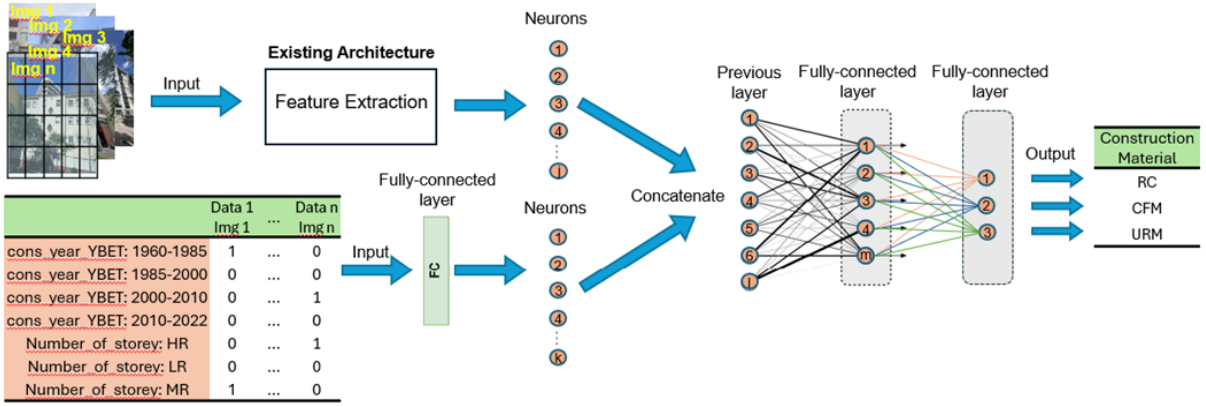


Figure 3: The proposed CNN architecture (n: total number of data; j, k, l, and m: Dense Unit).

Next, the entire dataset is split into two subsets: one for training and one for testing. During this process, it is essential to ensure that all feature combinations are stratified, meaning they are proportionally represented in both subsets. This approach helps maintain a balanced distribution, ensuring that all classes are adequately represented in both the trained model and the testing phase, leading to a more robust evaluation of the model's performance. The training subset is used for hyperparameter tuning and model development, while the testing subset is reserved for evaluating the model's final performance. Cross-validation should be applied during training to ensure robustness and generalizability. Cross-validation involves dividing the training dataset into multiple subsets and iteratively training the model on one subset while testing it on the others. This process is repeated several times to ensure each data point is used for both training and testing, and the results are averaged to provide a reliable estimate of model performance.

Finally, hyperparameter tuning and fine-tuning are essential to enhance model performance, prevent overfitting, and adapt a pre-trained model to a new task. Key hyperparameters that can be adjusted include the optimizer function, activation function, number of epochs, batch size, learning rate, and number of dense units in the data layer [31, 32]. Furthermore, the trainability of the top layers should be allowed by modifying the weights of an existing model for feature extraction [33, 34]. This process is critical for the model to learn new tasks from the given dataset. Optimizers such as Gradient Descent, Stochastic Gradient Descent, and Adaptive Gradient Descent are used to minimize loss and optimize model weights. Activation functions, like the sigmoid function, rectified linear unit (ReLU), and Gaussian function, help the model learn complex relationships between input features. Further details on some hyperparameters that can be tuned are provided in Table 1.

## 4 APPLICATION OF THE METHODOLOGY

The proposed method has been implemented in the selected pilot area, Bar Municipality, Montenegro. To identify building characteristics and typologies, both the EXPLORA-T sur-

| Hyperparameters | Descriptions |
|---|---|
| Number of Epoch | how many times the learning algorithm will pass through the entire training dataset during the training process |
| Batch Size | Number of training samples |
| Learning Rate | Step size at each epoch to minimize the loss function |
| Dense Unit | The number of neurons in the tabular layer |
| Trainable | Feature extraction freezes the weights of the top layers of the feature extraction layers |

Table 1: Some of the hyperparameters and their roles in model tuning.

vey and an onsite survey were conducted, as detailed in the EXPLORA-T Survey and Image Data Acquisition section. A CNN architecture integrated with tabular data was trained using building images alongside their respective attributes. The tabular data included the construction year, categorized into six groups: Unknown (U), Before 1964 (NC), Between 1964 and 1981 (LC), Between 1982 and 2001 (MC), Between 2002 and 2011 (MHC), and After 2011 (HC), as well as the number of stories, classified into low-rise (LR) for 1–3 storey buildings, mid-rise (MR) for 4–8 storey buildings, and high-rise (HR) for 9–19 storey buildings. The developed CNN model is designed to classify the construction material of the lateral load-resisting system, distinguishing between reinforced concrete, confined masonry, and unreinforced masonry based on the building's image and additional characteristics, i.e. construction age and number of stories; the latter data is available in the purposedly built database of tabular data acquired along with images as described in section 2.3.

The conditional probabilities of observed characteristics for each construction material are presented below to illustrate the relationship between inputs and outputs. An initial analysis of the data reveals several key insights.

As shown in Figure 4, a large part of RC and CFM buildings were built after 1982, whereas URM buildings were predominantly built before 1964 or between 1964 and 1981; among the collected data there are no URM buildings constructed after 2001. Additionally, the distributions reveal distinct construction trends and time periods associated with each material type of lateral load-resisting system. RC buildings are predominantly linked to more recent construction years, while URM buildings correspond to older construction periods.

