MULTI-TASK LEARNING FOR HUMANITARIAN DEMINING OPERATIONS: A COMPARATIVE ANALYSIS OF PERCEPTION ALGORITHMS

A Thesis Submitted to the Committee on Graduate Studies
in Partial Fulfillment of the Requirements for the Degree of Master of Science
in the Faculty of Arts and Science

TRENT UNIVERSITY

Peterborough, Ontario, Canada

Applied Modelling and Quantitative Methods M.Sc. Graduate Program

January 2026

ABSTRACT

Multi-Task Learning for Humanitarian Demining Operations:

A Comparative Analysis of Perception Algorithms

Waun Broderick

This thesis presents a comprehensive investigation into machine learning approaches for landmine detection using thermal imagery. It addresses both classification and precise localization challenges that are integral for humanitarian demining operations. The research encompasses two complementary methodological frameworks: comparative evaluation of traditional machine learning versus deep learning approaches, then followed by an implementation of hyperparameter optimization for enhanced safety performance.

The foundation of the study demonstrates that traditional machine learning methods achieve competitive classification performance. Conventional models achieved significant performance with the Random Forest and RestNet50 respectively scoring accuracies of 91.88% and 94.29%, though struggle to achieve >10% when tasked with classification and localization. Expanding on this foundation, we addresses this gap through multi-task learning frameworks that simultaneously optimize for both detection and precise localization. Through systematic hyperparameter tuning across 64 configurations, the optimized multi-task approach achieves 90% detection accuracy with 92% precision while providing precise bounding box localization, representing a 37.5% reduction in false negatives. These findings demonstrate that while traditional machine learning offers computational efficiency for basic detection, multi-task deep learning frameworks provide significant performance gains when requiring precise spatial localization, which is an important requirement in demining operations.

Keywords: Landmines, demining, YOLO, humanitarian, Computer-Vision, Multi-Task Learning, Support Vector Machines, SVMs, Random Forest, CNN, Convolutional Neural Network, Thermal Imagery, Drone, Perception

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Sabine McConnell, for her guidance and support throughout this research project. My partner Jane for countless hours spent helping me label and segment data. My cat Peanut Butter for reminding me to take breaks along the way, and the entire Trent Computer Science Department for encouraging me to continue stoking my flame of passion and curiosity.

DEDICATION

To the individuals and coalitions working towards a safer, landmine free future.

Contents

ABSTRACT ii

ACKNOWLEDGEMENTS iii

DEDICATION iv

LIST OF FIGURES viii

LIST OF TABLES ix

GLOSSARY x

Chapter 1: INTRODUCTION 1

1.1 Introduction 2

Chapter 2: LITERATURE REVIEW 4

2.1 Historical Landmine Detection Methodologies 4

2.2 Drone-Based Systems for Landmine Detection 5

2.3 Methodologies for Subsurface Landmine Detection 5

2.4 Thermographic Imaging for Landmine Detection 6

2.5 Optical Imaging and Deep Learning Approaches 7

2.6 Multi-Task Learning for Enhanced Object Detection 8

2.7 Evolution of YOLO Architectures for Object Detection 9

2.8 Transfer Learning for Specialized Detection Tasks 10

2.9 Performance Metrics and Evaluation Standards 11

2.10 Data Augmentation and Dataset Limitations 12
2.11 Sensor Fusion and Multi-modal Approaches 13
2.12 Research Gap and Proposed Approach 14

Chapter 3: METHODOLOGY **16**

3.1 Introduction 16
3.2 The Dataset 17
3.3 Data Labeling 17
3.4 Data Balancing 18
3.4.1 Random Forest Classification 21
3.4.2 Support Vector Machine Classification 24
3.4.3 Transfer Learning 27
3.5 Multi-Task Learning Mechanisms 29
3.6 YOLOv8 Loss Function Components 34
3.7 Grid Search Methodology for Multi-Task Learning Optimization 36
3.8 Integration and Comparison Framework 40

Chapter 4: RESULTS **42**

4.1 Introduction 42
4.2 Phase 1 Results: Traditional Machine Learning Performance 43
4.2.1 Random Forest Classification Results 43
4.2.2 Support Vector Machine Classification Results 46
4.2.3 Artificial Neural Network Classification Results 47
4.2.4 Phase 1 Comparative Analysis 50
4.3 Phase 2 Results: YOLOv8 Optimization and Performance 53
4.3.1 Hyperparameter Optimization Results 54
4.3.2 Performance Trade-off Analysis 60
4.3.3 Optimal Configuration Validation 62

4.3.4 Detailed Performance Analysis 63
4.4 Comprehensive Performance Comparison 64
4.4.1 Cross-Phase Performance Analysis 64
4.4.2 Computational and Practical Considerations 66
4.4.3 Integration of Multi-modal Information 67
4.4.4 Operational Deployment Implications 68
4.4.5 Statistical Significance and Robustness 68
4.5 Key Findings and Performance Insights 69

Chapter 5: DISCUSSION **70**

5.1 Interpreting Hyperparameter Effects 70
5.2 Trade-off Optimization for Safety-Critical Applications 71
5.3 Phase 1 Comparative Analysis and Interpretations 72
5.4 Cross-Phase Performance Evolution and Implications 73
5.5 Computational and Practical Implementation Considerations 74
5.6 Multi-modal Information Integration Strategies 74
5.7 Operational Deployment Decision Framework 75
5.8 Statistical Robustness and Generalization Analysis 75
5.9 Algorithmic Performance Hierarchy and Selection Criteria 76
5.10 Critical Success Factors and Implementation Guidance 76
5.11 Practical Implications for Deployment 77

Chapter 6: CONCLUSION AND FUTURE WORK **79**

6.1 Conclusion 79
6.2 Future Work 81

BIBLIOGRAPHY **84**

List of Figures

3.1 Landmine Bounding Box visualization 18

3.2 Imitation Legbraker Landmine Image 20

3.3 Random Forest Visualization 22

3.4 SVM Application Visualization 26

3.5 Yolo v8 Architecture Visualization 32

4.1 ROC Curve Mappings 44

4.2 Aggregated Confusion Matrix 45

4.3 SVM ROC Curves 47

4.4 SVM Confusion matrix 48

4.5 ResNet50 ROC Curves 50

4.6 ResNet50 Confusion Matrix 51

4.7 False Negative effects with varied weights 56

4.8 False Negative Rate Response Curves 57

4.9 mAP Effects of Loss Weights 58

4.10 mAP Response Curves 59

4.11 Precision Response Curves 59

4.12 Recall Response Curves 60

4.13 False Negative Rate vs Precision Trade-off 61

4.14 YOLOv8 Confusion Matrices Comparison 65

List of Tables

3.1 Loss Weight Parameters Search Space 37

4.1 Cross-Validation Performance Metrics for Random Forest (%) 43

4.2 Cross-Validation Performance Metrics for SVM (%) 46

4.3 Cross-Validation Performance Metrics for ResNet50 CNN (%) 49

4.4 Performance Comparison of Optimized vs. Standard YOLOv8 Configuration 62

GLOSSARY

Accuracy

The proportion of correct predictions (both true positives and true negatives) among the total number of cases examined.

AdamW Optimizer

An adaptive learning rate optimization algorithm that combines the benefits of Adam optimization with weight decay regularization.

Artificial Neural Network (ANN)

A computational model inspired by biological neural networks, consisting of interconnected nodes that process information.

Area Under the Curve (AUC)

A performance measurement for classification problems that represents the degree of separability between classes.

Augmentation

The process of artificially expanding a dataset by creating modified versions of existing data through transformations like rotation, scaling, or flipping.

Backbone

The feature extraction component of a neural network architecture, typically a pre-trained convolutional neural network.

Batch Size

The number of training examples processed together in one forward/backward pass during neural network training.

Binary Cross-Entropy (BCE)

A loss function used for binary classification tasks that measures the difference between predicted probabilities and actual binary labels.

Bounding Box

A rectangular frame that defines the spatial boundaries of an object in an image, specified by coordinates and dimensions.

Class Weight

A parameter that adjusts the importance of different classes during training to handle class imbalance.

Complete Intersection over Union (CIoU)

An advanced metric for evaluating bounding box regression that considers overlap, center distance, and aspect ratio.

Confusion Matrix

A table used to evaluate classification performance by showing actual versus predicted classifications.

Convolutional Neural Network (CNN)

A deep learning architecture particularly effective for image processing tasks, using convolutional layers to detect features.

Cross-Validation

A statistical technique for evaluating model performance by dividing data into multiple folds and training/testing on different combinations.

CSPDarknet

A convolutional neural network architecture that uses Cross Stage Partial connections to improve gradient flow and reduce computational redundancy.

Demining

The process of detecting and removing landmines and other explosive ordnance from an area to make it safe for civilian use.

Distribution Focal Loss (DFL)

A loss function that models coordinate predictions as probability distributions over discretized values to improve localization precision.

Early Stopping

A regularization technique that stops training when validation performance stops improving to prevent overfitting.

Ensemble Learning

A machine learning approach that combines multiple models to improve overall performance and robustness.

Epoch

One complete pass through the entire training dataset during neural network training.

F1-Score

The harmonic mean of precision and recall, providing a single metric that balances both measures.

False Negative (FN)

A case where the model incorrectly predicts the absence of a landmine when one is actually present.

False Negative Rate

The proportion of actual positives that are incorrectly classified as negatives, calculated as FN/(FN+TP).

False Positive (FP)

A case where the model incorrectly predicts the presence of a landmine when none is actually present.

Feature Extraction

The process of identifying and extracting relevant information from raw data for use in machine learning models.

Grid Search

A systematic approach to hyperparameter optimization that tests all combinations of specified parameter values.

Humanitarian Demining

Mine clearance operations conducted to protect civilian populations and enable safe access to previously contaminated areas.

Hyperparameter

Configuration settings for machine learning algorithms that are set before training begins and control the learning process.

Intersection over Union (IoU)

A metric for evaluating object detection that measures the overlap between predicted and ground truth bounding boxes.

K-Fold Cross-Validation

A validation technique that divides data into k subsets, using k-1 for training and 1 for validation, repeated k times.

Landmine

An explosive device designed to be triggered by the presence, proximity, or contact of a person or vehicle.

Learning Rate

A hyperparameter that controls how much model weights are adjusted during training based on the calculated gradient.

Mean Average Precision (mAP)

A comprehensive metric for object detection that averages precision across different recall levels and IoU thresholds.

Multi-Task Learning

A machine learning approach where a model simultaneously learns multiple related tasks, sharing representations between them.

Object Detection

A computer vision task that involves identifying and locating objects within images by predicting both class labels and bounding box coordinates.

Overfitting

A modeling error where a model learns the training data too well, resulting in poor performance on new, unseen data.

PANet

Path Aggregation Network, a feature fusion architecture that enhances information flow between different levels of a neural network.

Pareto Frontier

The set of optimal solutions in multi-objective optimization where improving one objective requires worsening another.

Precision

The ratio of true positive predictions to all positive predictions, measuring the accuracy of positive classifications.

Radial Basis Function (RBF)

A kernel function used in Support Vector Machines that measures similarity based on Euclidean distance.

Random Forest

An ensemble learning method that constructs multiple decision trees and combines their predictions for improved accuracy and robustness.

Recall

The ratio of true positive predictions to all actual positives, measuring the model's ability to identify all positive cases.

Receiver Operating Characteristic (ROC)

A curve that plots true positive rate against false positive rate at various classification thresholds.

ResNet50

A 50-layer residual neural network architecture known for its skip connections that help address the vanishing gradient problem.

Residual Connection

A neural network component that allows information to skip layers, helping to train very deep networks effectively.

Spatial Localization

The ability to determine the precise location of objects within an image or physical space.

Specificity

The ratio of true negative predictions to all actual negatives, measuring the model's ability to correctly identify negative cases.

Support Vector Machine (SVM)

A supervised learning algorithm that finds an optimal hyperplane to separate different classes in feature space.

Thermal Imagery

Images captured using infrared radiation that detect temperature differences rather than visible light.

Transfer Learning

A machine learning technique where a model trained on one task is adapted for a related task, often improving performance with limited data.

True Negative (TN)

A case where the model correctly predicts the absence of a landmine when none is actually present.

True Positive (TP)

A case where the model correctly predicts the presence of a landmine when one is actually present.

YOLOv8

You Only Look Once version 8, a state-of-the-art real-time object detection algorithm that can simultaneously classify and locate objects in images.

Chapter 1

INTRODUCTION

1.1 Introduction

The use of unmanned explosive devices (landmines), have been a persistent issue in warfare for decades, despite widespread global bans. Landmines, cluster munitions, and improvised explosive devices (IEDs) have been deployed in direct conflicts, while unexploded ordnances and remnants of war continue to pose indirect threats, causing indiscriminate harm long after hostilities have ceased. Though bounds have been made, the problem is far from resolved and remains a critical concern worldwide.

The global impact of landmines and unexploded ordnance (UXO) represents a persistent humanitarian challenge despite international efforts to ban their use. According to the Landmine Monitor 2022 report, greater than 5,544 casualties from mines and explosive remnants of war were recorded in 2021 alone, with civilians accounting for 80% of all casualties. More recent data from the Landmine Monitor 2023 shows the continued threat these devices pose to civilian populations worldwide, with the International Campaign to Ban Landmines (ICBL) reporting that approximately 4,710 people were killed or injured by mines and explosive remnants of war in 2022. According to recent data from the United Nations Mine Action Service (UNMAS), over 60 countries and territories are still contaminated with landmines, affecting the lives and livelihoods of millions of people worldwide.

While significant progress has been made in detecting their presence, the continued identification and localization of buried explosives continues to be a challenge faced in efficient clearance operations. In that challenge, it is important to note that there is a fundamental difference between determining the possible presence of a landmine and precisely locating its whereabouts. This distinction is particularly important in areas with high landmine density or complex terrain, where conventional detection methods may indicate presence but fail to provide the spatial precision necessary for systematic clearance.

The integration of emerging technologies presents new opportunities to address these persistent challenges through automated systems that can operate safely and efficiently

in these unsafe areas. Recent technological advances in computer perception have also opened pathways for developing solutions that can bridge the gap between detection and the precise localization required for systematic clearance operations.

A potential solution would requires overcoming significant implementation challenges, including the need for dynamic data processing, reliable deployment systems, and integration of visual perception frameworks. Additionally, the cost of deploying advanced systems and the need for appropriate regulatory frameworks remain important considerations for widespread adoption. Addressing these constraints and requirements will be essential for effectively making advancements in working towards a landmine-free world.

This thesis explores this challenge through a two phased approach to. It begins with a foundational observation of traditional machine learning models like; Random Forest, SVM, and ResNet50 for basic classification. Phase 2 sees the utilization of more modern architectural frameworks like YOLOv8 paired with tools for hyperparameter tuning with high interoperability. Through the use of high quality thermal imagery and important textual metadata systematic optimizations can be derived that focus on minimizing false negatives. It is with this thesis that we hope to pave a path forward for real world applications by allowing high-fidelity communications about model tuning and a strong case for why data generation can improve vital generalizability of models like these. Chapter 2 reviews existing detection methodologies and establishes current research gaps, Chapter 3 details the experimental methodology across both phases, Chapter 4 presents comparative performance results, and Chapter 5 discusses practical implications for operational deployment. Through this comprehensive evaluation, the research demonstrates a 37.5% reduction in false negative rates compared to standard configurations while providing the spatial localization capabilities required for real-world demining operations, contributing valuable insights for technology-assisted humanitarian mine clearance efforts.

