Project 1: Are you in a safe building?

CP217: Machine Learning for Cyber-Physical Systems August-Dec Semester 2024

Team: Saferide Squad

Vanshika Jindal-24721, Vandana Mourya-22969, Guha H-24639

1 Introduction

In the field of disaster risk management, accurately assessing the structural vulnerability of buildings to earthquakes is crucial for minimizing casualties and economic losses. Recent advances in data availability, particularly street-view images, have provided new opportunities for rapid and scalable seismic risk assessments. In this project, our goal is to develop a machine learning model capable of automatically classifying building types and materials from street-view images. This classification helps in identifying buildings that are more susceptible to collapse during an earthquake, thereby improving earthquake preparedness efforts. The task involves training an ML model on a dataset of building images categorized into five distinct classes. Our approach focuses on image pre-processing, feature extraction, and model selection to create a robust classifier that generalizes well to unseen data.

2 Dataset Observation

The dataset for this project contains over 2,500 grayscale images of buildings, categorized into five classes: A, B, C, D, and S. Each 400x300 pixel image has three identical channels, organized into folders by class. A key challenge identified was the visual similarity between building types, making classification difficult, along with obstructions like trees or tilted angles.

The dataset is also highly imbalanced, with the following distribution: - Class A: 299 images, Class B: 362 images, Class C: 731 images, Class D: 915 images, Class S: 210 images. This imbalance required careful handling to avoid biased predictions. The training data was split 80:20 for training and validation. Data augmentation techniques were applied, and images were normalized between 0 and 1 to reduce computational demands. The test set contained 478 additional images for final evaluation.

3 Model Development

3.1 Baseline Machine Learning models

We began with traditional machine learning approaches to establish baseline results for our image classification task, experimenting with Logistic Regression, Random Forest, and Support Vector Machines (SVM). Logistic Regression initially achieved 31% validation accuracy, but after incorporating Histogram of Oriented Gradients (HOG) for feature extraction, the validation accuracy improved to 44%, showing the importance of specialized feature extraction in image classification. Random Forest, known for handling feature importance, offered a slightly better result with 43% validation accuracy. The SVM model started at 32% accuracy but improved to 50% after hyperparameter tuning, particularly by adjusting the regularization parameter (C) and exploring different kernel types, indicating that non-linear structures were present in the data. Despite these improvements, traditional models faced significant limitations, particularly due to data imbalance, where some classes dominated predictions, and the small dataset size (2,000 images) made it challenging for models to generalize without overfitting. These models also lacked the complexity needed to capture intricate spatial features in images.

3.2 Deep Learning and Transfer Learning models

Given the limitations of traditional machine learning models, we transitioned to deep learning, particularly transfer learning, to leverage pre-trained models that had already learned complex features from large datasets. ResNet50 was the first deep learning model we explored. After applying basic data augmentation techniques, it initially achieved a validation accuracy of 47%, a notable improvement over traditional methods but still insufficient for our goals. VGG16 performed slightly better, with a validation accuracy of 47.6%. However, like ResNet50, it struggled with overfitting due to the relatively small size of our dataset.We further improved the performance of ResNet50 by fine-tuning the model, unfreezing the last few layers. This led to a validation accuracy improvement, reaching 56%. We also experimented with the Vision Transformer (ViT) model, which achieved a training accuracy of 57%, but its validation accuracy dropped to 43.85%, highlighting poor generalization. Transformers generally perform best on large datasets, so this result was expected given the data limitations. These findings confirmed that our dataset size was the primary bottleneck, and that we needed to focus on augmentations and explore more sophisticated models to boost performance further.

3.3 Final Model selection

Recognizing the need for a model suited to small datasets, we tested several architectures, including EfficientNetB0, DenseNet, and Inception, while incorporating data augmentation techniques to enhance generalization. EfficientNetB0, which was pre-trained on the ImageNet dataset, emerged as the standout model, achieving a training accuracy of 89.33% and a validation accuracy of 64.13%. Its balanced scaling of model depth, width, and resolution made it a robust choice for our relatively small dataset. In contrast, DenseNet significantly underperformed, with a validation accuracy of only 19.44%. This poor performance was likely due to its high complexity, which made it prone to overfitting on the limited data available. Inception, while performing better than DenseNet, achieved a validation accuracy of 32.14%, still falling short of the results obtained with EfficientNetB0.

To enhance the performance of the EfficientNetB0 model and address issues related to overfitting and class imbalance, we implemented several key techniques. First, data augmentation was employed, involving a variety of transformations such as horizontal flips, rotations, brightness adjustments, and Gaussian blur. These augmentations artificially increased the dataset's size and variability, thereby aiding the model in generalizing better to unseen data. Additionally, we incorporated Batch Normalization and Dropout layers into the EfficientNetB0 architecture to mitigate overfitting. The use of Dropout, with a rate of 0.5, proved particularly effective in preventing the model from memorizing the training data. Lastly, we applied a custom Focal Loss function to specifically tackle the class imbalance issue. This loss function emphasizes the learning from harder-to-classify samples, ensuring that the model pays adequate attention to minority classes and effectively learns to classify them.

