mlenthusiast at NakbaArchiveClassifier Shared Task: A Lightweight SVM-Gated Ensemble of EfficientNets for Image Classification

Md. Ajwad Hossain
Department of Electronics and Telecommunication Engineering
Chittagong University of Engineering & Technology (CUET)
Chattogram, Bangladesh
md.ajwadhossain@gmail.com

Abstract

Image classification under strict time constraints requires a delicate balance between feature complexity and computational overhead. This paper presents an optimized ensemble methodology developed for the NAKABA competition, focusing on identifying structural destruction. We propose a hybrid architecture that leverages two distinct Convolutional Neural Networks (EfficientNetB0 and EfficientNetB3) as base feature extractors, coupled with a Support Vector Machine (SVM) functioning as a meta-classifier. Instead of standard probability averaging or processing high-dimensional embeddings directly, the Meta-SVM acts as a learned gating mechanism to optimally combine the low-dimensional probability predictions of the base models. This ensures robust performance without the latency of heavier deep learning architectures. Empirical results demonstrate the efficacy of this approach. The model achieved a validation accuracy of 0.884 and a weighted F1-score of 0.885, with a notable F1-score of 0.839 on the challenging 'destruction' class. On the official NAKABA leaderboard test set, the ensemble maintained strong generalization, achieving an F1-score of 0.831 and an accuracy of 0.845, which secured the 12th position overall and proved the model's high effectiveness within the competition's strict operational constraints.

Keywords: Ensemble Learning, EfficientNet, Support Vector Machine, Stacking, Time-Constrained Inference

1. Introduction

Identifying critical states (e.g., complete structural "destruction") from photographs is a complex visual task. Importantly, predictions must be made under strict latency constraints. In real-world scenarios like post-disaster assessment or real-time structural health monitoring, models must deliver near-real-time inference on constrained hardware (Fan, 2024). In such settings, lightweight network architectures are essential, since they must execute rapidly on remote devices, such as drones, with limited computational capacity. This creates a critical trade-off: we require models with very high accuracy for safety-critical damage detection, alongside very low inference latency.

Standard deep neural networks (e.g., large ResNets or Transformers) can achieve high accuracy but typically incur prohibitive latency. Recent work has introduced model families that are far more compact: for example, EfficientNet models use compound scaling to balance depth, width, and resolution, yielding state-of-the-art accuracy with far fewer parameters. In particular, EfficientNet architectures achieve top-tier ImageNet accuracy while being significantly smaller and faster at inference than previous benchmarks (Tan and Le, 2019). Successive iterations further focus on parameter efficiency and training speed, producing models that execute even faster (Tan and Le, 2021).

Meanwhile, ensemble methods are known to improve robustness by combining multiple classifiers. As demonstrated by Dietterich (2000), aggregating diverse models frequently outperforms any single constituent model. In theory, larger ensembles yield better results, but naively ensembling multiple large Convolutional Neural Networks (CNNs) violates strict time budgets. Thus, we leverage ensembles strategically: we utilize multiple EfficientNet-based feature extractors (each relatively small and fast) and merge their outputs via a highly optimized, lightweight decision rule.

Recent advancements in ensemble fusion have explored geometric stacking methods, such as solving maximum-weight rectangle problems in high-dimensional spaces to separate classes (Wu et al., 2023, 2024). While effective, these can introduce unnecessary computational overhead. Instead, we propose an architecture that decouples heavy feature extraction from the final decision using a streamlined geometric approach. Multiple EfficientNet variants (B0 and B3) run in parallel to extract features, and a Support Vector Machine (SVM) (Cortes and Vapnik, 1995) meta-classifier combines their probabilistic outputs. By embedding the base-model predictions into a highly compressed, 4-dimensional space and solving for an optimal linear hyperplane, the meta-classifier effectively learns a robust gating mechanism. In summary, by combining efficient CNN feature extractors with a fast SVM meta-learner, our pipeline achieves high accuracy on the structural damage classification

Proceedings of the 2nd International Workshop on Nakba Narratives as Language Resources (Nakba-NLP 2026) @ LREC 2026, pages 217–220
11 May 2026. ©ELRA Language Resources Association (ELRA), 2026

task while strictly satisfying the competition's latency requirements.

2. Dataset and Task Description

The proposed ensemble architecture was developed and evaluated using the official dataset provided by the NAKABA competition (Abrahams et al., 2026). The primary objective of this task is the binary classification of structural images to rapidly identify critical damage.

2.1. Data Composition

The development dataset consists of a total of 2,001 labeled photographs categorized into two mutually exclusive classes: `not_destruction` (1,295 instances) and `destruction` (706 instances). This natural class imbalance reflects the realities of post-disaster structural assessments. To ensure robust model evaluation and tune our meta-classifier without overfitting, this available data was partitioned into:

- **Training Set:** 1,400 images utilized to train the base EfficientNet feature extractors and the Meta-SVM.
- **Validation Set:** 199 images used to evaluate local generalization and tune hyperparameters (e.g., the SVM regularization parameter C).