Chapter 2

LITERATURE REVIEW

2.1 Historical Landmine Detection Methodologies

Landmine detection has been a significant research focus for decades, with various methods including metal detectors, ground-penetrating radar, and infrared thermography being explored to address the persistent threat of explosive remnants of war. The practical implementation of landmine detection technologies presents unique challenges beyond theoretical advancements. The Catalogue of Advanced Technologies and Systems for Humanitarian Demining (Milan et al., 2005) details a comprehensive overview of operational demining technologies, including sensor-based detection systems in both person-operated and robotic delivery systems. Despite being published in 2005, this foundational resource remains valuable for understanding the technological landscape and implementation challenges in field deployment.

Traditional techniques have demonstrated limitations in terms of accuracy, speed, and safety, particularly when deployed in varied environmental conditions. Recent advances in artificial intelligence (AI) and machine learning have introduced new capabilities in feature extraction, classification, and automated threat assessment. One study investigates the potential of enhancing the effectiveness of conventional metal detectors using various machine learning methodologies. By pairing relevant sensor data alongside feature extraction through various models such as; support vector machines, neural networks, and

other classification algorithms, the study aims to propose a framework that can increase interoperability of models that to be employed for safety critical operations like landmine detection (Safatly et al., 2020).

2.2 Drone-Based Systems for Landmine Detection

Unmanned aerial vehicles (UAVs) have emerged as promising platforms for landmine detection operations due to their ability to operate without putting human lives at risk. Kowalski et al. conducted a comprehensive analysis of UAV systems for detecting explosive hazards, highlighting the significant advantages of aerial platforms in covering large areas efficiently while maintaining a safe distance from potential threats (Kowalski et al., 2020). Their research emphasized the importance of sensor integration and real-time processing capabilities in developing effective drone-based detection systems .

The deployment of UAVs offer particular advantages in post-conflict areas where terrain may be difficult to navigate on foot. Luna et al. observed that UAVs provide access to dangerous or inaccessible areas for human operators, which could potentially speed up demining operations while substantially reducing risks (Luna et al., 2020). Their system demonstrated the feasibility of integrating multiple sensing technologies on a single drone platform, creating a more comprehensive detection capability.

2.3 Methodologies for Subsurface Landmine Detection

The challenge of detecting buried landmines has led to the development of various methodologies with differing sensing modalities. Makki et al. examined multisensor approaches for landmine detection, comparing ground-penetrating radar (GPR), magnetometry, infrared thermography, and other techniques (Makki et al., 2019). Their analysis revealed that each modality has specific strengths and limitations depending on environmental conditions, mine composition, and burial depth.

One notable advancement in this field is the MAGnetometry Imagined based Classification System (MAGICS) approach described by Nikulin et al. (Nikulin et al., 2018). This system utilizes Convolutional Neural Networks (CNNs) to analyze magnetic anomaly data collected by UAVs, demonstrating significant potential for detecting metallic mines. The researchers reported detection rates of over 90% for specific types of metallic landmines under controlled conditions, highlighting the promise of machine learning approaches when applied to magnetometry data.

2.4 Thermographic Imaging for Landmine Detection

While magnetic anomaly detection shows promise for metallic mines, thermographic imaging offers distinct advantages, particularly for detecting both metallic and non-metallic explosive devices. A comprehensive meta-study conducted at Università di Roma (Santulli, 2007) analyzed multiple investigations and provided strong evidence supporting thermography as a viable technique for detecting buried objects in soil. This method leverages thermal variations in soil composition, which are influenced by environmental factors such as humidity and temperature, making it a promising approach for addressing the limitations of purely regular image-based detection systems.

Thermographic imaging excels in certain environments and conditions where other detection methods may struggle. In their comparison of various sensing modalities, Doe and Williams determined that infrared thermography performs especially well during daily temperature transitions when buried objects and surrounding soil exhibit maximum thermal contrast (Doe and Williams, 2021). Their research demonstrated that thermography can detect both surface and shallow-buried landmines with higher reliability than GPR in dry, sandy soils where radar signals may experience significant degrading in their efficacy.

The operational advantages of thermographic imaging extend beyond technical performance metrics. Based on Milan et al.'s comprehensive catalogue of demining technologies,

thermographic systems operate effectively from safe distances, without ground contact, and can quickly scan extensive areas, making them well-suited for UAV platform integration (Milan et al., 2005). These practical considerations further strengthen the case for thermographic approaches in humanitarian demining operations.

2.5 Optical Imaging and Deep Learning Approaches

With recent advancements in processing power and the development of convolutional neural networks (CNNs) for image analysis, object detection through image recognition has become a more viable and effective alternative. One notable study explores a wide range of object detection methodologies and their effectiveness in analyzing diverse datasets to enhance detection accuracy in novel applications (Wang et al., 2025). Many of these approaches have significantly contributed to the advancement and widespread adoption of object detection across various fields.

In recent years, significant advancements have been made in applying deep learning techniques to landmine detection challenges. Safatly et al. demonstrated the effectiveness of machine learning algorithms applied to metal detector data, achieving classification accuracies of over 85% for differentiating between landmines and metallic debris (Safatly et al., 2020). Their approach, while limited to metallic objects, established the viability of AI-based discrimination for reducing false positive rates in demining operations.

Many different types of land-borne IEDs can be deployed both on-terrain and subterraneanly, which requires a variety of different optical imaging methodologies. One study addressed this challenge by focusing on the detection of common surface landmines, specifically the PFM-1 (butterfly) and PMA-2 (starfish with tripwire) types (Vivoli et al., 2024). Their system, leveraging advanced image processing and neural networks, achieved high recall rates in real-time detection scenarios, demonstrating the potential of AI-driven computer vision approaches for landmine detection. However, this method does not ac-

count for the significant number of buried landmines, which remain a critical challenge in demining operations.

Moving specifically to thermographic imaging analysis, Mohamed et al. developed a deep learning-based thermal image processing approach for detecting buried objects and mines (Mohamed et al., 2021). Their research utilized Region-based Convolutional Neural Networks (R-CNNs) to analyze thermal imagery, reporting promising results in identifying thermal signatures consistent with buried explosive devices. The authors noted that deep learning models can identify subtle thermal patterns potentially invisible to human operators, which may increase detection reliability while reducing false alarms.

Further advancing the application of deep learning in this domain, Vivoli et al. implemented a real-time detection system for surface landmines using optical imaging (Vivoli et al., 2024). Their approach employed YOLOv5, demonstrating the efficacy of the YOLO (You Only Look Once) architecture family for object detection in the context of explosives detection. The authors found that their system achieved real-time processing capabilities with detection accuracies exceeding 90% (under favorable lighting conditions), establishing an important benchmark for detection speed and accuracy in above-ground field deployments.

2.6 Multi-Task Learning for Enhanced Object Detection

Multi-task learning (MTL) represents a paradigm where a single neural network simultaneously learns multiple related tasks, leveraging shared representations to improve performance across all objectives. In the context of object detection, MTL approaches can simultaneously optimize for classification, localization, and auxiliary tasks such as depth estimation or semantic segmentation. Zhang and Yang provided a comprehensive survey of multi-task learning methodologies, demonstrating that shared feature representations learned across related tasks often lead to improved generalization performance compared

to single-task approaches, particularly when training data is limited (Zhang and Yang, 2017).

The application of multi-task learning to object detection has shown particular promise in scenarios where multiple types of information must be extracted from the same input. Ruder observed that MTL can act as a form of regularization, reducing overfitting by encouraging the model to learn more general representations that are useful across multiple tasks (Ruder, 2017). This characteristic is especially valuable in landmine detection applications where the model must not only identify the presence of explosive devices but also provide precise localization information while potentially distinguishing between different types of threats.

Recent developments in multi-task learning for computer vision have demonstrated the effectiveness of hard parameter sharing architectures, where early layers of the network are shared across tasks while task-specific layers handle specialized processing. Kendall et al. showed that careful weighting of multiple loss functions in MTL frameworks can significantly improve performance on individual tasks compared to single-task baselines, establishing important principles for designing effective multi-task detection systems in safety-critical applications such as humanitarian demining (Kendall et al., 2018).

2.7 Evolution of YOLO Architectures for Object Detection

The YOLO family of object detection architectures has evolved significantly since its introduction, with each iteration bringing improvements in accuracy, speed, and feature extraction capabilities. Redmon and Farhadi introduced YOLOv3, which incorporated residual networks and feature pyramid networks to improve detection accuracy across objects of varying scales (Redmon and Farhadi, 2018). This architecture demonstrated particular strength in detecting small objects, a critical capability for landmine detection applications

where targets may occupy a small portion of the image frame.

Wang et al. further advanced the YOLO architecture with YOLOv7, which introduced trainable bag-of-freebies techniques that substantially improved model performance without increasing inference costs (Wang et al., 2022). Their architecture achieved state-of-the-art results on the MS COCO dataset (Lin et al., 2014), establishing new benchmarks for real-time object detection performance.

The latest evolution in this architectural family, YOLOv8, builds upon these advances with significant improvements in backbone design, loss function optimization, and hyperparameter sensitivity. According to Jocher et al. , YOLOv8 features a more efficient backbone, improved path aggregation neck, and anchor-free detection heads that enhance both accuracy and inference speed compared to previous versions (Jocher et al., 2023). These architectural improvements, combined with advanced data augmentation techniques, position YOLOv8 as a particularly promising candidate for landmine localization applications where both precision and computational efficiency are critical requirements.

2.8 Transfer Learning for Specialized Detection Tasks

Given the scarcity of large-scale labelled datasets in the landmine detection domain, transfer learning approaches offer substantial advantages by leveraging knowledge gained from models pre-trained on large general-purpose datasets. Gao et al. found that models previously trained on natural image datasets could be effectively adapted to thermal imagery classification tasks using relatively small amounts of domain-specific training data (Gao et al., 2018).

Liu et al. further explored the application of transfer learning techniques to infrared image object detection, concluding that kernel-based transfer learning methods can effectively connect the domains of visible and infrared imagery, allowing more efficient model training for specialized detection tasks (Liu et al., 2020). Their research highlighted the

importance of careful feature extraction and adaptation when transferring learned representations between different imaging modalities.

The application of transfer learning to explosive hazard detection has been further explored by Zhang et al. in their comprehensive survey of multi-modal deep learning approaches (Zhang et al., 2023). The researchers found that transfer learning provides more robust feature extraction by incorporating knowledge from diverse datasets, potentially improving adaptation to the varied environmental conditions encountered in real-world demining operations. This observation is particularly relevant for landmine detection systems that must operate reliably across different soil types, moisture levels, and diurnal temperature variations.

2.9 Performance Metrics and Evaluation Standards

Lekhak et al. demonstrated through field testing that false negative rates (missed landmines) carry significantly higher operational and safety costs than false positive rates (Lekhak et al., 2024). Their comparative analysis of handheld versus airborne metal detection systems achieved detection accuracies of 95.79% while emphasizing the critical importance of minimizing missed detections in humanitarian demining operations.

The International Mine Action Standards (IMAS) provide comprehensive technical and operational guidelines for mine action activities, including specific requirements for detection system performance. According to IMAS 09.42 on metal detector operations, detection systems must achieve a minimum probability of detection of 95% for all target mines under specified conditions, while maintaining false alarm rates below acceptable thresholds defined by operational requirements (Service, 2013). These standards establish baseline performance requirements that any automated detection system must meet for operational deployment.

Benchmarking methodologies for landmine detection systems have evolved to address

the unique challenges of comparing performance across different environmental conditions and target types. Geneva International Centre for Humanitarian Demining (GICHD) (for Humanitarian Demining, 2020) developed standardized testing protocols that evaluate detection systems under controlled and field conditions, incorporating variables such as soil composition, burial depth, and environmental factors. Their framework emphasizes the importance of statistical significance in performance evaluation, requiring multiple test iterations across diverse conditions to establish reliable performance baselines for operational decision-making.

2.10 Data Augmentation and Dataset Limitations

The development of robust machine learning models for landmine detection faces significant challenges due to the inherent scarcity and sensitive nature of landmine imagery datasets. Unlike conventional object detection applications where large datasets are readily available, landmine detection research is constrained by safety, legal, and ethical considerations that limit data collection opportunities.Vivoli et al. addressed this challenge by developing a dataset called "SurfLandmine" containing 47 videos with 6,640 annotated frames. With their dataset, they were able to demonstrate that through carefully designed data collection and augmentation strategies, they can significantly improve model performance. Despite limited training data, they were able to achieve 98.4% detection rates for existing objects in their test set (Vivoli et al., 2024).

The creation of large-scale landmine datasets presents unique logistical and safety challenges that distinguish this domain from other computer vision applications. Real-world data collection requires specialized expertise, safety protocols, and often involves coordination with military or humanitarian organizations. As documented by the humanitarian demining community, the controlled detonation or handling of actual explosive devices for research purposes is highly regulated and often impractical, leading researchers to rely on

inert replicas, historical datasets, or simulated environments that may not fully capture the complexity of operational conditions.

Ethical considerations in landmine data collection extend beyond safety concerns to include broader humanitarian and research ethics principles. The use of real landmine imagery raises questions about data sensitivity, potential misuse, and the responsibility of researchers to ensure their work contributes positively to humanitarian demining efforts. As outlined by the International Committee of the Red Cross (ICRC), research involving explosive hazards must balance scientific advancement with humanitarian principles, ensuring that technological developments ultimately serve to reduce rather than increase the risks associated with explosive remnants of war (of the Red Cross, 1997).

2.11 Sensor Fusion and Multi-modal Approaches

The integration of multiple sensing modalities represents a promising approach for enhancing landmine detection performance by leveraging the complementary strengths of different sensor types. Ground-penetrating radar (GPR) excels at detecting subsurface anomalies regardless of material composition, while electromagnetic induction provides high sensitivity for metallic objects, and thermal imagery offers advantages for detecting thermal anomalies caused by buried objects. Marsh et al. provided a comprehensive study of dual-modality landmine detection, demonstrating that integrating spectroscopic metal detection with ground-penetrating radar can spatially locate metallic clutter items and minimum-metal mine surrogates to within 20 mm of their known positions, with each technique offering distinct operational advantages that can be addressed through systematic sensor fusion approaches (Marsh et al., 2019).

Information fusion frameworks for landmine detection typically operate at three levels: data-level fusion, feature-level fusion, and decision-level fusion. Data-level fusion combines raw sensor measurements before processing, offering the highest information content

but requiring synchronized data acquisition and compatible sensor characteristics. Marsh et al. demonstrated through controlled testing that feature-level fusion of magnetic induction spectroscopy (MIS) and GPR sensors, using confidence values from both modalities, could improve detection performance by enabling all four buried test objects to be successfully identified when depth information was incorporated as an additional feature, compared to individual sensors operating alone (Marsh et al., 2019).

Feature-level fusion approaches extract characteristic features from each sensor modality before combining them in a unified feature space for classification or detection. This approach offers greater flexibility in handling sensors with different sampling rates or spatial resolutions. The feature-level sensor fusion strategy demonstrated by Marsh et al. involved extracting confidence values from both MIS and GPR systems, then using adaptive thresholds based on object depth to optimize detection performance (Marsh et al., 2019). Their research showed that when one sensor reports very high confidence, it is not necessary to require a high degree of confidence from the other sensor, leading to improved overall detection rates.

2.12 Research Gap and Proposed Approach

While previous studies have demonstrated the potential of deep learning for landmine detection using various imaging modalities, several key research gaps remain. First, most existing approaches have focused on detection (determining presence) rather than precise localization (determining exact coordinates), which is critical for efficient clearance operations. Second, the application of transfer learning and hyperparameter optimization specifically for thermographic landmine imagery remains relatively unexplored, particularly with the latest YOLO architectures.