Figure 5 illustrates the distribution of representative buildings based on the number of storeys and the material of the lateral load-resisting system. The "unknown" category (storey number 0) represents cases where specific building information is unavailable. Among buildings with a CFM lateral load-resisting system, the majority have 2 or 3 storeys, with occurrence probabilities of approximately 48% and 36%, respectively. The minimum number of storeys in this category is 1, with over 10% of buildings having only a single storey, while the maximum is 3. Additionally, a small proportion of buildings in this category falls under the "unknown" classification (3.4%).

RC buildings exhibit the greatest variation in the number of storeys, ranging from a minimum of 4 to a maximum of 11. The most frequent numbers of storeys are 10 (16%), 11 (16%), and 6 (20%). Additionally, a small portion of RC buildings (3.7%) falls under the "unknown" category.

In contrast, URM buildings predominantly have 2 or 1 storeys, with occurrence probabilities

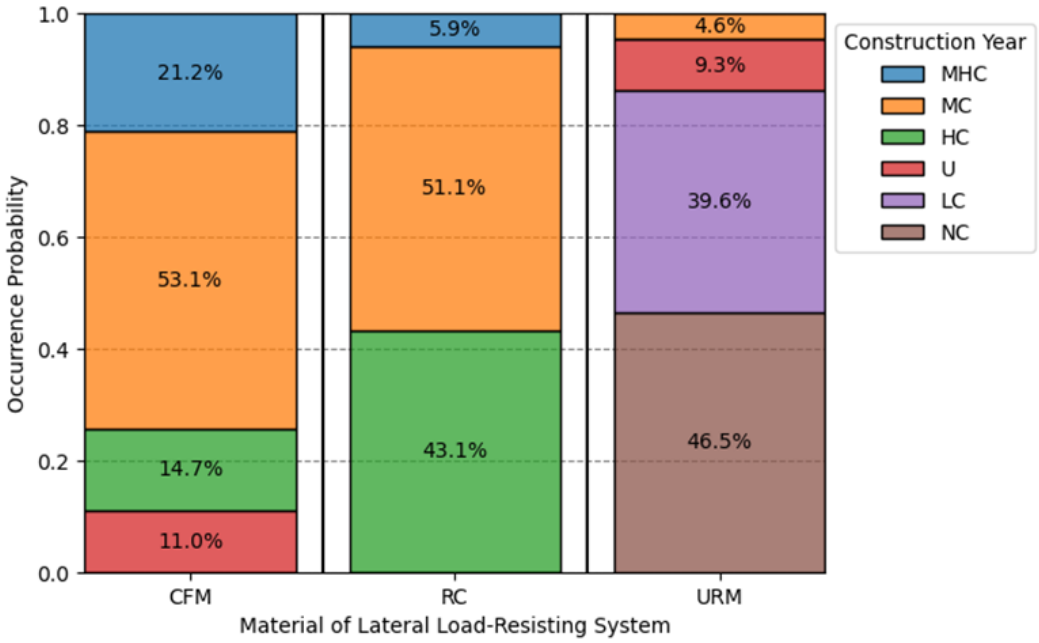


Figure 4: The distribution of the construction year conditioned on the material of the lateral load-resisting system.

of approximately 46% and 38%, respectively. The minimum number of storeys in this category is 1, while the maximum is 4. Only 1% of URM buildings exceed 3 storeys, and 2.6% are classified as "unknown."

Overall, these patterns indicate that RC buildings exhibit the highest variability in the number of storeys, whereas CFM and URM buildings tend to have lower and more consistent distributions.

However, these statistics are not considered fully reliable. This method primarily aims to represent general construction practices rather than produce statistically representative results. Therefore, the collected images and corresponding information are intended to explore the relationships between inputs and outputs rather than to reflect accurate distributions. Since the construction year and number of storey are categorical features, they were pre-processed using the one-hot encoding method before training the model.

The dataset was split into two subsets, with 80% allocated for training and 20% for testing. The testing subset was reserved for the final evaluation of the model's performance, while the training subset was used for hyperparameter tuning and model development. Additionally, Stratified K-fold cross-validation was applied to the training dataset to enhance the model's robustness and generalizability. This approach ensured a more comprehensive evaluation, minimizing the risks of overfitting and underfitting.

Data augmentation is a commonly used and effective technique for improving prediction performance. Increasing the dataset size typically enhances accuracy and reduces the risk of overfitting. Considering these advantages and the limited data available, data augmentation was implemented before training. The applied techniques included brightness adjustments, random rotations, vertical and horizontal shifting, contrast adjustment, and horizontal flipping.

