Dimensionality Reduction for Satellite Ship Detection: Accuracy, Speed, and Real-World Performance Trade-offs

Matty J. Maloni
Department of Electrical and Computer Engineering
University of Florida,
Gainesville, FL
mmaloni1@ufl.edu

***Abstract* – This study evaluates dimensionality reduction for satellite ship detection, comparing baseline, PCA, and ISOMAP approaches across Logistic Regression, Random Forest, and SVM using 4,000 Planet Labs satellite images. Test metrics suggested PCA Random Forest as optimal (97% accuracy, 9,090 images/second). However, operational validation revealed catastrophic failure: Random Forest detected only 1-2 ships per scene while PCA SVM consistently identified all ships with high confidence and 2-3 false positives. ISOMAP achieved extreme speeds but suffered 4-6% overfitting gaps. Despite 0.25% lower test accuracy, PCA SVM proves operationally superior, demonstrating that test performance is insufficient for deployment decisions. This work establishes PCA SVM as optimal for balancing accuracy (96.75%), speed (82× faster than baseline), and real-world reliability, while highlighting the critical importance of operational validation beyond benchmark metrics.**

I. Introduction

A. *Industry Context*

More than 80% of goods are transported by sea [1]. Maritime transport generates massive volumes of data through multiple channels: onboard sensors, satellite imagery, port management systems, and AIS (Automatic Identification System) transponders. Compared to the others, satellites provide independent, comprehensive, tamper-proof monitoring, something impossible with cooperative systems like AIS. Commercial imagery providers such as Planet Labs [2] operate constellations of small satellites that revisit regions daily at 3-meter resolution, making frequent, consistent monitoring feasible. This accessibility of data creates an operational need: automated ship detection algorithms that can process imagery faster than manual analysis to enable real-time supply chain monitoring, anomalous activity detection, and port capacity planning.

B. *Impact for Companies*

Translating this operational need into deployed systems introduces a practical constraint: computational efficiency. Processing satellite imagery at scale demands not only high accuracy but also fast inference and manageable computational costs. A detection system that achieves 95% accuracy but requires excess time or resources per image is limited in practical value. To create tools that can balance time and accuracy, one must explore and analyze the trade-offs.

C. *Research Goals*

To address this problem, this study applies dimensionality reduction to three classification methods (Logistic Regression, Random Forest, and Support Vector Machine) to explore the costs and benefits of baseline and dimensionality-reduction-enhanced versions. Through evaluation of accuracy and efficiency, this paper will determine which dimensionality reduction approach best balances performance with scalability for operational deployment.

Figure 1: Visualizing the detections on satellite imagery from San Pedro Bay using SVM (note the 2 false positives).

II. Data

A. *Dataset Introduction*

This study consists of images extracted from Planet satellite imagery collected over the San Fransico and San Pedro Bay areas of California. It includes 4000 80 x 80 RGB images labeled with either a “ship” or “no ship” classification. Images were derived from PlanetScope full-frame visual scene products, which are orthorectified to a 3-meter pixel size. There are 1000 images of ships of different sizes, orientations, and atmospheric collection conditions. There are 1000 images of a random sampling of land cover features - water, vegetation, bare earth, buildings, etc. - that do not include any portion of a ship. The next 1000 are partial ships, images which do not contain enough of a ship to meet the full definition of a “ship”. The last 1000 are images that have been previously mislabeled by machine learning models.

B. Data Preprocessing

Images are flattened into 19,200-dimensional feature vectors (80 × 80 × 3 channels). No additional feature engineering or augmentation was applied; raw pixel values are used directly. Where applicable, features are standardized to zero mean and unit variance before model training to ensure consistent scaling across classifiers. The dataset is split into training (80%) and testing (20%) sets, stratified by class label to ensure balanced representation of ship/no-ship cases in both partitions.