The current research addresses these gaps by employing transfer learning with the state-of-the-art YOLOv8 architecture for precise landmine localization using thermographic im-

agery. By leveraging pre-trained weights from large general-purpose datasets and systematically optimizing hyperparameters for the specific characteristics of thermal landmine signatures, this approach aims to achieve centimetre-level localization precision while maintaining computational efficiency suitable for deployment on UAV platforms with limited processing resources.

Furthermore, this research builds upon the thermal imagery dataset established by Białek et al. (Białek et al., 2023), which provides standardized thermal images of landmines and landmine-like objects in controlled environments. The availability of this specialized dataset enables more rigorous validation of the proposed approach compared to previous studies that relied on simulated data or limited field samples.

In summary, this literature review demonstrates the evolution of landmine detection methodologies from traditional sensor approaches to advanced AI-driven systems, establishing the foundation for the current research on transfer learning and hyperparameter optimization with YOLOv8 for precise landmine localization using thermographic imagery. The proposed approach represents a logical advancement of existing techniques, addressing critical research gaps while building upon established successes in related application domains.

Chapter 3

METHODOLOGY

3.1 Introduction

This research can be conceptually segmented into 2 phases designed to systematically evaluate and compare different algorithmic approaches. The first phase establishes baseline performance using traditional machine learning methods, while the second phase explores advanced deep learning architectures. This sequential methodology enables a comprehensive understanding of the problem domain while developing increasingly sophisticated solutions optimized for the safety-critical requirements of demining operations.

The initial phase focuses on Random Forest and Support Vector Machine (SVM) classifiers. This phase provides essential insights into the fundamental challenges of landmine detection and serves as a foundation for understanding the relative advantages of more advanced approaches.

The second phase leverages deep learning architectures, incorporating transfer learning techniques and multi-task learning. Transfer learning enables the adaptation of pre-trained models to the specific domain of landmine detection, capitalizing on learned representations from large-scale datasets while reducing computational requirements and training time. Multi-task learnings are employed to enhance the model's ability to focus on relevant spatial features within the input data, potentially improving both identification accuracy and localization precision.

Within both phases, we address two correlated yet analytically distinct tasks: identification and localization. During the evaluation of these methodologies, we identified that achieving optimal performance on both tasks simultaneously presents a significant challenge.

3.2 The Dataset

The dataset used in this study is comprised of 2,700 thermographic images of terrain containing buried antipersonnel landmines, acquired at Universidad del Valle (Cali, Colombia) using a Zenmuse XT infrared camera (7-13 μm spectral range) mounted on a DJI Matrice 100 drone Tenorio-Tamayo et al. (2023). Images were captured at altitudes ranging from 1 to 10 meters in 1-meter intervals over a controlled 10m × 10m test site. The experimental site featured nine marked zones with imitation landmines buried at depths of 0, 1, 5, and 10 cm, plus one mine-free control zone. Each imitation mine consisted of a PVC cylinder (8.7 cm diameter, 10.2 cm height) filled with anthracite coal to simulate TNT thermal properties, with thermal characteristics closely matching actual antipersonnel mines (specific heat difference: 0.11 J/g·K; thermal conductivity difference: 0.03 W/m·K). Following established protocols, images were acquired on cloudless days between 17:00-18:00 hours under optimal thermal contrast conditions, with wind speeds of 3-5 m/s and ambient temperatures of 26-30°C. The terrain composition included non-saline soil (pH 5.73) with 35.6% sand, 35.4% clay, and 29% silt, selected to maximize thermal signature detection while minimizing environmental interference.

3.3 Data Labeling

Each image was annotated with it's respective metadata including temperature, depth, and altitude measurements as shown in Fig. 3.1. For images containing landmines, bounding boxes were drawn to closely follow the contours of the detected objects. These bounding

box annotations precisely delineated the landmine locations from their surrounding areas.

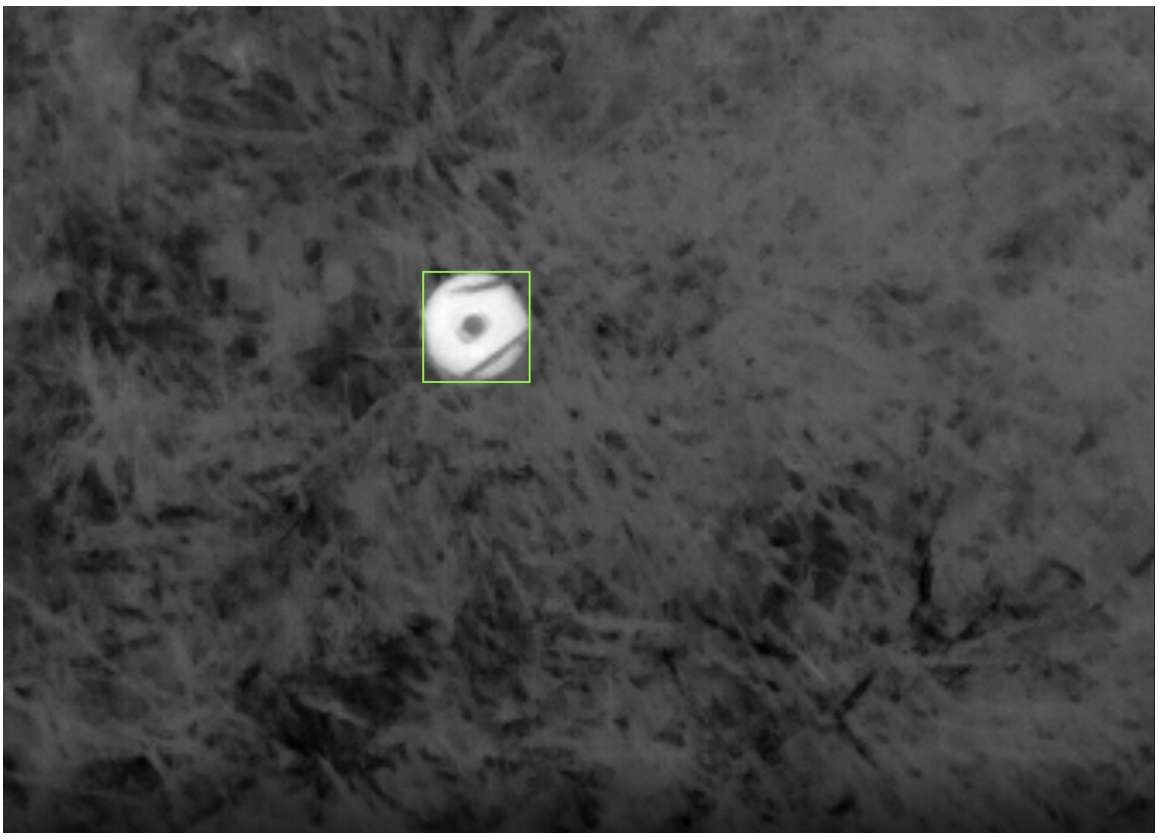

Figure 3.1: Landmine And Its created Bounding Box (Tenorio-Tamayo et al., 2023)

3.4 Data Balancing

The dataset being utilized contained an inherent imbalance, the images with landmines significantly outnumbered those without, necessitating a comprehensive data balancing strategy through controlled image augmentation. This process aimed to achieve two key objectives: preserving the integrity of landmine signatures while ensuring a balanced dataset for unbiased training.

For images containing landmines (I_m) and those without (I_n), we calculated appropriate augmentation multipliers (M_m, M_n) using the ratio between classes:

$$r = \frac{|I_m|}{|I_n|} \tag{3.1}$$

where $|I_m|$ and $|I_n|$ represent the number of images with and without landmines, respectively. The multipliers were then determined as:

$$M_m = \max(1, T - 1) \tag{3.2}$$

$$M_n = \max(1, \lceil T \cdot r \rceil - 1) \tag{3.3}$$

The target multiplier T was determined to achieve approximate class balance between images containing landmines (Im) and those without (In). The original dataset of 1,116 images contained an inherent imbalance where 846 images contained landmines while only 270 images had no landmines present, resulting in a class ratio r = |Im|/|In| = 846/270 = 3.13. This significant imbalance, with landmine images outnumbering non-landmine images by more than 3:1, necessitated a comprehensive data balancing strategy.

Data Splitting and Usage All experiments were conducted using a balanced dataset obtained through the augmentation process described above. The dataset was split into training and testing sets using a 70/30 ratio, following industry-standard practices for machine learning evaluation. Additionally, 5-fold cross-validation was employed to ensure robust and unbiased performance assessment across different data partitions.

This approach ensures that the models are trained and evaluated on representative and balanced data, providing reliable insights into their detection capabilities. For images containing landmines, we implemented careful augmentation that preserved bounding box integrity. The augmentation process included:

Horizontal flips along the vertical axis (y-axis) and vertical flips along the horizontal axis (x-axis), each applied with a 50% probability.

The decision to use minimal transformations was made to ensure the integrity of iden-

tification of the ’legbreaker landmines specific characteristics, see Fig. 3.2. At a certain altitude and depth these landmines will have a consistent size and signature that if distorted can cause issues with detecting the specific characteristics of the landmine. In a system that would be tasked to identify a myriad of different ordinates, the ability to distinguish between one another can be paramount. Regardless of this, we employed several fail-safes to ensure that our transformations did not distort of dataset beyond our intended limts.

Figure 3.2: Imitation Legbreaker landmine (Tenorio-Tamayo et al., 2023)

To ensure bounding box validity post-augmentation, we applied strict validation criteria:

Area ratio constraint: $0.8 \leq \frac{A_{\text{new}}}{A_{\text{orig}}} \leq 1.2$

Coordinate normalization: $\{x, y, w, h\} \in [0, 1]$

Spatial validity: $x + w \leq 1$ and $y + h \leq 1$

where A_{new} and A_{orig} represent the areas of the augmented and original bounding boxes, respectively, and (x, y, w, h) are the normalized coordinates and dimensions of the bounding boxes.

To maintain data quality, we employed the Interquartile Range (IQR) method for outlier detection in bounding box areas:

$$\text{IQR} = Q_3 - Q_1 \tag{3.4}$$

$$[L_{\text{bound}}, U_{\text{bound}}] = [Q_1 - 1.5 \cdot \text{IQR}, Q_3 + 1.5 \cdot \text{IQR}] \tag{3.5}$$

where Q_1 and Q_3 represent the first and third quartiles, respectively, and $[L_{\text{bound}}, U_{\text{bound}}]$ defines the acceptable range for bounding box areas.

The final balanced dataset maintains the original data integrity while providing sufficient sample diversity for robust model training. All augmented images were verified to preserve the essential characteristics of landmine signatures, ensuring the augmentation process did not introduce outliers that could compromise the model's learning objectives.

3.4.1 Random Forest Classification

A Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The algorithm creates a forest of uncorrelated trees by using bootstrap aggregating (bagging) and feature randomness. For each tree in the forest, a bootstrap sample is taken from the training data, and at each node of the tree, a random subset of features is considered for splitting. This approach reduces the variance of the model without increasing bias, effectively mitigating overfitting

issues commonly found in single decision trees. Random Forests are particularly effective when dealing with high-dimensional data, can handle both categorical and continuous variables, and automatically rank feature importance (Breiman, 2001).

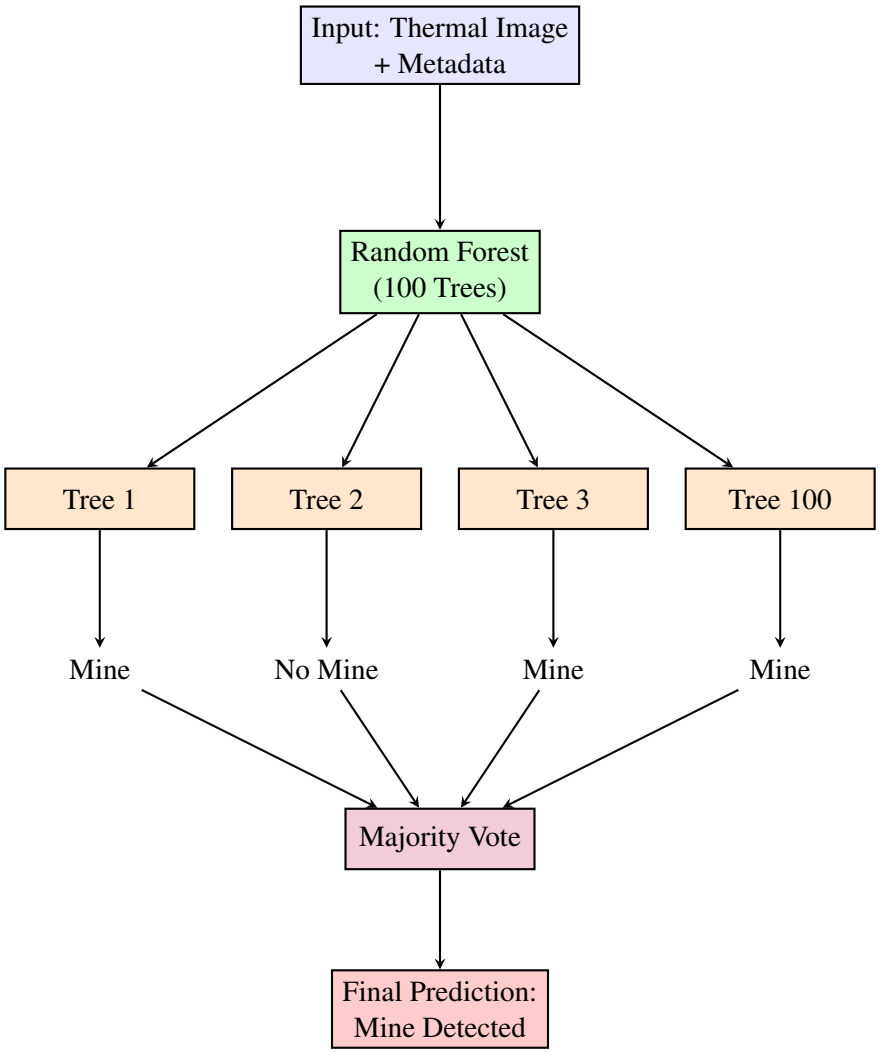

Figure 3.3: Random Forest ensemble methodology for landmine detection showing the aggregation of multiple decision trees to produce a final classification.

Figure 3.3 illustrates the Random Forest ensemble approach implemented for landmine detection. Each tree in the forest is trained on a different bootstrap sample of the data and considers only a random subset of features at each split, creating diversity among the individual classifiers. This diversity is crucial for landmine detection as it enables the model to capture different aspects of the thermal signatures and environmental conditions

that may indicate the presence of landmines.

Random Forest is particularly well-suited for this landmine detection study for several key reasons. First, the algorithm excels at handling high-dimensional data, which is essential given our input vector of 268,206 features (299×299×3 flattened pixels plus 3 metadata features). The random feature selection at each split prevents overfitting to specific pixel patterns while still capturing meaningful thermal signatures. Second, Random Forest naturally handles the class imbalance present in our dataset through balanced class weights, ensuring that the model does not bias toward the more frequent class. Third, the ensemble nature provides robust uncertainty quantification through vote distributions, which is critical in safety-critical applications where understanding prediction confidence can inform operational decisions. Finally, the algorithm's inherent resistance to overfitting and ability to provide feature importance rankings make it an ideal baseline method for understanding which thermal and environmental features are most indicative of landmine presence, providing valuable insights for subsequent deep learning approaches.