As described in the Methodology section, the proposed approach consists of two branches: one for feature extraction from images and another for incorporating additional information from supplemental data. For the feature extraction process, well-established CNN architectures

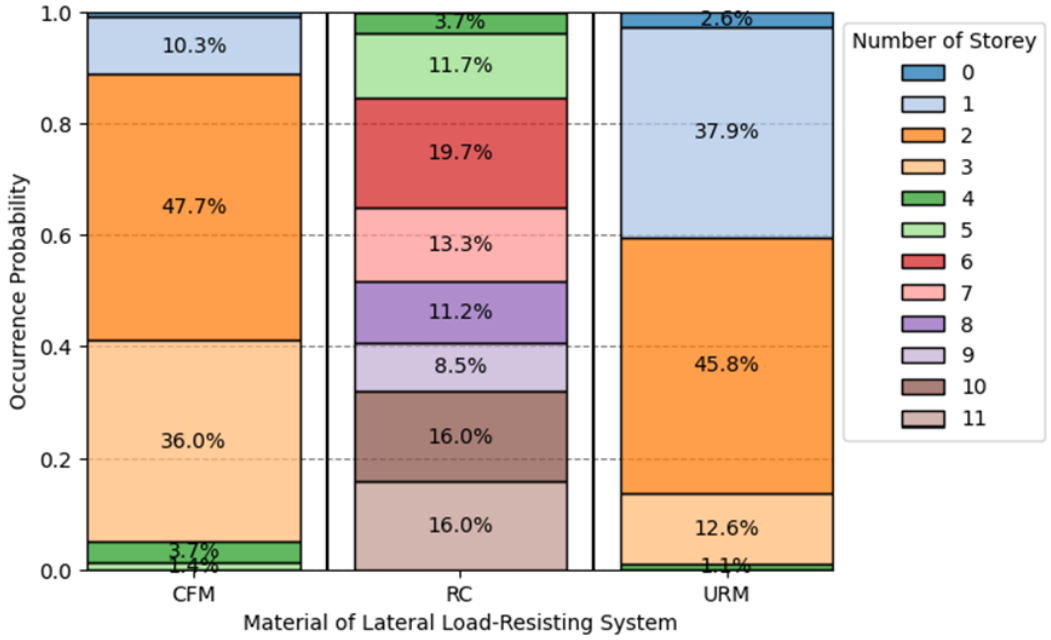


Figure 5: The distribution of the number of storeys conditioned on the material of the lateral load-resisting system.

that have demonstrated strong performance on the ImageNet dataset were employed, including ResNet50V2 [27], Xception [35], and EfficientNetV2 [36]

The proposed method was tuned by optimizing key hyperparameters, including the optimizer, activation function, number of epochs, batch size, learning rate, and dense unit, which determines the number of neurons in the supplemental data branch. A detailed explanation of other hyperparameters is provided in Table 1.

Approximately 1,500 models were trained using random search to identify the optimal set of hyperparameters. To prevent overfitting and encourage the learning of more generalized features, a dropout rate of 0.3 was applied, randomly deactivating 30% of neurons after each fully connected layer. After hyperparameter tuning, the final set of selected hyperparameters was determined based on the model that achieved the highest average F1 score across cross-validation sets.

All models were trained on a system equipped with a 13th-generation Intel Core i9 processor, 32 GB of RAM, and an NVIDIA GeForce RTX 4060 GPU. A summary of the selected hyperparameters and the top five F1 scores, along with their corresponding hyperparameters, are presented in Table 2 and Table 3, respectively.

Moreover, feature extraction from images was performed using transfer learning from the ImageNet dataset, providing the pre-trained model with a foundational understanding of visual patterns. However, since the target task in this project differs from ImageNet's, the model must adapt to the new dataset. To achieve this, fine-tuning was applied to the pre-trained model (base model) by unfreezing the top few layers and using a lower learning rate, allowing it to learn task-specific patterns, update weights, and refine trainable layers accordingly.

The optimal number of layers to unfreeze depends on factors such as dataset size, similarity to ImageNet, and model complexity. Given our relatively small dataset, excessive unfreezing could lead to overfitting. An automated selection process was implemented to determine the optimal number of layers, iterating through different numbers of trainable layers. As a result,

| Hyperparameters | Selected Hyperparameters |
|---|---|
| CNN Architecture | EfficientNetV2S |
| Optimizer | Adaptive Moment Estimation (Adam) |
| Activation Function | Rectified Linear Unit (ReLu) |
| Number of Epoch | 25 |
| Batch Size | 64 |
| Learning Rate | 0.005 |
| Dense Unit | 512 |

Table 2: Selected hyperparameters.

| CNN Architecture | Optim. | Activ. Func. | Number Epoch | Batch Size | Learning Rate | Dense Unit | F1 Score |
|---|---|---|---|---|---|---|---|
| EfficientNetV2S | adam | relu | 25 | 64 | 0.005 | 512 | 0.91 |
| ResNet50v2 | adam | relu | 25 | 32 | 0.005 | 512 | 0.87 |
| Xception | adam | relu | 10 | 16 | 0.005 | 512 | 0.85 |
| EfficientNetV2S | adam | relu | 50 | 16 | 0.001 | 256 | 0.84 |
| EfficientNetV2S | adam | tanh | 50 | 64 | 0.005 | 1024 | 0.82 |

Table 3: Five best models' hyperparameters and F1 scores.

unfreezing and training the top 10 layers yielded the best performance, improving the F1 score from 0.91 to 0.97 on the validation dataset. This indicates that fine-tuning effectively enhanced the model's performance and generalization for the target task.