The final evaluation was conducted on an unseen hidden test set maintained by the competition organizers on the official leaderboard.

2.2. Preprocessing and Optimization

Given the variance in real-world structural photographs, a strict preprocessing pipeline was applied prior to feature extraction. To align with the compound scaling principles of the EfficientNet family (Tan and Le, 2019), architecture-specific input resolutions were utilized to maximize feature extraction efficiency and preserve pre-trained weight integrity.

Images processed by the EfficientNetB0 baseline were resized to 224×224 pixels, providing a lightweight input footprint for rapid foundational pattern recognition. Conversely, images processed by the EfficientNetB3 baseline were resized to 300×300 pixels, allowing the deeper network to capture more complex, fine-grained spatial features without distortion. Furthermore, all pixel values were normalized using standard ImageNet mean and standard deviation metrics to ensure stable convergence and consistent feature mappings across both parallel extraction branches.

3. Methodology

The core of the proposed architecture is a meta-learning ensemble utilizing a stacking mechanism to combine predictive signals. The pipeline is divided into feature extraction, base probability generation, and meta-classification.

3.1. Base Feature Extractors

We utilize two pre-trained models from the EfficientNet family (Tan and Le, 2019), chosen for their optimal scaling of depth, width, and resolution:

- **EfficientNetB0:** Serves as a highly lightweight, rapid feature extractor capable of capturing foundational visual patterns with minimal computational overhead.
- **EfficientNetB3:** A moderately larger variant that captures more complex, fine-grained spatial features without suffering from the extreme latency associated with heavier architectures.

In our pipeline, the final classification heads of both networks are bypassed. Instead, the models are used to output dense spatial feature embeddings (1280 and 1536 dimensions, respectively) for each input image.

3.2. SVM Meta-Classifier via Stacking

Instead of employing simple soft-voting or training an additional dense neural network layer on the high-dimensional embeddings—which would significantly increase inference latency and risk overfitting—we utilize a Support Vector Machine (SVM) (Cortes and Vapnik, 1995) in a stacked generalization architecture.

Figure 1 outlines this approach. To minimize latency, we do not concatenate the massive EfficientNet feature vectors directly. Instead, we train two independent Base SVMs that reduce these dense features into 2-dimensional probability vectors representing the likelihood of the 'destruction' and 'not_destruction' classes.

These outputs are then concatenated into a highly compact, 4-dimensional vector. Finally, a linear Meta-SVM ($C = 0.1$) evaluates this combined input. In this stacked configuration, the Meta-SVM effectively acts as a learned gating mechanism. By finding the optimal hyperplane in this highly compressed probability space, the meta-classifier learns exactly which base model's predictions to trust under varying conditions, ensuring the final classification is both highly accurate and exceptionally fast.

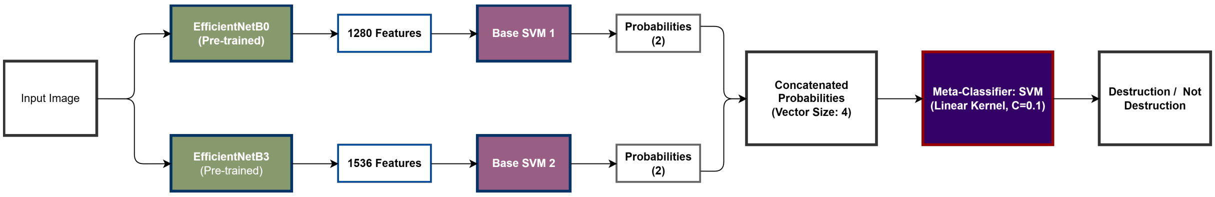

Figure 1: Proposed Hybrid Architecture: Input images are processed in parallel by EfficientNet variants to extract dense feature embeddings. These features are reduced to probability vectors by Base SVMs, which are then concatenated and classified by the final SVM meta-classifier.

3.3. Implementation Details and Hardware

To ensure complete reproducibility and optimal computational efficiency, all experiments were conducted within a Google Colaboratory (Colab) environment, accelerated by an NVIDIA T4 GPU.

A global random seed of 42 was strictly enforced across all computational libraries (including data partitioning, feature extraction, and SVM initialization). During the preliminary processing and training phases, the dataset was ingested using a batch size of 32.

4. Experiments and Results

4.1. Validation Performance

During local evaluation, the ensemble demonstrated marked improvements over the standalone models. The meta-classifier successfully filtered out individual model errors.