Implementation The balanced dataset was used to train a Random Forest classifier for landmine detection. Each input sample consisted of a flattened RGB image (299×299×3) concatenated with three additional features: depth, altitude, and temperature, resulting in an input vector $x_i \in \mathbb{R}^{268,206}$ where:

$$x_i \in \mathbb{R}^{268,206} \text{ where } x_i = [I_{\text{flat}}, d, a, t] \tag{3.6}$$

where I_{flat} represents the flattened image pixels, and d, a, and t represent depth, altitude, and temperature respectively.

The model was evaluated using 5-fold cross-validation, with 70:30 training to testing splits throughout all references to k-fold, to ensure robust performance assessment. For each fold k, the dataset was split into training ($\mathcal{D}^k_{\text{train}}$) and validation ($\mathcal{D}^k_{\text{val}}$) sets. The Random Forest classifier was configured with 100 estimators and balanced class weights to

handle any remaining class imbalance:

$$w_c = \frac{n_{\text{samples}}}{n_{\text{classes}} \cdot n_{\text{samples}_c}} \tag{3.7}$$

Performance metrics were calculated for each fold:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.9}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3.10}$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.11}$$

where TP, TN, FP, and FN represent True Positives, True Negatives, False Positives, and False Negatives respectively.

3.4.2 Support Vector Machine Classification

A Support Vector Machine (SVM) is supervised learning algorithm used for classification and regression tasks. The fundamental principle of SVM is to find an optimal hyperplane that maximally separates different classes in the feature space. For linearly separable data, SVM finds the hyperplane with the maximum margin between the closest points (support vectors) from each class. For non-linearly separable data, SVM employs the "kernel trick" to implicitly map input data into higher-dimensional spaces where linear separation becomes possible. Common kernel functions include linear, polynomial, and Radial Basis Function (RBF). The RBF kernel, also known as the Gaussian kernel, is particularly effective for complex classification tasks as it measures the similarity between points based on the Euclidean distance, allowing the algorithm to capture non-linear relationships in the data (Cortes and Vapnik, 1995). SVMs are known for their effectiveness in high-

dimensional spaces, memory efficiency, and robustness when the number of dimensions exceeds the number of samples.

Implementation The balanced dataset was similarly utilized to train a Support Vector Machine (SVM) classifier. The input vector structure remained consistent with the Random Forest implementation, where each sample comprised a flattened RGB image concatenated with depth, altitude, and temperature features:

$$x_i = [I_{\text{flat}}, d, a, t] \tag{3.12}$$

The SVM classifier was configured with a Radial Basis Function (RBF) kernel, which has demonstrated superior performance in thermal image analysis and object detection tasks (Gao et al., 2018; Liu et al., 2020). The RBF kernel's effectiveness in handling the non-linear relationships present in thermographic data makes it particularly suitable for binary object detection in thermal imagery, with previous studies showing its robustness in processing infrared data for defect detection (Gao et al., 2018). This kernel choice is further supported by comprehensive reviews of kernel methods in infrared image processing, which highlight RBF's particular strengths in handling the complex feature relationships present in thermal imagery (Liu et al., 2020).

The comprehensive visualization in Figure 3.4 illustrates the complete SVM methodology progression from basic linear classification to advanced RBF kernel implementation.

Panel (a) demonstrates the fundamental linear SVM concept with linearly separable data, clearly showing the optimal hyperplane (solid black line), margin boundaries (dashed gray lines), and support vectors (red circles) that define the maximum-margin decision boundary.

Panel (b) exposes the critical limitations of linear SVM when applied to non-linearly separable thermal data, where the linear decision boundary (red line) fails to adequately separate complex thermal signatures from background patterns—a common challenge in

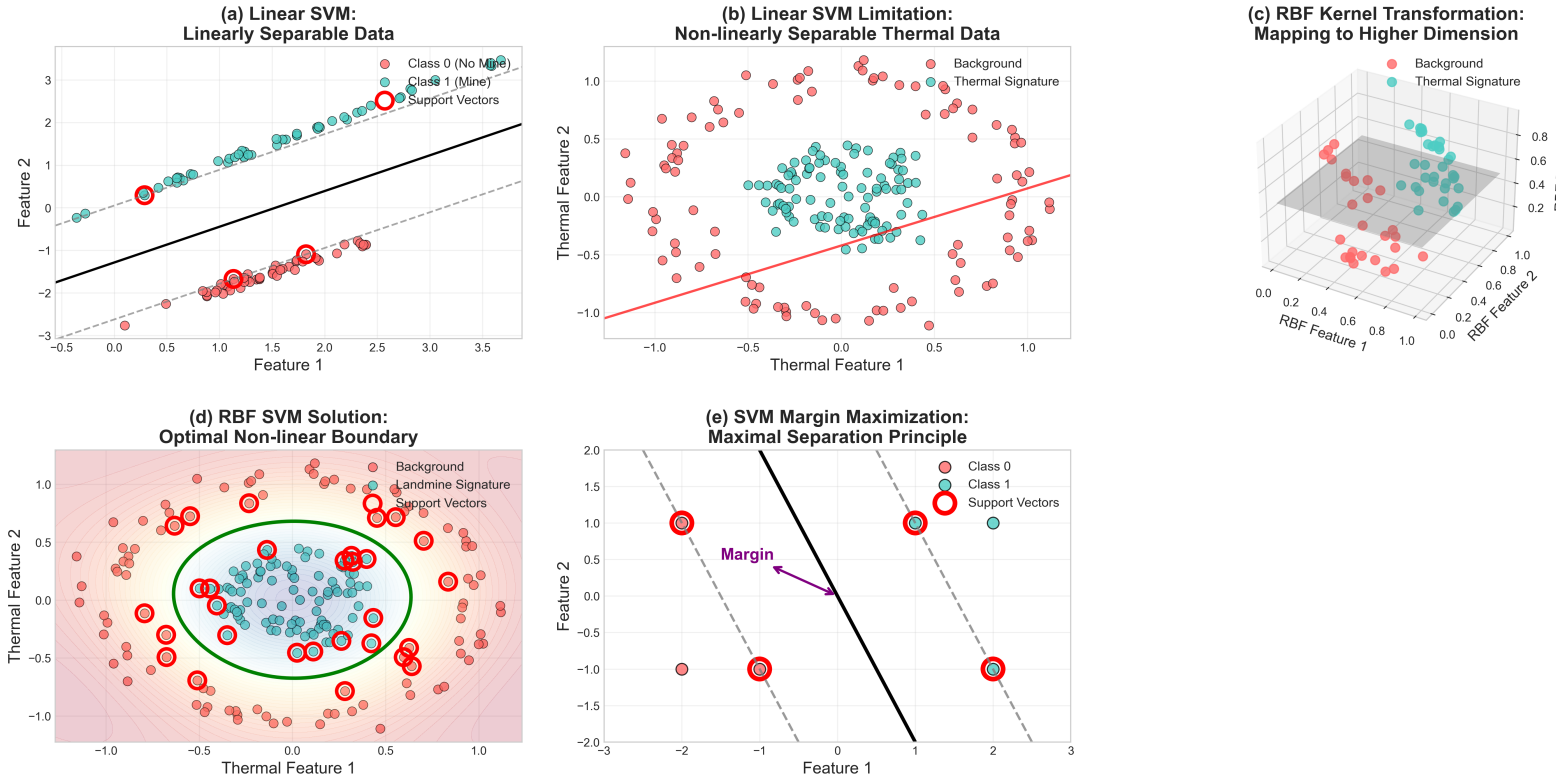

Figure 3.4: Support Vector Machine (SVM) methodology visualization: (a) Linear SVM with linearly separable data, (b) Linear SVM limitations with non-linearly separable thermal data, (c) RBF kernel transformation to higher-dimensional space, (d) RBF SVM solution with optimal non-linear boundary, and (e) SVM margin maximization principle.

thermal landmine detection.

Panel (c) visualizes the RBF kernel transformation concept through a 3D representation, illustrating how the kernel function maps non-linearly separable 2D thermal features into a higher-dimensional space where linear separation becomes mathematically feasible. This transformation is crucial for handling the complex, non-linear thermal signatures characteristic of buried landmines.

Panel (d) presents the RBF SVM solution applied to the same complex thermal data from panel (b), showcasing the optimal non-linear decision boundary (green contours) that successfully separates landmine thermal signatures from background noise, with support vectors (red circles) defining the decision boundary.

Panel (e) provides a detailed illustration of the SVM margin maximization principle, highlighting the maximum-margin hyperplane and the critical support vectors that determine the optimal separating boundary—a key advantage of SVM algorithms in maintaining robust classification performance. This comprehensive visualization demonstrates the theoretical foundation and practical implementation of RBF kernel selection for thermal landmine detection applications.

The model evaluation followed the same 5-fold cross-validation strategy used in the Random Forest implementation, with the dataset split into training ($\mathcal{D}^k_{\text{train}}$) and validation ($\mathcal{D}^k_{\text{val}}$) sets for each fold k. Performance metrics including precision, recall, specificity, and F1-score were calculated using identical formulations to ensure direct comparability with the Random Forest results.

3.4.3 Transfer Learning

Transfer learning involves repurposing a pre-trained model (typically trained on large-scale datasets like ImageNet) for a new task. This approach allows leveraging knowledge gained from solving one problem to improve generalization in another, particularly valuable when the target task has limited training data such as ours. In practice, transfer learning often

involves freezing early layers of the pre-trained model (which capture general features) while fine-tuning later layers to adapt to the specific requirements of the new task.

Pre-trained models like ResNet, which introduced residual connections to address the vanishing gradient problem in deep networks, have become standard backbones for transfer learning due to their robust feature extraction capabilities and generalizability across diverse computer vision tasks (He et al., 2016).

Transfer Learning Implementation for Landmine Detection

We implemented a deep learning solution using a modified ResNet50 architecture for landmine detection. Transfer learning with ResNet50 has shown particular success in object detection tasks (Vivoli et al., 2024), making it a suitable choice for landmine detection. The network was adapted by incorporating our additional sensor data (depth, altitude, and temperature) alongside the thermal imagery, building on established multi-modal deep learning approaches (Zhang et al., 2023).

Architecture Design The network architecture includes two fully connected layers with dropout for regularization, following the feature extraction from ResNet50. The addition of metadata features through parallel input branches allows the network to leverage both spatial and sensor information for improved detection performance.

$$f(x) = \sigma(W_2\delta(W_1[\text{ResNet50}(I_{\text{img}}), d, a, t] + b_1) + b_2) \tag{3.13}$$

where ResNet50(I_{img}) represents the feature extraction from the pre-trained backbone, $[\cdot, \cdot]$ denotes concatenation, δ is the ReLU activation function, and σ is the sigmoid activation for binary classification.

The network architecture includes:

- A ResNet50 backbone (pretrained on ImageNet) for image feature extraction
- Additional input branches for depth, altitude, and temperature data

- Two fully connected layers (1024 and 512 units) with dropout (0.5)
- Binary classification output with sigmoid activation

Training Strategy Training was performed using the Adam optimizer with a learning rate of 1×10^{-4} and binary cross-entropy loss. To prevent overfitting, we employed early stopping with a patience of 3 epochs, monitoring validation loss. The final 20 layers of the ResNet50 backbone were made trainable while earlier layers remained frozen to preserve learned low-level features while allowing adaptation to our specific task.

The same 5-fold cross-validation strategy employed in the traditional machine learning approaches was utilized to ensure consistent evaluation methodology. Performance metrics including precision, recall, specificity, and F1-score were calculated using identical formulations to enable direct comparison with the Random Forest and SVM implementations.

3.5 Multi-Task Learning Mechanisms

Multi-task learning enables models to simultaneously optimize multiple related objectives through shared representations and specialized task-specific heads. This approach leverages the principle that related tasks can benefit from shared feature learning, where lower-level representations capture common patterns while higher-level task-specific heads focus on distinct objectives. In object detection architectures like YOLOv8, multi-task learning is implemented through multiple prediction heads that jointly optimize classification, localization, and objectness scores, thereby improving overall detection performance through shared feature extraction and coordinated loss optimization.

The core principle of multi-task learning lies in its ability to exploit task relatedness through shared representations while maintaining task-specific specialization. In the context of computer vision tasks such as landmine detection, this capability is particularly valuable as it enables the model to simultaneously learn object classification, bounding

box regression, and confidence estimation, leading to more robust and accurate detection performance through coordinated optimization of multiple complementary objectives.

The mathematical foundation of multi-task learning involves optimizing a combined loss function that incorporates multiple task-specific objectives. For YOLOv8's multi-task architecture, the total loss function combines classification loss (L_{cls}), box regression loss (L_{box}), and distribution focal loss (L_{dfl}):

$$L_{total} = \lambda_{cls} L_{cls} + \lambda_{box} L_{box} + \lambda_{dfl} L_{dfl} \tag{3.14}$$

where λ_{cls}, λ_{box}, and λ_{dfl} are weighting parameters that balance the contribution of each task. The multi-task heads process shared backbone features through task-specific layers:

$$\text{Head}_i(x) = f_i(F_{shared}(x)) \tag{3.15}$$

where $F_{shared}(x)$ represents the shared feature extraction and f_i represents the task-specific transformation for head i.

Building upon the insights gained from traditional machine learning approaches and transfer learning implementations, the second phase of this research focused on implementing and optimizing state-of-the-art object detection capabilities using YOLOv8's multi-task learning framework. This transition from classification-based approaches to multi-task object detection represents a fundamental shift in methodology, moving from binary classification of entire images to precise localization and detection of landmines within complex scenes while leveraging multiple specialized heads to jointly optimize detection, classification, and localization objectives.

The integration of multi-task learning mechanisms within the YOLOv8 framework addresses key limitations observed in previous approaches, particularly regarding the model's ability to simultaneously optimize multiple detection objectives while maintaining computational efficiency. The YOLOv8 framework was selected for its superior real-time per-

formance capabilities and proven effectiveness in multi-task object detection, making it particularly suitable for the demanding requirements of operational demining scenarios where accurate classification, precise localization, and reliable confidence estimation are all critical for identification and localization success.

The decision to implement multi-task YOLOv8 was informed by the limitations observed in the traditional approaches, particularly regarding the need for simultaneous optimization of multiple detection objectives and the ability to distinguish subtle thermal signatures from environmental clutter through coordinated learning. While the Phase 1 methods provided valuable baseline performance metrics, the need for precise bounding box predictions, improved classification accuracy, reliable confidence estimation, and in a real world implementation, fast inference times necessitated the adoption of modern multi-task object detection architectures. This phase not only implements YOLOv8's multi-task framework but also conducts extensive hyperparameter optimization to balance the various loss components and achieve optimal performance for the specific safety oriented challenges of landmine detection.

Architecture Overview

YOLOv8 builds upon previous YOLO versions with an improved architecture for faster and more accurate object detection, enhanced in our implementation with integrated multi-task learning capabilities to improve spatial feature extraction and object localization through coordinated optimization of multiple detection objectives.