The accuracy and loss over epochs for the models without fine-tuning and with fine-tuning are presented in Figures 6 and 7, respectively. The accuracy plot in Figure 6 (without fine-tuning) indicates that the model learns well during training, achieving high training accuracy, while the validation accuracy closely follows. This suggests that the model generalizes well, with no severe overfitting issues, despite minor fluctuations in the validation accuracy. Importantly, there is no significant drop in validation accuracy as training progresses. The loss plot in Figure 6 further supports this observation, showing effective learning with an initial steep drop in loss, confirming that the model optimizes efficiently.

In contrast, the accuracy and loss plots in Figure 7 (with fine-tuning) demonstrate a significant improvement. The fine-tuned model achieves nearly perfect accuracy, with both training and validation accuracy remaining close to 1. The corresponding loss plot shows minimal loss values for both training and validation, with only a slight gap between them, indicating that the model generalizes exceptionally well. The stability of the validation loss further confirms that fine-tuning has enhanced the model's performance without introducing overfitting.

## 5 RESULTS

As a result, the base model was trained using the EfficientNetV2 architecture with transfer learning, and its hyperparameters are provided in Table 2. A dropout rate of 0.3 was applied after each fully connected layer, randomly deactivating neurons during training to prevent overfitting and enhance generalisation.

For further improvement, the base model was fine-tuned by unfreezing the last 10 layers of the EfficientNetV2 model and training with a learning rate of 0.0005, which is lower than that

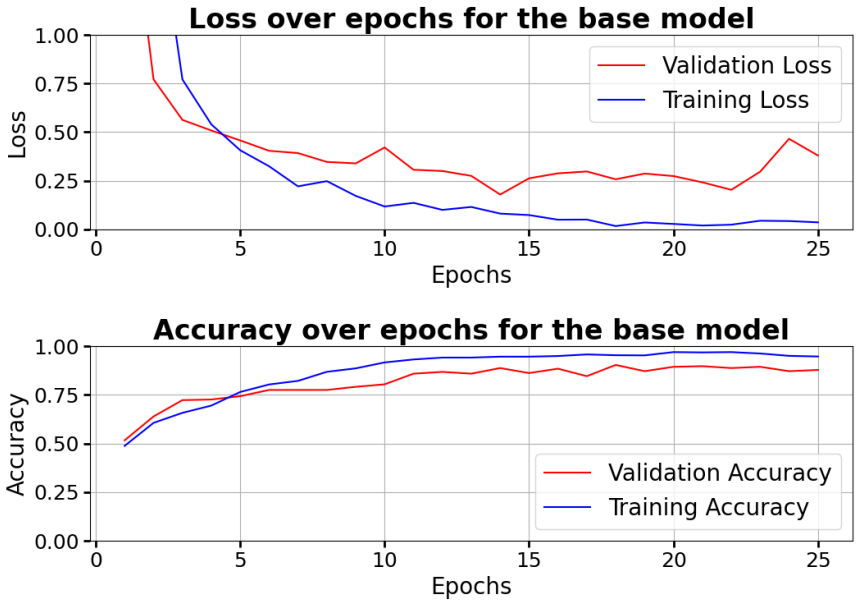


Figure 6: Loss and accuracy over epochs for the base model (without fine-tuning).

of the base model. This allowed the final model to better adapt to the target task. Additionally, early stopping, a regularization technique, was employed to prevent overfitting by monitoring the validation loss and stopping training if no improvement was observed for five consecutive epochs.

The final model was evaluated using a test dataset comprising 20% of the initial dataset. During the dataset split, we ensured that the test set adequately represented all class combinations based on the number of stories, construction year, and construction material, maintaining a balanced distribution for reliable model evaluation.

The performance of the final model was evaluated using precision, recall, and F1 score metrics, as summarized in Table 4. Among the three construction material classes, Reinforced Concrete (RC) achieved the highest performance, with an F1 score of 0.96. This indicates that the model is robust in identifying RC buildings, likely due to distinctive building characteristics that make this class easier to classify. Confined Masonry (CFM) and Unreinforced Masonry (URM) also have a high performance with an F1 score of 0.94. The weighted average F1 score of 0.95 reflects the overall balance of the model's performance across the three classes.

| | **Precision** | **Recall** | **F1 Score** |
|---|---|---|---|
| Confined Masonry | 0.93 | 0.96 | 0.94 |
| Reinforced Concrete | 0.93 | 1.00 | 0.96 |
| Unreinforced Masonry | 0.98 | 0.91 | 0.94 |
| Weighted Average | 0.95 | 0.95 | 0.95 |

Table 4: Performance metrics of the final model corresponding to the test dataset.

The confusion matrix in Figure 8 provides additional insights into the model's classification performance. Reinforced Concrete (RC) buildings were perfectly classified with no misclassifications. This aligns with the high precision and recall observed for RC in Table 4. Confined