III. Baseline Models

A. Logistic Regression

Logistic Regression was implemented to create a simple, interpretable baseline. Using hyperparameter tuning and 5-fold cross-validation, it achieved a peak accuracy of 93% with a C (regularization parameter) of .001. Strong regularization working best suggests that the decision boundary may be simple or that the model struggles with noise. The high accuracy, however, suggests the patterns are largely linearly separable and that the ships in the images have consistent enough visual characteristics. The model achieved higher recall on the majority no-ship class (0.96) than the minority ship class (0.82), reflecting class imbalance in the dataset. The singular model took 2.34 seconds to train and an average of 2.86 milliseconds to interpret per image.

B. Random Forest

To capture more nonlinear patterns, Random Forest was applied. Using hyperparameter tuning and 5-fold cross-validation, it achieved a peak accuracy of 95.4% with a max depth of 20, minimum samples split of 2, and 100 estimators. The model benefits from complexity to capture nonlinear patterns in ship detection but doesn't require extremely deep trees or massive ensembles. Like Logistic Regression, it achieved higher recall on the majority no-ship class (0.98) than the minority ship class (0.87), again reflecting class imbalance in the dataset. It had a training time of 17.5 seconds and an average interpretation time of 19.31 milliseconds.

C. Support Vector Machine

To attack the issue of high dimensionality, for the final classifier SVM (support vector machine) was used. Using hyperparameter tuning and 5-fold cross-validation, it achieved a peak accuracy of 97.1% with a C of 50, gamma of 'scale', and the RBF kernel. It had .92 recall on ships (the best yet) and a .99 recall on non-ships. The minimal class performance gap suggests SVM handles the imbalance better. However, this model took 20.5 seconds to train and had an average interpretation time of 23.95 milliseconds.

IV. Dimensionality Reduction

A. Principal Component Analysis

To experiment how to make the models faster and potentially more effective, PCA (principal component analysis) was applied in different ways to the dataset. PCA finds the most important patterns in the data and throws away the noise. It does this by deciding which combinations of pixels actually matter in the data. A 5-fold Grid Search combined with visual inspection of reconstructed images determined the optimal number of PCA components for balancing speed and accuracy. It was also found that 108 components explain 90% of the variance (Figure 2).

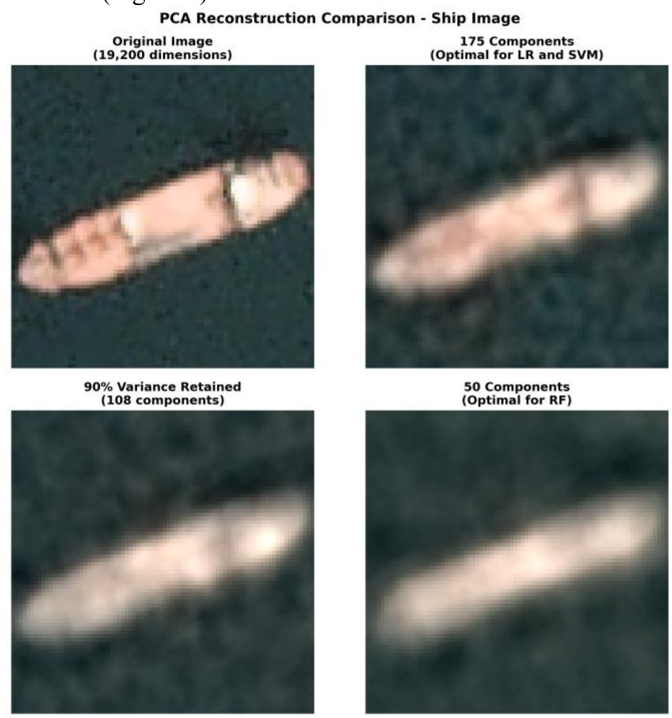

Figure 2: Visualizing PCA. Reconstructed images after PCA using different numbers of components and variance levels.

For Logistic Regression, 175 components were used with a C again of .001. This model achieved an accuracy of 92.8%. It took 8.33 seconds to train and an average inference time of 0.21 milliseconds per image. Although slightly worse accuracy and training time, the quick inference time makes this optimal for quick, real-time systems.