The enhanced YOLOv8 architecture (Figure 3.5) employs a CSPDarknet backbone for feature extraction, integrated multi-task learning modules for coordinated objective optimization, a PANet-based neck for feature fusion across multiple scales, and a decoupled detection head that separately predicts objectness, classification, and bounding box coordinates. This architecture represents a significant advancement over both the traditional machine learning approaches implemented in Phase 1 and standard object detection frame-

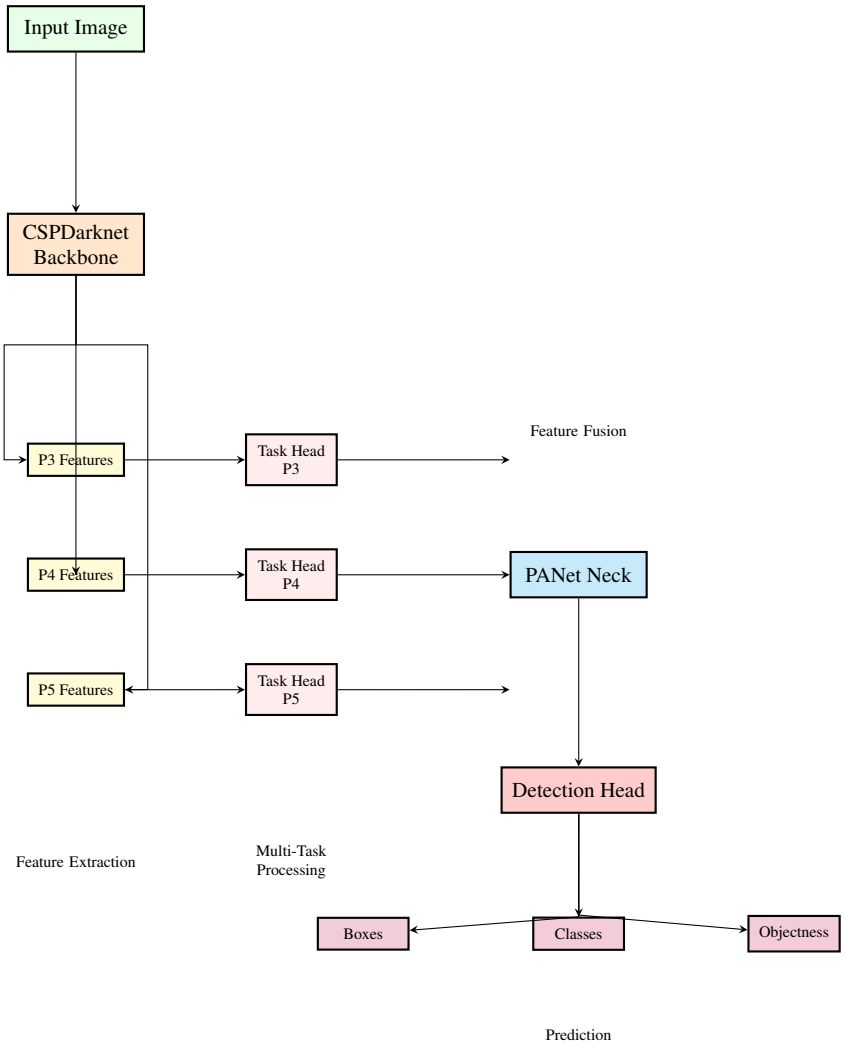

Figure 3.5: YOLOv8 architecture with multi-task learning feature processing, neck, and head components.

works, offering integrated spatial localization capabilities with multi-task learning for coordinated optimization and end-to-end training for object detection tasks.

The backbone network utilizes Cross Stage Partial (CSP) connections within the DarkNet architecture, enabling efficient gradient flow and reducing computational redundancy. The integrated multi-task learning modules process the extracted features at multiple scales (P3, P4, P5), allowing the network to simultaneously optimize multiple detection objectives while leveraging shared feature representations. The PANet neck facilitates feature fusion across multiple scales, allowing the network to detect objects of varying sizes effectively. The decoupled detection head separates the prediction of objectness, classification, and bounding box regression, enabling each component to be optimized independently for improved overall performance.

Multi-Task Learning Integration

The integration of multi-task learning mechanisms within the YOLOv8 architecture serves to enhance the model's ability to simultaneously optimize multiple detection objectives within thermal imagery. Multi-task learning modules are strategically placed after the backbone feature extraction at multiple scales (P3, P4, P5), enabling the network to coordinate the optimization of classification, localization, and confidence estimation tasks before features are passed to the neck for multi-scale fusion.

Each multi-task learning module implements coordinated optimization of multiple objectives using shared feature representations. For a feature map $F \in \mathbb{R}^{H \times W \times C}$, the multi-task processing is computed by applying task-specific transformations while maintaining shared representations:

$$F_{multitask} = \sum_{i=1}^{T} \lambda_i \cdot \text{TaskHead}_i(F_{shared}) + F_{shared} \tag{3.16}$$

where T is the number of tasks, λ_i are task-specific weights, and the residual connection preserves the original feature information while allowing the multi-task learning

mechanism to refine the representations through coordinated optimization.

This multi-task learning-enhanced feature processing enables the model to adaptively balance multiple detection objectives while leveraging shared feature representations, thereby improving both detection accuracy and localization precision compared to standard CNN-based approaches that optimize each objective independently.

3.6 YOLOv8 Loss Function Components

The YOLOv8 model utilizes a composite combination of loss functions that consists of: box loss (CIoU), classification loss (BCE), and distribution focal loss (DFL). Each loss function serves a different function and aids the other in object detection and classification. This multi-component loss function design addresses the unique challenges of object detection by simultaneously optimizing spatial localization, classification accuracy, and coordinate precision.

Box Loss ($\mathcal{L}_{\text{box}}$) Box loss is used to measure the accuracy of the overall bounding box predictions. YOLOv8 uses the Complete IoU (CIoU) loss, defined as:

$$\mathcal{L}_{\text{box}} = 1 - \text{CIoU} = 1 - \left(\text{IoU} - \frac{\rho^2(b, b^{gt})}{c^2} - \frac{v^2}{(1 - \text{IoU}) + v^2} \right) \tag{3.17}$$

where IoU is the Intersection over Union between the predicted box b and ground truth box b^{gt}, $\rho(b, b^{gt})$ is the Euclidean distance between the centers of the predicted and ground truth boxes, c is the diagonal length of the smallest enclosing box covering both boxes, v is a measure of aspect ratio consistency. This can be also defined as:

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \tag{3.18}$$

The CIoU loss simultaneously optimizes: Overlap area (through IoU), Central point distance (through the ρ^2 term), and Aspect ratio consistency (through the v^2 term). This

comprehensive approach to bounding box optimization represents a significant advancement over the simple binary classification outputs of the traditional machine learning methods employed in Phase 1.

Classification Loss ($\mathcal{L}_{\text{cls}}$) YOLOv8 uses Binary Cross-Entropy (BCE) loss for multi-label classification. For each grid cell prediction, the classification loss is calculated as:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}[y_{i,c}\log(p_{i,c}) + (1-y_{i,c})\log(1-p_{i,c})] \tag{3.19}$$

where: N is the number of predicted objects, C is the number of classes, $y_{i,c}$ is the ground truth (1 if object i belongs to class c, 0 otherwise), $p_{i,c}$ is the predicted probability that object i belongs to class c.

Distribution Focal Loss (DFL) The DFL is a novel component introduced to improve bounding box coordinate precision. Instead of directly predicting coordinate values, YOLOv8 models each coordinate as a distribution over a set of discretized values.

$$\mathcal{L}_{\text{dfl}} = -\sum_{i=1}^{n}\sum_{j=1}^{d} y_{i,j}\log(p_{i,j}) \tag{3.20}$$

where: n is the number of coordinates (4 for a bounding box: x, y, width, height), d is the number of discretization bins (usually 16 in YOLOv8), $y_{i,j}$ is the ground truth distribution for coordinate i at bin j, $p_{i,j}$ is the predicted probability for coordinate i at bin j.

The final coordinate value is computed as the expectation over this distribution:

$$\hat{v} = \sum_{j=1}^{d} j \cdot p_j \tag{3.21}$$

This approach allows the model to express uncertainty in coordinate predictions and generally leads to more accurate localization. The distribution-based approach to coordi-

nate prediction represents a sophisticated advancement over traditional regression-based methods, providing the model with the ability to capture and express uncertainty in its predictions.

Combined Loss Function The overall complete and combined loss function in YOLOv8 is a weighted combination of these three components:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{box}} \cdot \mathcal{L}_{\text{box}} + \lambda_{\text{cls}} \cdot \mathcal{L}_{\text{cls}} + \lambda_{\text{dfl}} \cdot \mathcal{L}_{\text{dfl}} \tag{3.22}$$

where λ_{box}, λ_{cls}, and λ_{dfl} are the weights for the box loss, classification loss, and distribution focal loss respectively. The optimization of these weighting parameters formed the core of our grid search methodology, recognizing that the relative importance of each loss component could significantly impact performance in the specific context of landmine detection.

3.7 Grid Search Methodology for Multi-Task Learning Optimization

The transition from traditional machine learning approaches to multi-task learning-enhanced architecture required a systematic approach to hyperparameter optimization, particularly given the critical nature of landmine detection where false negatives could have severe consequences. The grid search methodology was designed to explore the optimal balance between the three loss components while also optimizing multi-task learning parameters, maintaining the rigorous evaluation standards established in Phase 1.

Model Architecture Modifications To optimize our multi-task learning model for the specific requirements of demining operations, we implemented architectural modifications to the base model. The final layer of the network was adjusted to focus specifically on mine detection.

This modification reconfigured the model to output a single class prediction, optimizing the network specifically for mine detection rather than multi-class object detection. This architectural adjustment aligns with the binary classification framework established in Phase 1 while leveraging the superior spatial localization capabilities of the YOLOv8 architecture enhanced with multi-task learning for improved coordinated optimization of detection objectives.

Hyperparameter Grid Search Given the critical nature of demining operations where false negatives could have severe consequences, we conducted an extensive grid search across 64 different model variants to identify optimal hyperparameter configurations. This comprehensive approach builds upon the evaluation framework established in Phase 1, extending the rigorous cross-validation methodology to the more complex hyperparameter space of deep learning architectures.

Search Space Definition The grid search explored combinations of the three primary loss function weights in the YOLOv8 architecture:

Table 3.1: Loss Weight Parameters Search Space

Parameter	**Range**	**Step Size**
Box Loss Weight (λ_{box})	5.0–12.5	2.5
Class Loss Weight (λ_{cls})	0.5–2.0	0.5
DFL Loss Weight (λ_{dfl})	1.0–3.0	0.5

This resulted in a total of 64 distinct model configurations (4 box weight values × 4 class weight values × 4 DFL weight values). The parameter ranges were selected based on preliminary experiments and established best practices in object detection literature, while the step sizes were chosen to provide sufficient granularity for identifying optimal configurations without excessive computational overhead.

Training Protocol Each of the 64 model variants was trained using a standardized protocol that maintained consistency with the rigorous evaluation framework established in

Phase 1:

- **K-Fold Cross-Validation**: Each model was trained using 5-fold cross-validation with shuffling enabled and a fixed random state to ensure reproducibility, maintaining consistency with the evaluation approach used for traditional machine learning methods
- **Consistent Data Splitting**: Training and validation data directories were specified through YAML configurations for each fold, ensuring identical data partitioning across all model variants
- **Hardware Configuration**: All models were trained on identical hardware configurations to eliminate variability and ensure fair comparison
- **Training Duration**: Each model was trained for 100 epochs with an early stopping patience of 20 epochs to prevent overfitting while allowing sufficient time for convergence
- **Batch Size**: A consistent batch size of 16 was used across all training runs, balancing memory requirements with training stability
- **Optimizer**: All models used the AdamW optimizer with an initial learning rate of 0.001 and cosine learning rate scheduling to ensure optimal convergence behavior

This comprehensive training protocol ensures that performance differences between model variants can be attributed to hyperparameter choices rather than implementation inconsistencies or random variations in the training process.

Evaluation Framework The evaluation framework for YOLOv8 optimization extended the comprehensive assessment approach established in Phase 1 while incorporating metrics specific to object detection tasks.

Performance Metrics Each model variant was evaluated using a comprehensive set of metrics that provide both absolute performance measures and insights into the trade-offs between different aspects of detection performance:

- **Mean Average Precision (mAP)**: Standard detection performance at IoU threshold of 0.5, providing a comprehensive measure of detection accuracy that accounts for both localization precision and classification accuracy
- **Precision**: The ratio of true positive detections to all positive detections, indicating the reliability of positive predictions
- **Recall**: The ratio of true positive detections to all ground truth objects, measuring the model's ability to identify all instances of landmines
- **False Negative Rate (FN Rate)**: The proportion of mines that went undetected, calculated as $1 -$ Recall, with particular importance given the critical consequences of missed detections in demining operations
- **Balanced Score**: A composite metric calculated as mAP $\times$ Recall to identify models with good overall performance while maintaining high recall, recognizing the particular importance of minimizing false negatives in landmine detection

Statistical Analysis For each model configuration, we aggregated performance metrics across all five folds to calculate comprehensive statistical measures that ensure robust comparison between different hyperparameter configurations:

- Mean performance values across all folds to provide central tendency measures
- Standard deviations to assess stability and consistency of performance across different data partitions
- 95% confidence intervals for robust comparison, enabling statistical significance testing between different model configurations

This statistical framework provides the foundation for making informed decisions about optimal hyperparameter configurations while accounting for the inherent variability in deep learning model training.

Visual Analysis We implemented comprehensive visualization techniques to analyze the grid search results and identify optimal hyperparameter configurations:

- **Parameter Response Curves**: Visualized the relationship between individual hyperparameters and performance metrics, enabling identification of optimal ranges for each parameter
- **Heatmaps**: Created multi-dimensional visualizations to identify interactions between pairs of hyperparameters while holding the third constant, revealing complex interdependencies in the parameter space
- **3D Scatter Plots**: Plotted the entire parameter space with color-coding by performance metrics, providing an intuitive understanding of the performance landscape
- **Trade-off Analysis**: Visualized the relationship between false negative rate and precision to identify optimal trade-off points that balance detection accuracy with operational requirements

These visualization techniques enable comprehensive analysis of the hyperparameter optimization results and provide insights into the underlying relationships between different loss function weights and their impact on landmine detection performance.

3.8 Integration and Comparison Framework

The sequential implementation of traditional machine learning approaches followed by YOLOv8 optimization provides a comprehensive framework for evaluating different algorithmic approaches to landmine detection. This two-phase methodology enables direct

comparison of performance metrics while highlighting the advantages and limitations of each approach. The consistent evaluation framework, including 5-fold cross-validation and standardized performance metrics, ensures fair comparison across all implemented methods.

The progression from Phase 1 to Phase 2 represents not only a technological advancement but also a methodological evolution from classification-based approaches to sophisticated object detection frameworks. This comprehensive methodology provides the foundation for identifying optimal solutions for landmine detection while contributing to the broader understanding of machine learning applications in critical security domains.

Chapter 4

RESULTS

4.1 Introduction

This chapter presents a comprehensive evaluation of landmine detection performance across two distinct phases of investigation. The results demonstrate the evolution from traditional machine learning classification approaches to advanced object detection methodologies, providing critical insights into the relative strengths and limitations of each approach. The sequential presentation allows for direct performance comparison while highlighting the perfromance gains achieved through the systematic progression from Phase 1 to Phase 2 implementations.

The evaluation framework maintains consistency across both phases through standardized 5-fold cross-validation procedures and comprehensive performance metrics, ensuring fair comparison between fundamentally different algorithmic approaches. This rigorous assessment provides the foundation for understanding not only absolute performance levels but also the practical trade-offs between computational complexity, implementation requirements, and detection accuracy in real-world demining scenarios.

4.2 Phase 1 Results: Traditional Machine Learning Performance

The initial phase of investigation focused on establishing baseline performance using traditional machine learning approaches. All experiments in this phase were conducted using a balanced dataset obtained through augmentation, with a 70/30 training/testing split and 5-fold cross-validation to ensure robust and unbiased evaluation. This comprehensive evaluation of Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN) classifiers provides essential insights into the fundamental capabilities and limitations of conventional methods for landmine detection. The results from this phase served as the foundation for understanding the problem complexity and informed the subsequent development of more sophisticated approaches in Phase 2.