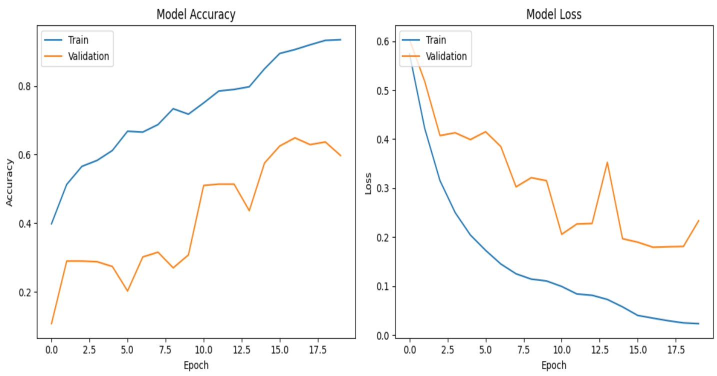

Figure 1: Training, Validation accuracy and loss plot

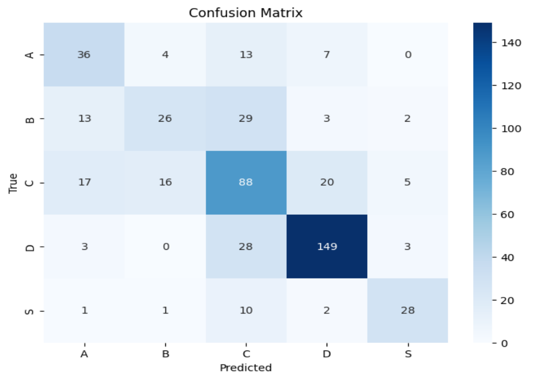

Figure 2: Confusion matrix - EfficientNetB0 model

4 Computation Resource used

While simpler models, such as Logistic Regression and Random Forest, were run on the CPU, the deep learning models leveraged the computational power of accelerator GPUs provided by Kaggle, specifically utilizing GPU T4. This significant increase in processing capability allowed for faster training and more efficient handling of the complex architectures required for image classification tasks, enabling the models to learn from the data more effectively.

5 Further attempts to improve the model

To further improve our model, we extracted features from the last hidden layer of EfficientNet and applied PCA to reduce the dimensionality. We found that 1,500 features were sufficient to explain 95% of the variance in the data. PCA revealed that the data was highly non-linear, and linear classifiers like SVM with RBF kernel performed well in this space, giving us an outstanding accuracy of 97%.Our next step involves exploring additional non-linear classifiers and experimenting with ensembling methods to push the performance further.

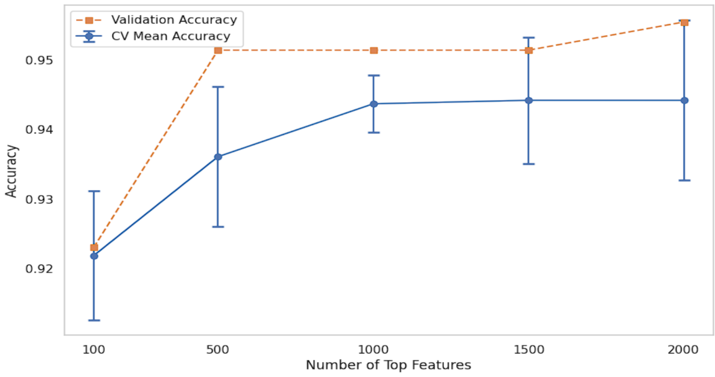

Figure 3: Accuracy vs Number of top features

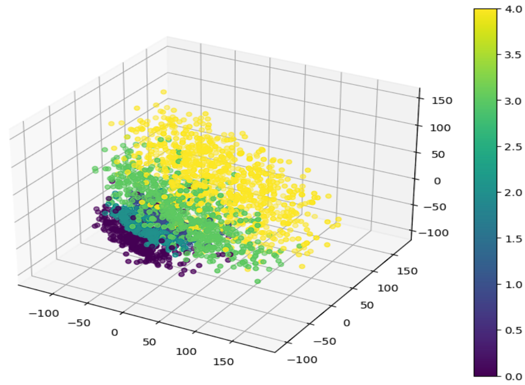

Figure 4: PCA features in 3D

6 Conclusion

In this project, we developed a machine learning model to classify building types and materials from street-view images, addressing a key need in disaster risk management. After exploring various machine learning and deep learning architectures, we found traditional methods limited by data imbalance and small dataset size, prompting a shift to deep learning.

Using transfer learning with models like ResNet50, VGG16, and ultimately EfficientNetB0, we achieved notable improvements. EfficientNetB0, pre-trained on ImageNet, was the most effective, reaching 89.33% training accuracy and 64.13% validation accuracy. Data augmentation, Batch Normalization, Dropout, and a custom Focal Loss function helped address overfitting and class imbalance.

Our findings highlight the importance of deep learning and data-handling strategies in assessing building vulnerability. This research sets the stage for improving seismic risk assessments, contributing to better earthquake preparedness in urban environments.

References

[1] Murdoch, Ryan & Al-Habashna, Ala'a. (2023). Residential building type classification from street-view imagery with convolutional neural networks. *Signal, Image and Video Processing*, 18, 1-10. `https://doi.org/10.1007/s11760-023-02882-8`

[2] Metana Blog. (2023). Deep Learning Models for Classification: A Comprehensive Guide. `https://metana.io/blog/deep-learning-models-for-classification-a-comprehensive-guide/`

[3] ChatGPT. (2023). ChatGPT Homepage. `https://chatgpt.com/`