For Random Forest, 50 components produced the fastest and most accurate model with the hyperparameters of no max depth, a minimum of 2 samples per split, and 100 estimators. It achieved 97% accuracy while training in only 5.3 seconds, with an average inference time of 0.11 milliseconds per image. This represents a dramatic operational improvement over baseline: 70% faster training and 175× faster inference, enabling real-time processing of approximately 9,000 images per second compared to baseline's 52 images per second.

SVM, also found considerable improvement. The optimal parameters were 175 components, with a C of 10, a gamma of 'scale' and the rbf kernel. The model took 8.85 seconds to train. PCA-enhanced SVM sacrifices only 0.35 percentage points of accuracy (96.75% vs 97.1%) but achieves approximately 82x faster inference (0.29ms vs 23.95ms per image).

B. ISOMAP Manifold Learning

To identify optimal hyperparameters for ISOMAP-enhanced classifiers, a systematic grid search was conducted over two key parameters: the number of principal components (50, 100, 175) and the number of neighbors used for manifold construction (5, 10, 15). Each configuration was evaluated using 5-fold cross-validation, with performance measured by classification accuracy across all four classifiers. The grid search revealed that 100 components with 15 neighbors provided the best performance across classifiers.

ISOMAP transformation was fitted once on the training data and subsequently applied to both validation and test sets, avoiding redundant manifold computation while maintaining proper train-test separation.

For Logistic Regression, using a tuned C of 0.1 (grid search), the trained in 9.2 milliseconds with an accuracy of 91.5%. This model had an average inference time of 1.39 microseconds, or the ability to process approximately 720,000 thousand images per second.

For Random Forest, using a grid search-tuned max depth of 20, minimum sample splits of 5, and 20 estimators, the model achieved a 93.9% accuracy, with a training time of 0.59 seconds and an (again crazy) inference time of 2.56 microseconds. It favoring these lower parameters tell signs of overfitting, an issue explored after presenting the SVM results.

For SVM, again using grid-search-tune a C of 7, using scale for gamma, and the rbf kernel was found to be optimal. This achieved a test accuracy of 94.6%. The training time took 0.16 seconds while the inference time took 65.58 microseconds. Although not as fast as Logistic Regression, that still gives the ability to handle approximately 15,000 images per second. The lower C also suggests the model is trying harder to avoid overfitting.

Both SVM and RF suffer from overfitting when using the ISOMAP trained dataset. Despite achieving high test accuracies, both models exhibited significant train-test performance gaps. Hyperparameter tuning reduced these gaps but could not eliminate the underlying overfitting. Random Forest exhibited a training accuracy of 99.8% percent, a gap of 5.9% from its test accuracy. SVM exhibited also a very high training accuracy, 99.4%, a gap of 3.9% from its test accuracy. This overfitting behavior, combined with the modest test accuracies, suggests ISOMAP's non-linear manifold learning may be encoding training-specific noise patterns that fail to generalize—a fundamental limitation addressed in the comparative analysis below.

C. *PCA vs ISOMAP Manifold Learning*

To visualize how each dimensionality reduction method captures class structure, Figures 3 and 4 show the training data

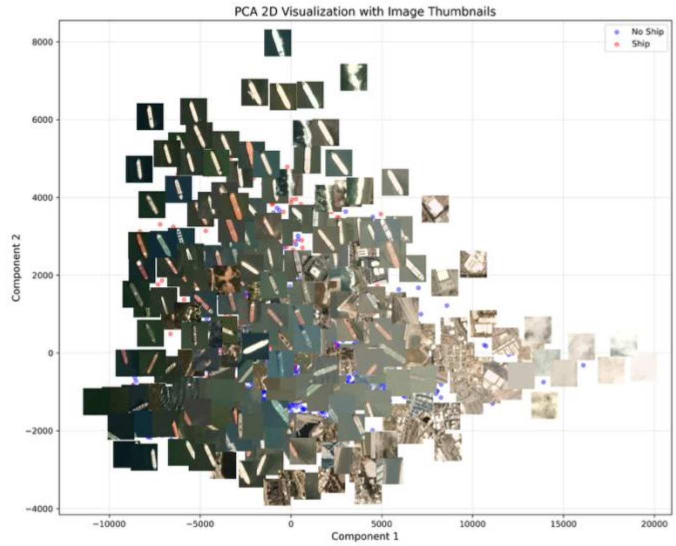

Figure 3: PCA Component PCA projection of training data onto first two principal components. Classes are more clearly separated and clustered.