4.2.1 Random Forest Classification Results

The Random Forest classifier achieved consistent performance across all cross-validation folds, demonstrating the robustness and reliability characteristic of ensemble methods. Table 4.1 summarizes the key performance metrics for the Random Forest model across all evaluation folds.

Table 4.1: Cross-Validation Performance Metrics for Random Forest (%)

Fold	Accuracy	Precision	Recall	F1-Score	AUC
1	91.52	87.79	97.46	92.37	97.35
2	89.73	86.18	94.64	90.21	95.72
3	92.41	87.50	98.13	92.51	97.77
4	93.75	89.29	98.04	93.46	97.94
5	91.96	87.16	95.96	91.35	96.49
Mean	91.88	87.58	96.85	91.98	97.05
Std	1.31	1.01	1.35	1.11	0.83

The Random Forest model demonstrated strong overall performance with an average accuracy of 91.88% across all folds. The high recall score (96.85%) indicates the model's

effectiveness in identifying positive instances (images containing landmines), which is particularly important in landmine detection applications where false negatives can have severe consequences. The precision of 87.58% shows that the model maintains a good balance between detecting landmines and avoiding false positives. The low standard deviation across all metrics (ranging from 0.83% to 1.35%) demonstrates the consistency and reliability of the Random Forest approach across different data partitions.

Figure 4.1 illustrates the ROC curves for all cross-validation folds, showing the trade-off between True Positive Rate and False Positive Rate. The high area under the curve (AUC) of 97.05% further confirms the model's strong discriminative ability and provides confidence in its potential for practical deployment scenarios.

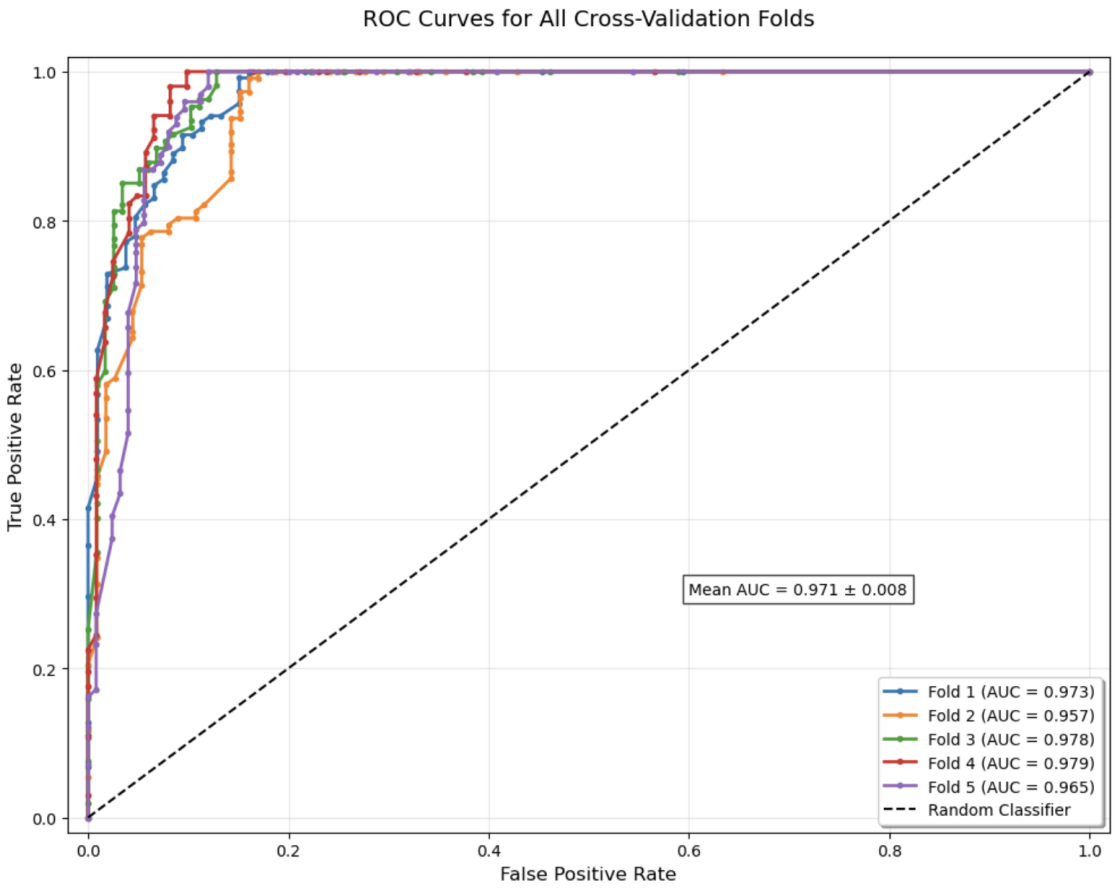

Figure 4.1: ROC curves for all Random Forest cross-validation folds, showing the trade-off between True Positive Rate and False Positive Rate.

Figure 4.2 presents the aggregated confusion matrix across all folds, providing a visual representation of the model's predictions versus ground truth values. The confusion matrix reveals that most false predictions occur as false positives rather than false negatives, which aligns with the preference for higher recall in safety-critical detection applications. This pattern suggests that the Random Forest approach errs on the side of caution, which is appropriate for humanitarian demining operations where missing a landmine has far more severe consequences than investigating a false positive.

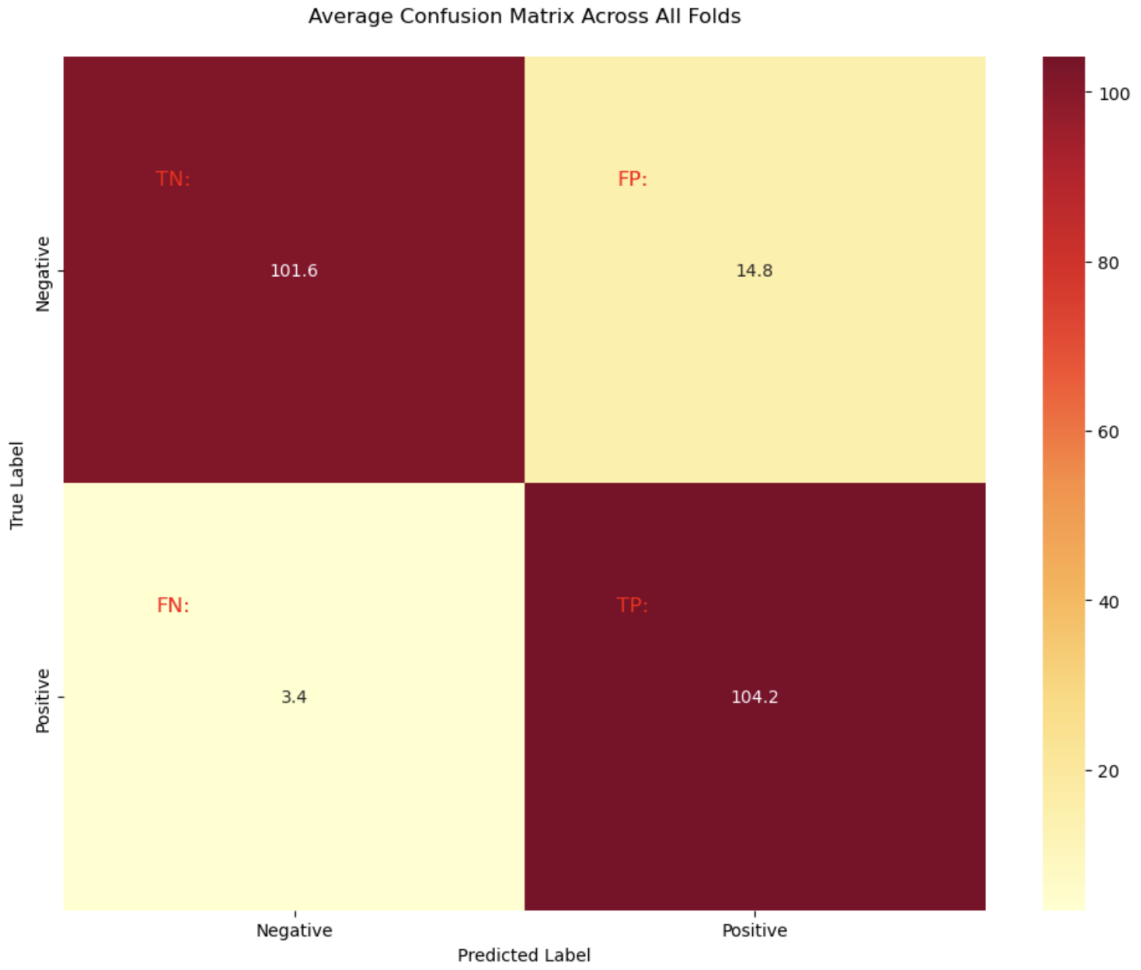

Figure 4.2: Aggregated confusion matrix across all folds, showing the distribution of predictions versus ground truth for the Random Forest model.

4.2.2 Support Vector Machine Classification Results

The Support Vector Machine with RBF kernel showed competitive performance across all evaluation metrics, demonstrating the effectiveness of kernel-based approaches for complex pattern recognition in thermal imagery. Table 4.2 summarizes the cross-validation results for the SVM classifier, revealing consistent performance characteristics that complement the Random Forest findings.

Table 4.2: Cross-Validation Performance Metrics for SVM (%)

Fold	Accuracy	Precision	Recall	F1-Score	AUC
1	89.29	86.15	94.92	90.32	95.45
2	90.62	85.83	97.32	91.21	94.79
3	90.18	84.55	97.20	90.43	95.49
4	92.41	88.29	96.08	92.02	97.03
5	91.07	84.35	97.98	90.65	96.35
Mean	90.71	85.83	96.70	90.93	95.82
Std	1.14	1.58	1.21	0.69	0.85

The SVM classifier achieved an average accuracy of 90.71%, slightly lower than the Random Forest model but still demonstrating strong performance for the challenging task of landmine detection. Similar to the Random Forest model, the SVM maintained high recall (96.70%) across all folds, indicating its effectiveness in identifying landmine-containing images while minimizing the critical false negative errors. The precision (85.83%) was marginally lower than the Random Forest, suggesting a slightly higher rate of false positives, which represents an acceptable trade-off given the safety-critical nature of the application.

Figure 4.3 shows the ROC curves for all SVM cross-validation folds, demonstrating the trade-off between True Positive Rate and False Positive Rate. The high AUC (95.82%) confirms the model's robust discriminative ability, though slightly lower than the Random Forest model. The consistency of the ROC curves across folds indicates stable performance characteristics that would be valuable in operational deployments.

Figure 4.4 illustrates the aggregated SVM confusion matrix across all folds, showing

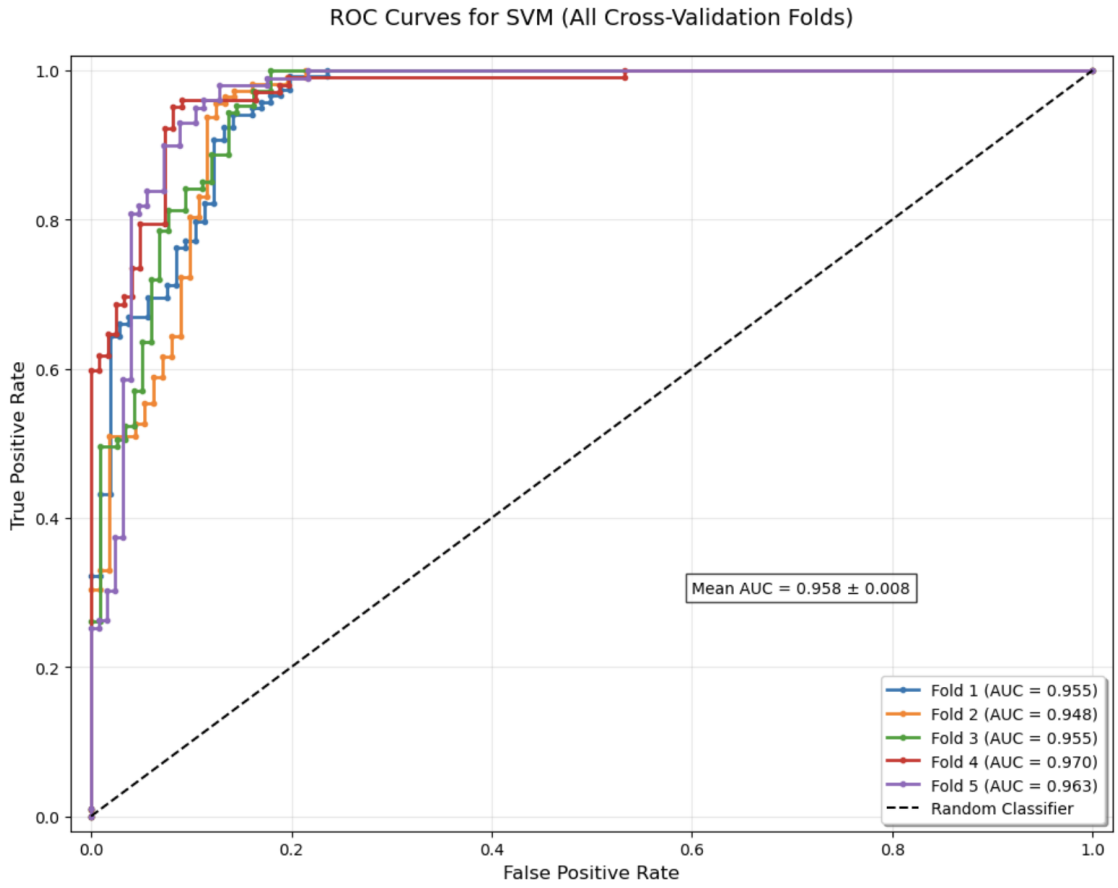

Figure 4.3: ROC curves for all SVM cross-validation folds, demonstrating the trade-off between True Positive Rate and False Positive Rate.

the distribution of predictions versus ground truth. The confusion matrix pattern is similar to the Random Forest, with false positives being more common than false negatives. This consistency across different algorithmic approaches suggests that this error pattern may be inherent to the dataset characteristics and the nature of thermal landmine detection, providing valuable insights for operational planning and decision-making protocols.