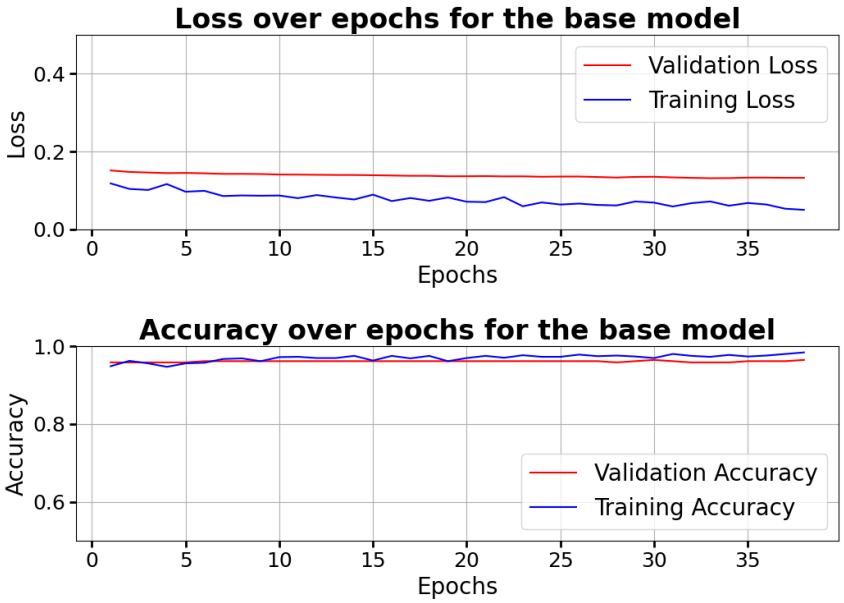


Figure 7: Loss and accuracy over epochs for the proposed model (with fine-tuning).

Masonry (CFM) buildings were correctly classified in 95.6% of cases, with small misclassification rates of 2.6% as RC and 1.8% as Unreinforced Masonry (URM). This indicates strong performance in distinguishing CFM from other categories.

However, URM buildings showed higher misclassification rates, with 9.1% misclassified as CFM. While the majority (90.9%) of URM buildings were correctly classified, the misclassification may stem from similar building characteristics or visual patterns present in the dataset.

The overall model accuracy, calculated at 94.6%, demonstrates a high level of generalizability in classifying building types, supported by 97.0% validation accuracy and 97.4% training accuracy. This performance indicates that the model effectively differentiates between Reinforced Concrete (RC), Confined Masonry (CFM), and Unreinforced Masonry (URM) buildings.

However, further improvements could focus on reducing misclassifications between CFM and URM, as well as occasional misclassification of CFM as RC. Potential enhancements include expanding the dataset with more diverse samples, incorporating advanced feature extraction techniques, or integrating additional relevant features. These refinements could help improve the model's robustness and further reduce classification errors.

## 6 EXPLAINABILITY

Misclassifications are an inherent challenge in machine learning models. However, despite the complexity of convolutional neural networks (CNNs) making their decision-making process less interpretable, several techniques exist to enhance explainability. These include Linear Proxy Models [37], automatic rule extraction [38, 39, 40], and saliency mapping methods such as Layer-wise Relevance Propagation (LRP) [41], DeepLIFT [42], Class Activation Mapping (CAM) [43], and Gradient-weighted Class Activation Mapping (Grad-CAM) [44].

By analyzing which regions of an image contribute to a model's classification at different layers, researchers can better understand how CNNs distinguish between Reinforced Concrete (RC), Confined Masonry (CFM), and Unreinforced Masonry (URM) buildings. CNNs base

| | | Predicted Class | | | |
|---|---|---|---|---|---|
| Actual Class | | Confined Masonry | Reinforced Concrete | Unreinforced Masonry | Total |
| | Confined Masonry | 109 (95.6%) | 3 (2.6%) | 2 (1.8%) | 114 |
| | Reinforced Concrete | 0 (0.0%) | 38 (100%) | 0 (0.0%) | 38 |
| | Unreinforced Masonry | 8 (9.1%) | 0 (0.0%) | 80 (90.9%) | 88 |
| | Total | 117 | 41 | 82 | Accuracy: 94.6% |

Figure 8: Confusion matrix.

their predictions on distinctive patterns such as shapes, textures, and structural features, and recognizing these cues can help improve both model accuracy and interpretability.

To enhance classification accuracy and minimize errors, CNNs offer interpretability techniques that highlight the key features influencing their predictions. Grad-CAM visually identifies the most relevant regions of an image for decision-making. For example, Grad-CAM can pinpoint which façade elements or textures contributed to classifying a building as RC, CFM, or URM. By analysing these visual explanations, researchers can better understand why the model struggles to distinguish between CFM and URM, which is a challenge reflected in Figure 8, where 9.1% of URM buildings were misclassified as CFM. These visualization techniques improve model transparency and can inform data refinement strategies, such as incorporating more representative samples or applying feature engineering to address overlapping characteristics. Despite inherent challenges, CNNs' ability to provide interpretable outputs demonstrates their potential for enhancing classification performance in future iterations.

Figures 9, 10, and 11 display Grad-CAM visualizations for both correct and incorrect classifications of CFM and URM, while only correct classifications are shown for RC buildings, as no misclassifications occurred in this category, respectively. These visualizations reveal certain challenges affecting classification accuracy. Some building images were partially obstructed by objects such as trees, fences, or surrounding structures, concealing key features necessary for accurate classification. Additionally, photographs taken from considerable distances limited the model's ability to capture crucial textures, shapes, and patterns associated with construction materials. These limitations significantly affected the model's performance in correctly identifying the lateral load-resisting system. Furthermore, similarities between construction materials, such as balconies in RC and CFM buildings, added to the classification complexity, contributing to misclassifications.