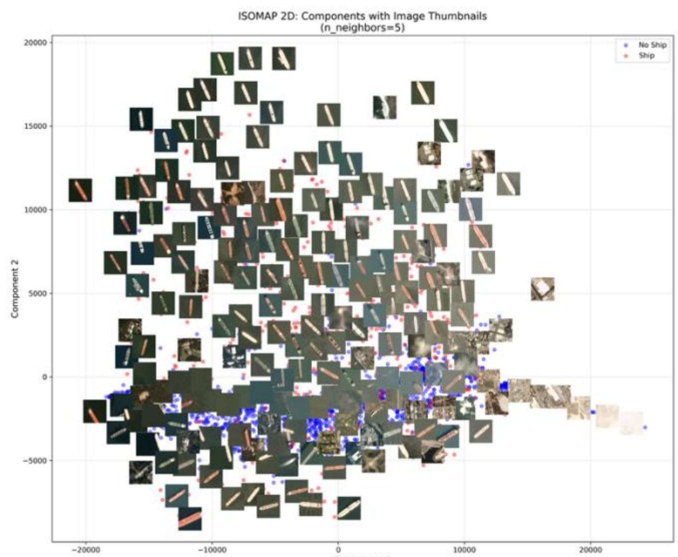

Figure 4: ISOMAP projection of training data onto first two manifold dimensions. Increased class overlap as well as spread out ship class.

projected onto the first two components of PCA and ISOMAP respectively. PCA (Figure 3) reveals a more clear separation between ship and no-ship classes with distinct clustering, while ISOMAP (Figure 4) shows greater class overlap and less defined boundaries. This visual difference suggests PCA's linear components align better with the discriminative features needed for ship detection, foreshadowing the performance differences observed in model evaluation.

V. Results and Model Comparisons

A. Test Set Performance Summary

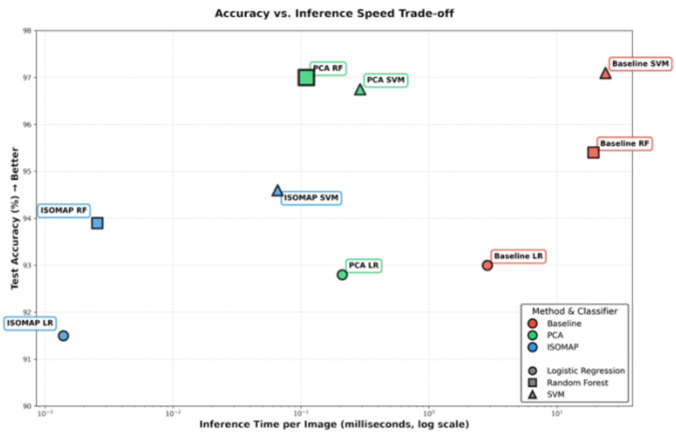

Figure 5: Accuracy vs. inference speed trade-off for baseline and dimensionality-reduced classifiers. PCA Random Forest (highlighted) achieves optimal balance with 97% accuracy and 0.11ms inference time.

B. Key Findings

PCA dramatically improves computational efficiency while maintaining accuracy. PCA Random Forest achieves 97% accuracy with 175× faster inference than baseline, enabling real-time processing at 9,090 images/second. PCA SVM achieves 82× speedup with only 0.35 percentage point accuracy loss compared to baseline.