4.2.3 Artificial Neural Network Classification Results

The modified ResNet50 architecture achieved strong performance across the evaluation metrics, demonstrating the effectiveness of deep learning approaches even within the classification framework of Phase 1. Table 4.3 summarizes the cross-validation results for the

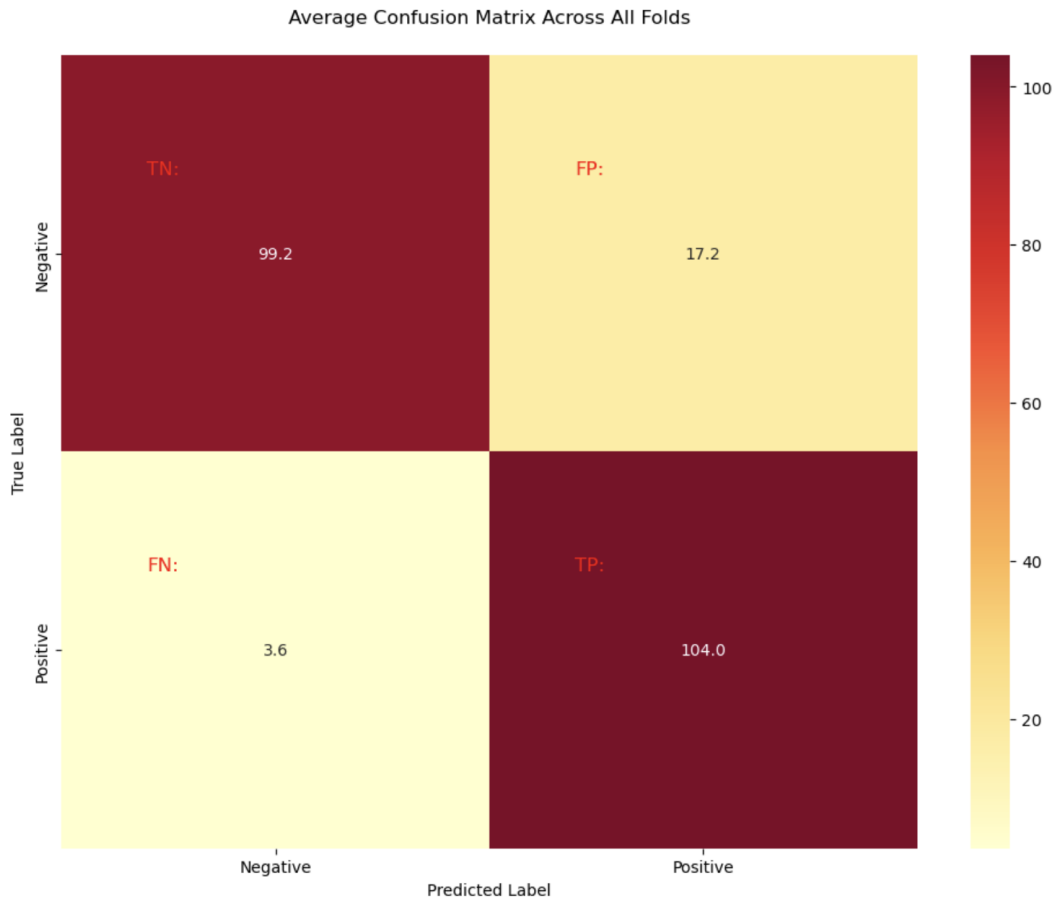

Figure 4.4: Aggregated SVM confusion matrix across all folds, illustrating the distribution of predictions versus ground truth.

neural network classifier, revealing superior performance characteristics that foreshadowed the potential for even greater improvements in the subsequent object detection phase.

The ResNet50-based model achieved an average accuracy of 94.29% across all folds, outperforming both traditional machine learning approaches and establishing the potential for deep learning methods in landmine detection. The model maintained high precision (95.69%) while achieving good recall (92.31%), indicating effective balance between false positives and false negatives. The exceptional AUC score of 99.11% demonstrates the model's strong discriminative ability and suggests that the hierarchical feature learning capabilities of deep neural networks are particularly well-suited for thermal imagery analysis.

Figure 4.5 shows the ROC curves for all cross-validation folds, demonstrating the trade-

Table 4.3: Cross-Validation Performance Metrics for ResNet50 CNN (%)

Fold	Accuracy	Precision	Recall	F1-Score	AUC
1	96.88	94.40	100.00	97.12	98.32
2	82.59	91.95	71.43	80.40	98.73
3	98.21	98.13	98.13	98.13	99.45
4	97.32	97.06	97.06	97.06	99.91
5	96.43	96.91	94.95	95.92	99.15
Mean	94.29	95.69	92.31	93.73	99.11
Std	6.22	2.51	11.54	7.31	0.61

off between True Positive Rate and False Positive Rate. The consistently high AUC values across all folds indicate robust model performance, though there is notable variation in fold 2's performance compared to the others. This variability, while concerning from a consistency standpoint, provided important insights into the sensitivity of deep learning approaches to training data composition and motivated the more sophisticated approaches implemented in Phase 2.

Figure 4.6 presents the aggregated confusion matrix across all folds, providing a visual representation of the model's predictions versus ground truth values. The confusion matrix reveals that the model achieves a good balance between minimizing both false positives and false negatives, with particularly strong performance in correctly identifying positive cases. This balanced performance profile differs from the traditional machine learning approaches and suggests that deep learning methods may offer more flexible optimization capabilities.

The integration of multiple input modalities proved particularly effective in the neural network approach. The model's architecture successfully combined features extracted from thermal imagery with metadata (depth, altitude, and temperature), leveraging the hierarchical feature learning capabilities of deep neural networks. However, this sophisticated modeling approach comes with increased computational requirements compared to traditional machine learning methods, a consideration that influenced the design of the Phase 2 investigations.

Notably, while the model achieved superior overall performance metrics, it showed

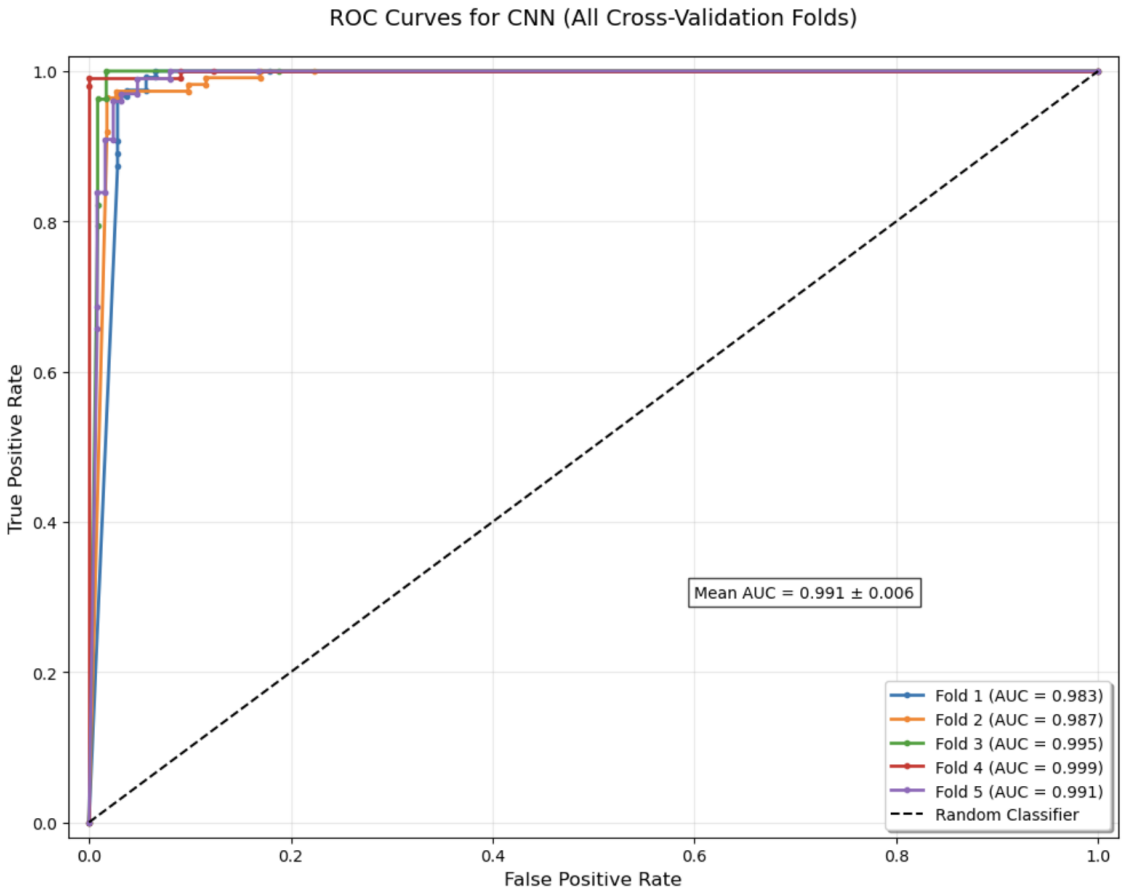

Figure 4.5: ROC curves for all ResNet50 cross-validation folds, demonstrating the trade-off between True Positive Rate and False Positive Rate.

more variation across folds compared to the Random Forest and SVM approaches, particularly in fold 2 where accuracy dropped to 82.59%. This suggests that the neural network model might be more sensitive to variations in the training data, a consideration that motivated the development of more robust training protocols and hyperparameter optimization strategies in the subsequent phase of research.

4.2.4 Phase 1 Comparative Analysis

Comparing the performance of all three traditional machine learning models reveals several important insights that informed the direction of Phase 2 investigations:

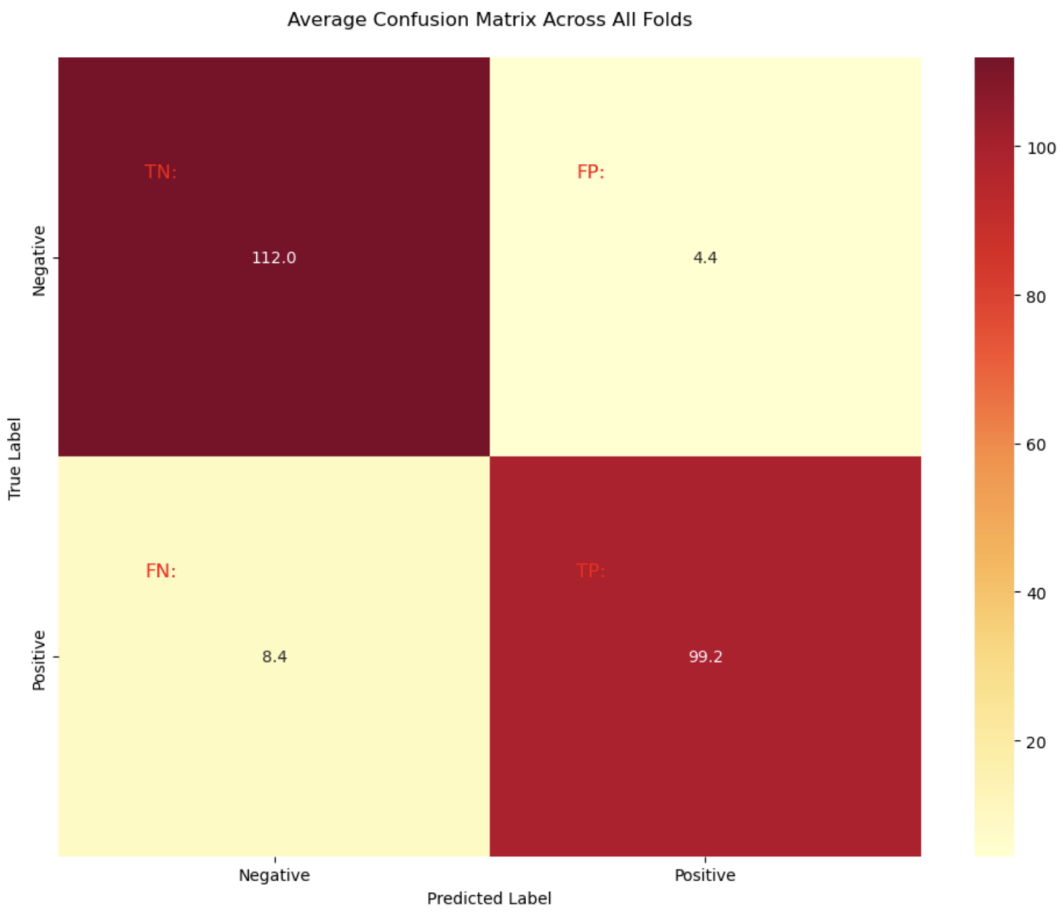

Figure 4.6: Aggregated ResNet50 confusion matrix across all folds, illustrating the distribution of predictions versus ground truth.

Overall Performance Hierarchy The ResNet50-based neural network demonstrated the highest overall accuracy (94.29%), followed by the Random Forest classifier (91.88%), and the SVM (90.71%). This hierarchy suggests that deep learning approaches can capture more complex patterns in thermal imagery data, but also highlighted the need for more sophisticated architectures specifically designed for object detection rather than classification.

Precision-Recall Trade-off Characteristics Each model showed different strengths in the precision-recall trade-off, revealing fundamental algorithmic differences:

- ResNet50 achieved the best balance with high precision (95.69%) and recall (92.31%),

demonstrating the potential for deep learning optimization.

- Random Forest prioritized recall (96.85%) while maintaining good precision (87.58%), showing consistent conservative behavior.
- SVM showed similar characteristics to Random Forest with high recall (96.70%) but lower precision (85.83%), indicating kernel-based approaches favor sensitivity over specificity.

Consistency and Reliability The ResNet50 model showed the highest variability across folds, particularly in fold 2, while the Random Forest demonstrated more consistent performance with a standard deviation in accuracy of 1.31%. This finding highlighted the need for more robust training strategies and motivated the extensive hyperparameter optimization conducted in Phase 2.

Discriminative Capability All models achieved excellent AUC values, indicating strong underlying discriminative capabilities:

- ResNet50: 99.11
- Random Forest: 97.05
- SVM: 95.82

These results confirmed that the fundamental approach of using thermal imagery with metadata was sound, but suggested that specialized object detection architectures could potentially achieve even better performance.

Computational Considerations The integration of additional features (depth, altitude, and temperature) alongside the image data proved beneficial for all approaches, allowing the models to leverage multi-modal information for improved detection accuracy. However,

the computational overhead varied significantly between approaches, with neural networks requiring substantially more resources than traditional methods.

While the neural network approach provided superior accuracy and AUC performance, it came with significantly higher computational requirements. The traditional machine learning models (particularly Random Forest) demonstrated excellent performance with lower computational overhead, suggesting they may be suitable for resource-constrained environments. However, these Phase 1 results also revealed a critical limitation: all approaches were fundamentally classification-based and could not provide the spatial localization information required for practical demining operations. This limitation, combined with the promising performance of deep learning approaches, motivated the transition to object detection methodologies in Phase 2.

4.3 Phase 2 Results: YOLOv8 Optimization and Performance

Building upon the insights gained from Phase 1, the second phase of investigation focused on implementing and optimizing YOLOv8 for landmine detection. Experiments in this phase also utilized the balanced dataset with a 70/30 training/testing split and 5-fold cross-validation, following industry-standard evaluation protocols. This phase represents a fundamental shift from classification to object detection, enabling precise spatial localization while maintaining the high-performance standards established in the traditional machine learning evaluation. The comprehensive grid search optimization process yielded significant insights into optimal configurations for mine detection applications and demonstrated substantial improvements over both standard YOLOv8 configurations and the Phase 1 baseline results.

The transition to object detection methodology addresses the critical limitation identified in Phase 1 regarding spatial localization while leveraging the superior pattern recogni-

tion capabilities demonstrated by deep learning approaches. The systematic hyperparameter optimization ensures that the YOLOv8 architecture is specifically tuned for the unique challenges of landmine detection, rather than relying on general-purpose object detection configurations.

4.3.1 Hyperparameter Optimization Results

The comprehensive grid search of 64 hyperparameter configurations yielded significant insights into the optimal settings for mine detection with YOLOv8. Through systematic testing and analysis, we identified clear parameter effects on key performance metrics and their trade-offs, providing a thorough understanding of how different loss function weights impact detection performance in the specific context of landmine identification.

Effects on False Negative Rate The false negative rate, critical in mine detection operations where missed detections can have life-threatening consequences, showed strong dependence on hyperparameter settings. The analysis revealed clear patterns that directly inform optimal configuration selection:

- **Class Loss Weight ($\lambda_{\mathbf{cls}}$)**: Exhibited the strongest influence on false negative rate, with a clear inverse relationship that demonstrates the importance of classification accuracy in overall detection performance. As shown in Fig. 4.7, increasing the class loss weight from 0.3 to 2.0 resulted in a substantial and consistent reduction in false negative rate from a median of approximately 0.17 to 0.13. This represents a relative improvement of approximately 23.5%, demonstrating that prioritizing classification accuracy directly translates to fewer missed detections.