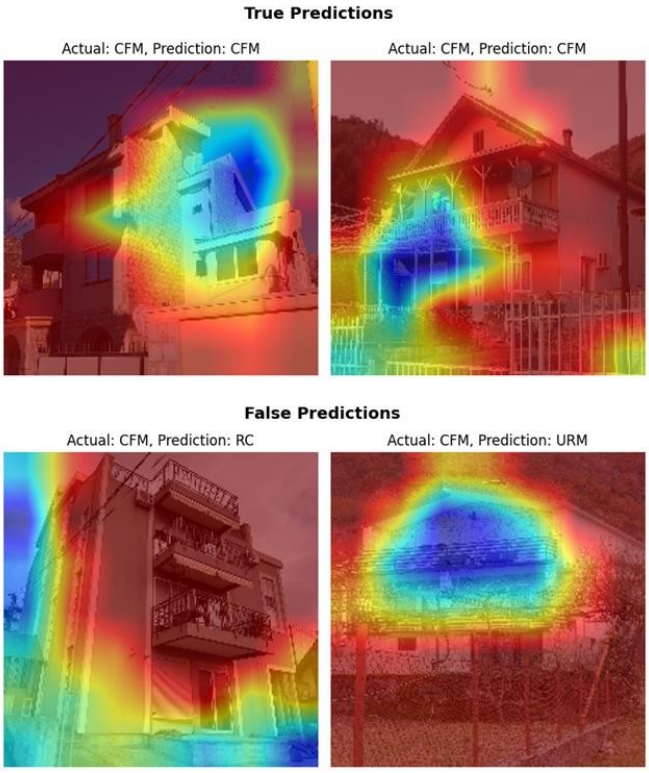


Figure 9: Examples of Grad-CAM maps representing true and false predictions for CFM.

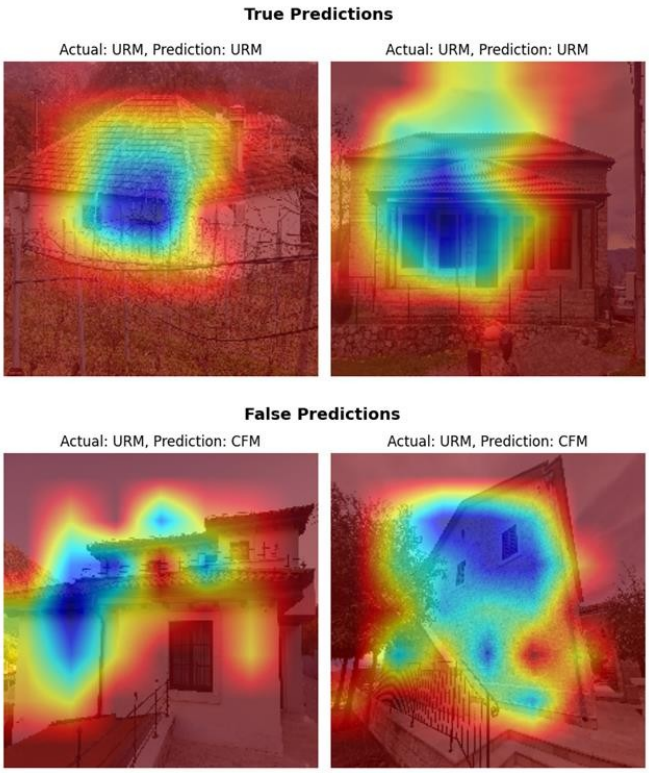


Figure 10: Examples of Grad-CAM maps, representing true and false predictions for URM.

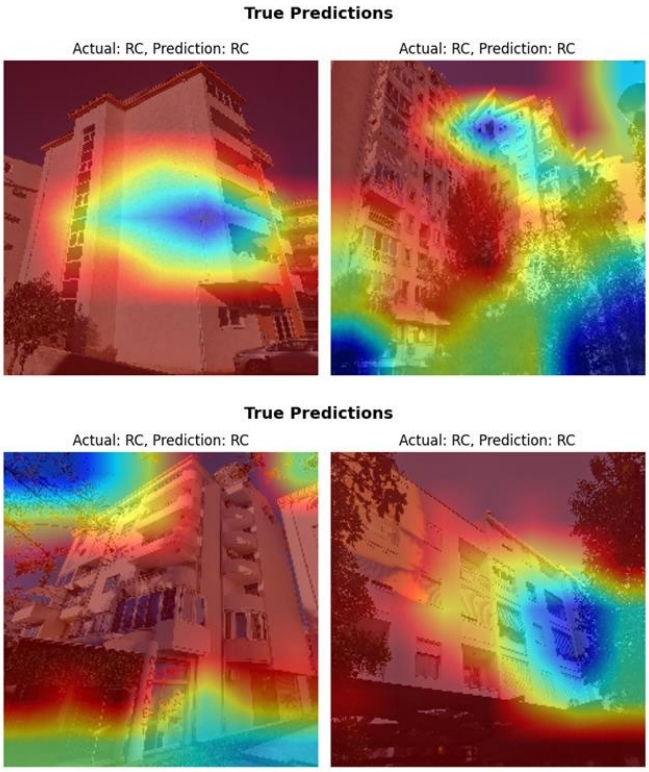


Figure 11: Examples of Grad-CAM maps, representing true predictions for RC.