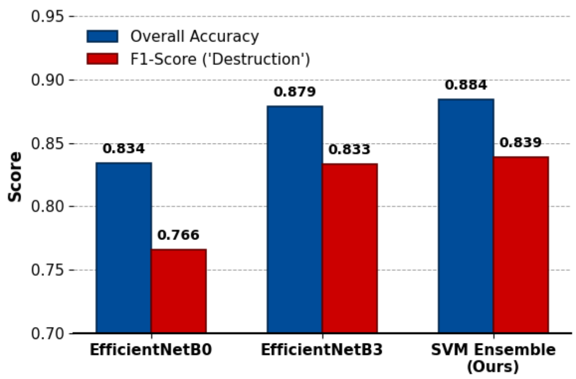

Figure 2: Performance Comparison: The ensemble model shows a distinct improvement in identifying the 'destruction' class compared to the standalone baselines.

The F1-score for the critical 'destruction' class improved to 0.839 compared to the best individual base model (EfficientNetB3), which achieved 0.833.

4.2. Leaderboard Test Performance

The ultimate validation of the time-constrained ensemble was its performance on the official hidden test set of the NAKABA competition. The model generated predictions efficiently and yielded highly competitive results on the leaderboard:

Metric	Score
Test Accuracy	0.84577
Test F1-Score	0.83123

Table 1: Official NAKABA Leaderboard Test Results

4.3. Discussion of Ensemble Effectiveness

The proposed ensemble demonstrated a distinct advantage over individual feature extractors by successfully leveraging their complementary strengths. While the EfficientNetB3 base model provided a strong foundation (validation accuracy of 0.879 and weighted F1-score of 0.880), it struggled slightly more with the critical 'destruction' class.

By extracting the output probabilities of both EfficientNetB0 and EfficientNetB3 and feeding them into an SVM meta-classifier—tuned with a linear kernel and a regularization parameter of $C = 0.1$—the architecture effectively learned to correct individual model errors. This meta-learning approach yielded a cross-validation weighted F1-score of 0.9900 during training, indicating a highly effective integration of the base models' probabilistic signals.

The final validation performance confirmed this advantage. The ensemble pushed the overall accuracy to 0.884 and achieved a highly balanced predictive capability across classes. Specifically, the F1-score for 'not_destruction' reached 0.910, while the performance on the more challenging 'destruction' class saw a meaningful improvement to 0.839. This confirms that simple linear combinations of probabilities via a meta-classifier can be exceptionally powerful when the base models capture distinct yet valuable spatial hierarchies.

5. Conclusion and Future Work

This study details an efficient, SVM-gated ensemble methodology tailored for the time-constrained NAKABA image classification challenge. By decoupling heavy feature extraction from the final decision-making layer, the architecture maximizes the utility of EfficientNet variants without incurring prohibitive latency.

The empirical results validate the robustness of this approach. The ensemble model not only surpassed the standalone base models—achieving an overall accuracy of 0.884 and a weighted F1-score of 0.885—but also provided a critical performance uplift in identifying structural destruction. The final system successfully generated rapid predictions for the unseen test dataset, yielding highly competitive leaderboard scores that secured the 12th position overall.

Future work will explore the integration of additional lightweight base extractors and the potential of non-linear gating mechanisms to further refine decision boundaries under strict operational constraints.

6. Limitations

While the proposed SVM-gated ensemble achieves high accuracy, its performance is fundamentally constrained by the diversity of the training data. The model was optimized specifically for the structural damage profiles present in the NAKABA dataset. Deploying this system to novel geographic regions or vastly different architectural styles may require domain adaptation or fine-tuning of the base EfficientNet feature extractors.

7. Acknowledgements

The author would like to express gratitude to the organizers of the NAKABA challenge for providing a rigorous and stimulating platform for applied machine learning.

8. Bibliographical References

Alexei Abrahams, Shadi Abudalfa, Mustafa Jarrar, and George Mikros. 2026. The nakbaarchive-classifier shared task on nakba image classification. In *Proceedings of the 2nd International Workshop on Nakba Narratives as Language Resources (Nakba-NLP 2026), co-located with the Language Resources and Evaluation Conference (LREC 2026)*, Palma, Mallorca, Spain.

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning*, 20(3):273–297.

Thomas G. Dietterich. 2000. Ensemble methods in machine learning. In *Multiple Classifier Systems*, pages 1–15, Berlin, Heidelberg. Springer Berlin Heidelberg.

Ching-Lung Fan. 2024. Deep neural networks for automated damage classification in image-based visual data of reinforced concrete structures. *Heliyon*, 10(19):e38104.

Mingxing Tan and Quoc Le. 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR.

Mingxing Tan and Quoc V. Le. 2021. Efficientnetv2: Smaller models and faster training. *CoRR*, abs/2104.00298.

Shengli Wu, Jinlong Li, and Weimin Ding. 2023. A geometric framework for multiclass ensemble classifiers. *Machine Learning*, 112:4929–4958.

Wei Wu, Liang Tang, Zhongjie Zhao, and Chung-Piaw Teo. 2024. Enhancing binary classification: A new stacking method via leveraging computational geometry.