ISOMAP achieves extreme speeds but suffers severe overfitting. Random Forest shows 5.9% train-test gap (99.8% vs 93.9%), and SVM shows 4.8% gap (99.4% vs 94.6%), indicating the non-linear manifold learning encodes training-specific noise rather than generalizable patterns. Figure 5 shows the accuracy-speed trade-off. Based on test metrics alone, PCA Random Forest appears optimal. However, operational validation reveals a different story.

VI. Operational Deployment Validation

A. Sliding Window Testing

Models were validated on full satellite scenes from San Francisco and San Pedro Bay using sliding windows (80×80 pixels, stride=10, confidence threshold=0.8, NMS threshold=0.1). Results revealed critical failures invisible in test metrics.

B. Real-World Performance

PCA SVM: Consistently detects all ships (occasionally missing 1-2), high confidence (>83%), 1-2 false positives per scene. Operationally reliable.

PCA Random Forest: Detects only 1-2 ships with medium confidence. Lowering the threshold floods the system with false positives. Unusable for deployment.

This gap reveals Random Forest overfits to clean, centered training examples. When confronting partial ships, varied positions, and complex backgrounds from sliding windows, RF fails catastrophically. SVM's margin-based approach generalizes robustly.

C. Final Recommendation

Despite 0.25% lower test accuracy, PCA SVM is the clear operational choice. It achieves near-complete detection, high confidence predictions, minimal false positives, and practical inference speed (3,448 img/s).

This demonstrates a critical lesson: test accuracy is necessary but not sufficient. Models must be validated under realistic deployment conditions. PCA RF's superior test metrics proved meaningless when it failed in production scenarios.

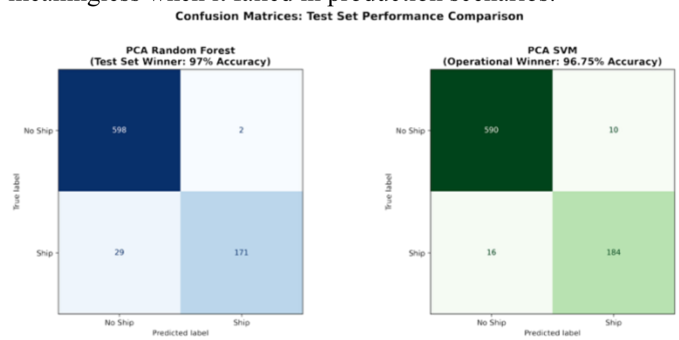

Figure 6: Confusion matrices for PCA Random Forest and PCA SVM. Both achieve 97% accuracy with different error patterns: RF has perfect no-ship recall but misses 11% of ships; SVM shows balanced performance across both classes.

Conclusion

This study evaluated dimensionality reduction approaches for satellite ship detection, comparing baseline, PCA, and ISOMAP methods across Logistic Regression, Random Forest, and SVM classifiers. Test metrics suggested PCA Random Forest as optimal (97% accuracy, 9,090 img/s throughput). However, operational validation using sliding window detection revealed PCA Random Forest's catastrophic failure in real-world conditions, detecting only 1-2 ships per scene.

PCA SVM emerges as the definitive operational solution, achieving 96.75% test accuracy, 82× faster inference than baseline, and—most critically—robust real-world performance with near-complete ship detection and minimal false positives.

Key contributions:

Demonstrated PCA's superiority over ISOMAP for this task, with ISOMAP showing 4-6% overfitting gaps.

Revealed the critical importance of operational validation beyond test metrics.

Established PCA SVM as the optimal balance of accuracy, speed, and real-world reliability.

Future work should explore ensemble methods combining SVM's robustness with Random Forest's speed and investigate why Random Forest fails to generalize to sliding window scenarios despite strong test performance.

References

[1] World Bank, "Sustainable Development in Shipping and Ports," World Bank Group, 2024. [Online]. Available: https://www.worldbank.org/en/topic/transport/brief/sustainable-development-in-shipping-and-ports

[2] Planet Labs PBC, "Satellite Monitoring," Planet Labs, 2024. [Online]. Available: https://www.planet.com/products/satellite-monitoring/