## 7 CONCLUSION

The results indicate that convolutional neural networks (CNNs) integrated with supplemental data present a promising approach for the automated collection of exposures. While some studies, such as Gouveia et al. (2024), highlight the challenges of predicting construction materials, particularly due to façades not always providing clear identification or renovations altering the original materials, these limitations can be mitigated by leveraging reliable image datasets and supplemental information. Training the model on high-quality images alongside relevant supplemental data can identify common visual patterns and incorporate additional contextual information to improve predictions. Consequently, concerns about historic renovations altering material identification become less relevant, as similar renovation practices in historic sites would maintain a level of consistency in the dataset.

Furthermore, hyperparameter tuning and fine-tuning processes enable models to learn effectively even with relatively small datasets through transfer learning, making the model more adaptable to its intended task.

The model's performance can be further enhanced by incorporating additional supplemental data, which could introduce expert-driven insights into material classification. Additionally, instead of predicting a single most probable construction material, the model could estimate the probability distribution of different materials, thereby supporting probabilistic exposure models. These models would offer valuable insights into the variability and uncertainty of construction material classification.

As previously emphasized, explainability is critical for decision-makers to understand the reasoning behind AI-driven predictions and adjust the model based on expert knowledge. Instead of solely relying on image-based pattern recognition, the model could be enhanced by

predicting the presence of specific architectural features associated with construction materials. For instance, long cantilevers or wide openings in façades may indicate RC structures, while arched entrances may suggest URM buildings. Incorporating such explicit feature-based predictions would increase model interpretability and make the system more transparent and practical for real-world applications.

## 8 ACKNOWLEDGEMENT

This study was performed within the EXPLORA project, co-financed by Ministero degli Affari Esteri e della Cooperazione Internazionale in Italy and by the Ministarstvo nauke i tehnološkog razvoja in Montenegro.

## REFERENCES

[1] B. Cetiner, C. Wang, F. McKenna, S. Hornauer, J.Zinyan, C. Perez, Y. Guo, *NHERI-SimCenter/BRAILS: Version 3.1.2 (v3.1.2)*. 2024.

[2] S. Tyagunov, M. Pittore, M. Wieland, S. Parolai, D. Bindi, K. Fleming, J. Zschau, Uncertainty and sensitivity analyses in seismic risk assessments on the example of Cologne, Germany. *Nat Hazard*, **14(6)**, 1625–1640, 2014.

[3] A.B. Acevedo, C. Yepes, D. Gonzalez, V. Silva, M. Mora, G. Posada, M. Arcila, C. Rosales, Seismic risk assessment for the residential buildings of the major three cities in Colombia: bogotá. Medellín Cali. *Earthq Spectra*. 2019.

[4] H. Crowley, S. Ozcebe, H. Baker, R. Foulser-Piggott, R. Spence, D7. 2 State of the knowledge of building inventory data in Europe *NERA Deliverable,7, v3*. 2014.

[5] V. Silva, H. Crowley, R. Pinho, H. Varum, Seismic risk assessment for mainland Portugal. *Bull Earthq Eng*, **13(2)**, 429–457, 2014.

[6] F. Meroni, T. Squarcina, V. Pessina, M. Locati, M. Modica, R. Zoboli, A damage scenario for the 2012 Northern Italy earthquakes and estimation of the economic losses to residential buildings. *Int J Disaster Risk Sci*, **8(3)**, 326–341, 2017.

[7] F. Cacace, G. Zuccaro, D. De Gregorio, F.L. Perelli, Building Inventory at National scale by evaluation of seismic vulnerability classes distribution based on Census data analysis: BINC procedure. *Int J Disaster Risk Reduct*, **28**, 384–393, 2018.

[8] FEMA 154 – ATC-21, Rapid Visual Screening of Buildings for Potential Seismic Hazards: A Handbook, Second Edition. *Federal Emergency Management Agency*, Washington D.C., USA, 1988.

[9] FEMA 154, Rapid Visual Screening of Buildings for Potential Seismic Hazards: A Handbook, Second Edition. *Federal Emergency Management Agency*, Washington D.C., USA, 2002.

[10] H. Sucuoglu, U. Yazgan, A. Yakut, A screening procedure for seismic risk assessment in urban building stocks. *Earthquake Spectra*, **23(2)**, 441–458, 2007.

[11] A. Ilki, M. Comert, C. Demir, K. Orakcal, D. Ulugtekin, M. Tapan, N. Kumbasar, Performance Based Rapid Seismic Assessment Method (PERA) for Reinforced Concrete Frame Buildings, *Advances in Structural Engineering*, **17(3)**, 439–459, 2014.

[12] C. Del Gaudio, P. Ricci, G.M. Verderame, G. Manfredi, Development and urban-scale application of a simplified method for seismic fragility assessment of RC buildings, *Engineering Structures*, **91**, 40–57, 2015.

[13] M. Polese, M. Marcolini, M. Gaetani d'Aragona, E. Cosenza, Reconstruction policies: explicitating the link of decisions thresholds to safety level and costs for RC buildings, *Bulletin of Earthquake Engineering*, **15(2)**, 759–785, 2017.

[14] M. Dolce, G. Zuccaro, F. Papa, *Protocollo d'intervista*, Dipartimento della Protezione Civile, 2002.

[15] P. Guéguen, C. Michel, and L. LeCorre, A simplified approach for vulnerability assessment in moderate-to-low seismic hazard regions: application to Grenoble (France), *Bulletin of Earthquake Engineering*, **4(3)**, 467–490, 2007.

[16] G. Zuccaro, M. Dolce, F.L. Perelli, D. De Gregorio, E. Speranza, CARTIS: a method for the typological-structural characterization of Italian ordinary buildings in urban areas, *Frontiers in Built Environment*, **9**, 2023.

[17] Y. Li, Y. Chen, A. Rajabifard, K. Khoshelham, M. Aleksandrov, Estimating building age from Google Street View images using deep learning. S. Winter, A. Griffin, M. Sester eds. *10th International Conference on Geographic Information Science (GIScience 2018)*, Leibniz International Proceedings in Informatics (LIPIcs), vol. 114, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, August 28–31, 2018.

[18] Q. Yu, C. Wang, F. McKenna, S.X. Yu, E. Taciroglu, B. Cetiner, K.H. Law, Rapid visual screening of soft-story buildings from street view images using deep learning classification, *Earthquake Engineering and Engineering Vibration*, **19**, 827–838, 2020.

[19] G. Zhang, Y. Pan, and L. Zhang, Deep learning for detecting building façade elements from images considering prior knowledge, *Automation in Construction*, **133**, 104016, 2022.

[20] C. Wang, Q. Yu, K.H. Law, F. McKenna, S.X. Yu, E. Taciroglu, A. Zsarnóczay, W. Elhaddad, B. Cetiner, Machine learning-based regional scale intelligent modeling of building information for natural hazard risk management, *Automation in Construction*, **122**, 103474, 2021.

[21] C. Wang, S.E. Antos, J.G.G. Goldsmith, L.M. Triveno, Visual perception of building and household vulnerability from streets, 2022.

[22] F. Gouveia, V. Silva, J. Lopes, R.S. Moreira, J.M. Torres, M. Simas Guerreiro, Automated identification of building features with deep learning for risk analysis, *Discover Applied Sciences*, **6(9)**, 2024.

[23] A.S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, CNN features off-the-shelf: an astounding baseline for recognition. *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*, IEEE, 2014.

[24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009.

[25] J. Pejovic, N. Serdar, I. Drobnjak, M. Jevric, S. Slavica, S. Vuckovic, M. Polese, F. Aloschi, G. Tocchi, M. Gaetani d'Aragona, REPORT ON EXPOSURE DATA AND VULNERABILITY MODELS, Deliverable D1.1 of EXPLORA project, 2023.

[26] V. Silva, R. Sousa, F.R. Gouveia, J. Lopes, M.J. Guerreiro, A building imagery database for the calibration of machine learning algorithms, *Earthquake Spectra*, **40(2)**, 1577–1590, 2024.

[27] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[28] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, *arXiv*, Cornell University, 2014.

[29] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely Connected Convolutional Networks, *arXiv*, Cornell University, 2016.

[30] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, **86(11)**, 2278–2324, 1998.

[31] E. Tuba, N. Bačanin, I. Strumberger, M. Tuba, Convolutional Neural Networks Hyperparameters Tuning. *Studies in Computational Intelligence*, Springer International Publishing, 65–84, 2021.

[32] H.N. Fakhouri, S. Alawadi, F.M. Awaysheh, F. Hamad, Novel hybrid success history intelligent optimizer with Gaussian transformation: application in CNN hyperparameter tuning, *Cluster Computing*, **27(3)**, 3717–3739, 2023.

[33] F. Radenovic, G. Tolias, O. Chum, Fine-Tuning CNN Image Retrieval with No Human Annotation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **41(7)**, 1655–1668, 2019.

[34] Y. Zheng, R. Zhang, J. Zhang, Y. YeYanhan, Z. Luo, LlamaFactory: Unified Efficient Fine-Tuning of 100+ Language Models, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, 400–410, 2024.

[35] F. Chollet, Xception: Deep Learning with Depthwise Separable Convolutions, *arXiv*, 2016.

[36] M. Tan, Q.V. Le, EfficientNetV2: Smaller Models and Faster Training, *arXiv*, 2021.

[37] M.T. Ribeiro, S. Singh, C. Guestrin, 'Why Should I Trust You?' *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, ACM, 2016.

[38] R. Andrews, J. Diederich, A.B. Tickle, Survey and critique of techniques for extracting rules from trained artificial neural networks, *Knowledge-Based Systems*, **8(6)**, 373–389, 1995.

[39] J.R. Zilke, Extracting Rules from Deep Neural Networks, Master's thesis, Technische Universität Darmstadt, 2016.

[40] L. Fu, Rule generation from neural networks, *IEEE Transactions on Systems, Man, and Cybernetics*, **24(8)**, 1114–1124, 1994.

[41] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, W. Samek, On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation, *PLOS ONE*, **10(7)**, e0130140, 2015.

[42] A. Shrikumar, P. Greenside, A. Kundaje, Learning important features through propagating activation differences, *arXiv preprint*, arXiv:1704.02685, 2017.

[43] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on*, IEEE, 2016.

[44] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, *arXiv*, arXiv:1610.02391, 2016.