- **Box Loss Weight ($\lambda_{\mathbf{box}}$)**: Demonstrated a non-monotonic relationship with false negative rate, with the lowest FN rates observed at box weight values of 5.0. Increasing box weight beyond this value generally led to higher false negative rates, suggesting

that overemphasis on box accuracy may come at the expense of detection completeness. This finding indicates that while spatial accuracy is important, excessive focus on bounding box precision can detract from the fundamental detection task.

- **DFL Loss Weight (λ_{dfl})**: Showed the weakest correlation with false negative rate, with a slight upward trend as the weight increased. This indicates that while coordinate precision contributes to overall performance, it has less impact on detection recall than the other parameters, suggesting that classification and basic box localization are more critical than fine-grained coordinate precision for landmine detection.

Figure 4.8 presents the detailed response curves for false negative rate across each parameter dimension, highlighting both individual data points and the statistical trends with confidence intervals. The visualization clearly demonstrates the robustness of the identified relationships and provides confidence intervals that support the statistical significance of the observed trends.

Effects on Mean Average Precision (mAP) Model accuracy as measured by mAP@0.5 showed distinct patterns across the parameter space, providing insights into the relationship between different loss components and overall detection performance. The mAP metric is particularly important as it provides a comprehensive assessment of both detection accuracy and localization precision:

- **Class Loss Weight**: Demonstrated the strongest positive correlation with mAP, with values increasing steadily from 0.52 at $\lambda_{\text{cls}} = 0.3$ to approximately 0.54 at $\lambda_{\text{cls}} = 2.0$. This consistent positive relationship suggests that increasing emphasis on correct classification directly improves overall detection performance, reinforcing the finding that classification accuracy is fundamental to successful landmine detection.

- **Box Loss Weight**: Showed a slight negative trend with mAP, with the highest values observed at the lowest box weight setting of 5.0. This counter-intuitive result sug-

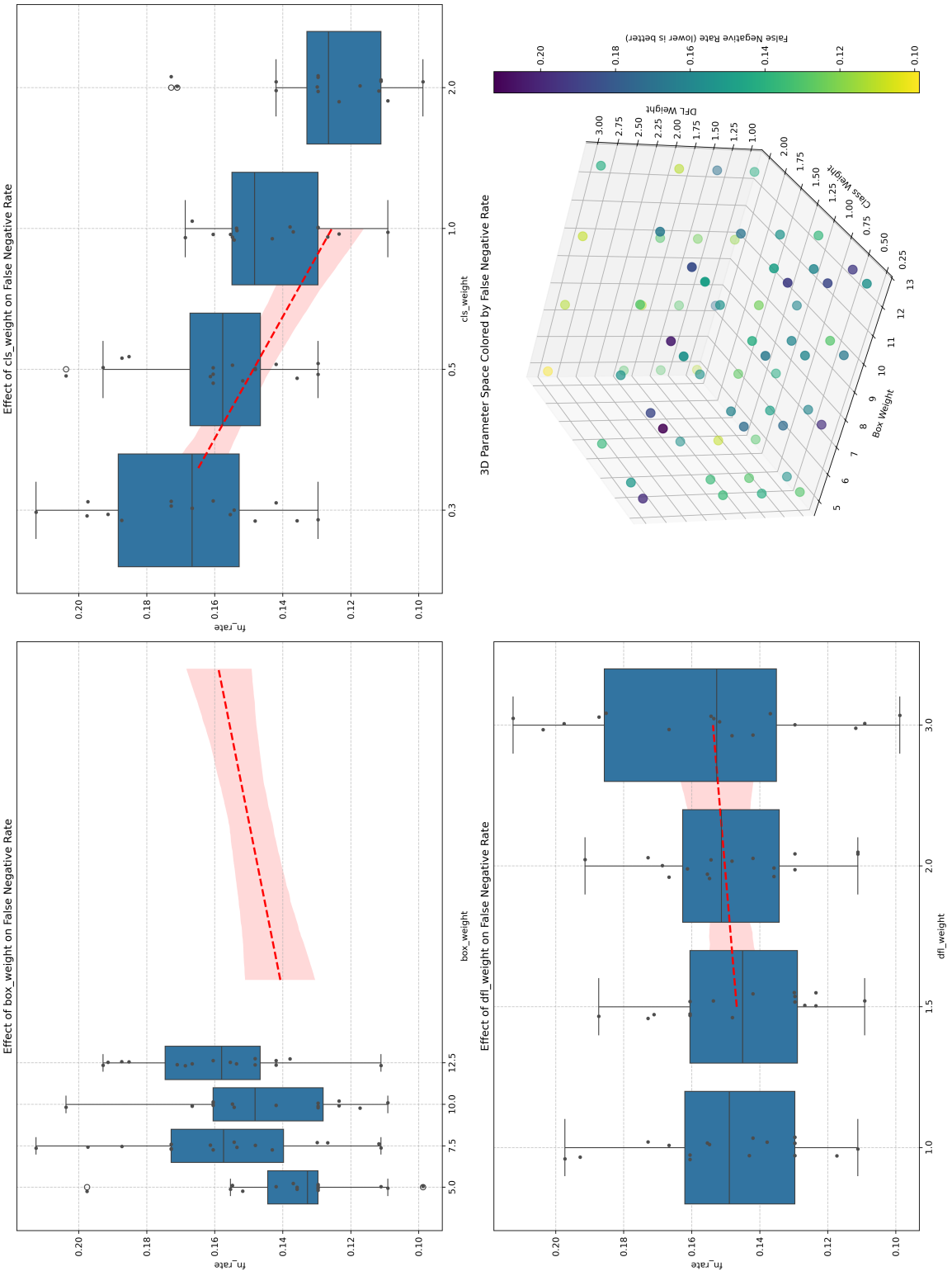

Figure 4.7: Effects of box weight, class weight, and DFL weight on false negative rate. The bottom right panel shows the 3D parameter space visualization colored by false negative rate.

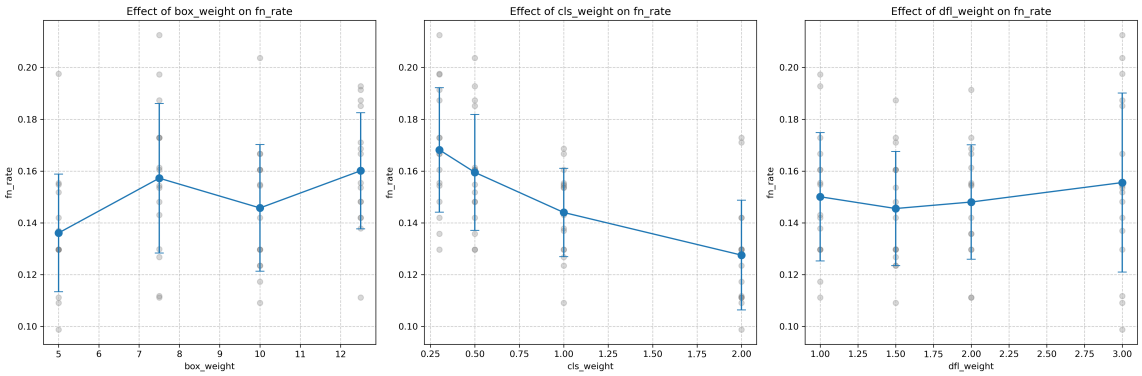

Figure 4.8: Detailed response curves showing the effect of each hyperparameter on false negative rate. The visualization includes individual data points and confidence intervals to illustrate the statistical significance of the observed trends.

gests that overemphasizing bounding box accuracy can potentially distract the model from overall detection performance, indicating that a balanced approach to spatial localization is more effective than excessive precision in box coordinates.

- **DFL Loss Weight**: Had minimal impact on mAP, with a slight negative correlation as weight increased, supporting the finding that coordinate precision plays a secondary role in overall detection performance. This result suggests that while the distribution focal loss contributes to coordinate accuracy, its impact on the fundamental detection task is limited compared to classification and basic box localization.

The response curves in Fig. 4.10 further illustrate these relationships, with the positive trend for class weight contrasting with the relatively flat or slightly negative trends for box and DFL weights. This visualization provides clear guidance for hyperparameter selection, emphasizing the importance of classification weight while suggesting more moderate values for the other parameters.

Effects on Precision and Recall Precision and recall metrics revealed additional insights into model behavior, providing a more detailed understanding of the trade-offs between different types of detection errors and their relationship to hyperparameter settings:

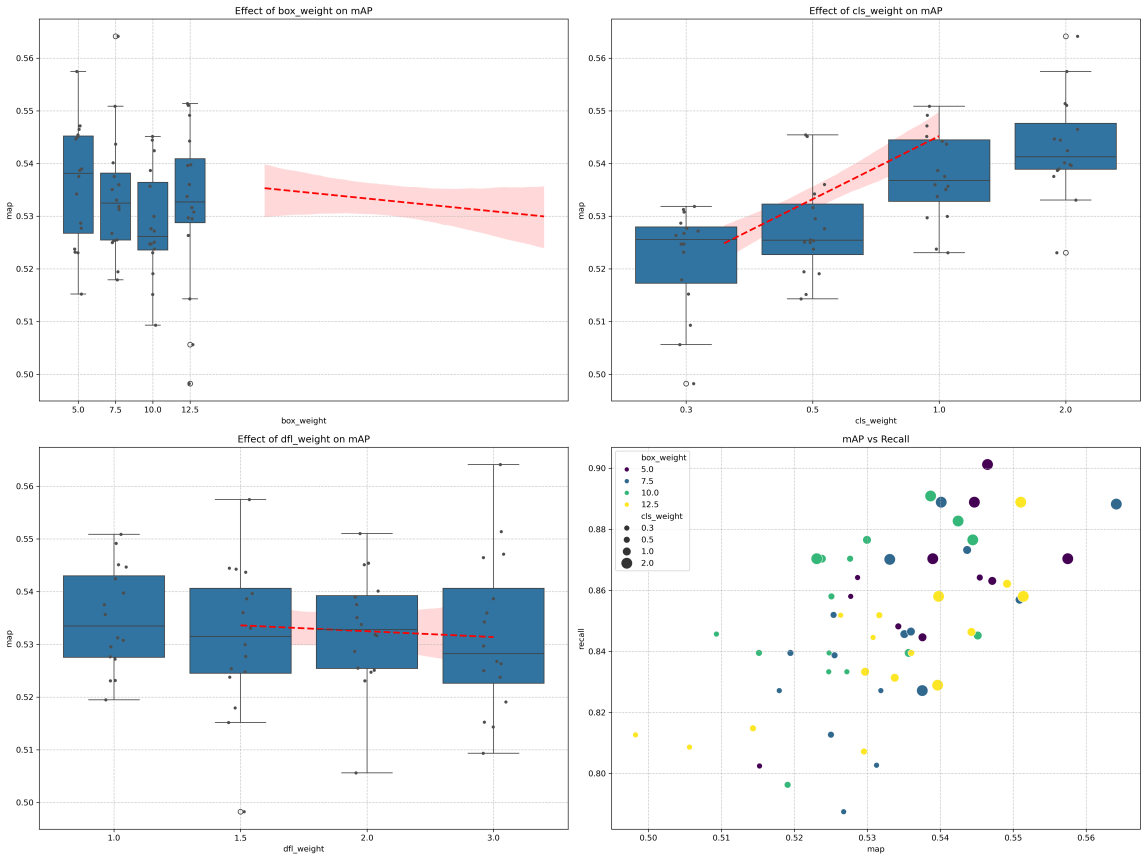

Figure 4.9: Effects of box weight, class weight, and DFL weight on mean average precision (mAP). The bottom right panel shows the relationship between mAP and recall across different configurations.

- **Precision**: Box weight showed a non-monotonic relationship with precision, peaking at a value of 7.5, suggesting an optimal balance point for spatial accuracy that maximizes the reliability of positive detections. Class weight demonstrated a positive correlation with precision, particularly at higher values, reinforcing the importance of classification accuracy. DFL weight exhibited a complex relationship with precision, peaking at a value of approximately 1.75, as shown in Fig. 4.11, indicating that moderate coordinate precision optimization is beneficial but excessive focus can be counterproductive.

- **Recall**: Directly related to false negative rate (recall = 1 - FN rate), recall showed the strongest positive correlation with class weight, confirming the critical importance of

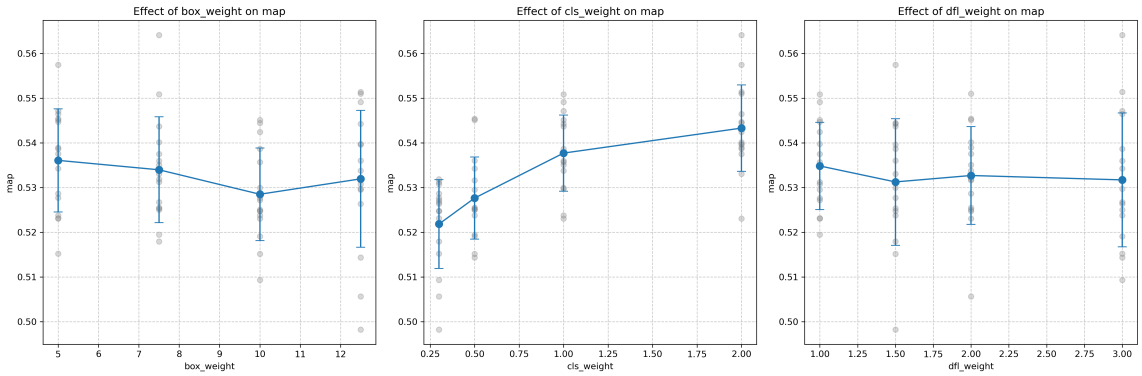

Figure 4.10: Response curves showing the effect of each hyperparameter on mean average precision (mAP). The clear positive trend for class weight contrasts with the relatively flat or slightly negative trends for box and DFL weights.

classification accuracy for detecting all landmine instances. Figure 4.12 demonstrates that class weight values of 2.0 achieved the highest recall of approximately 0.87, compared to 0.83 at class weight 0.3, representing a relative improvement of 4.8%. This improvement is particularly significant in landmine detection applications where maximizing recall is critical for safety.

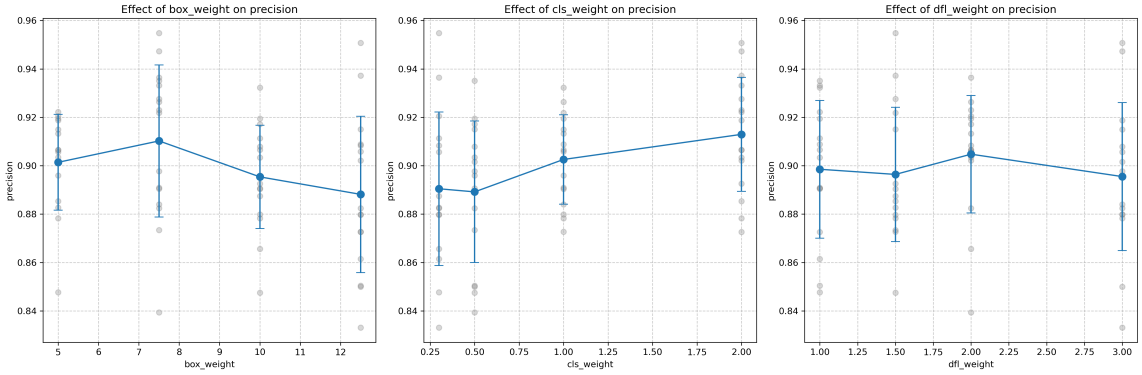

Figure 4.11: Response curves showing the effect of each hyperparameter on precision. Grey dots represent individual model configurations, while blue lines with error bars show the mean trend with confidence intervals.

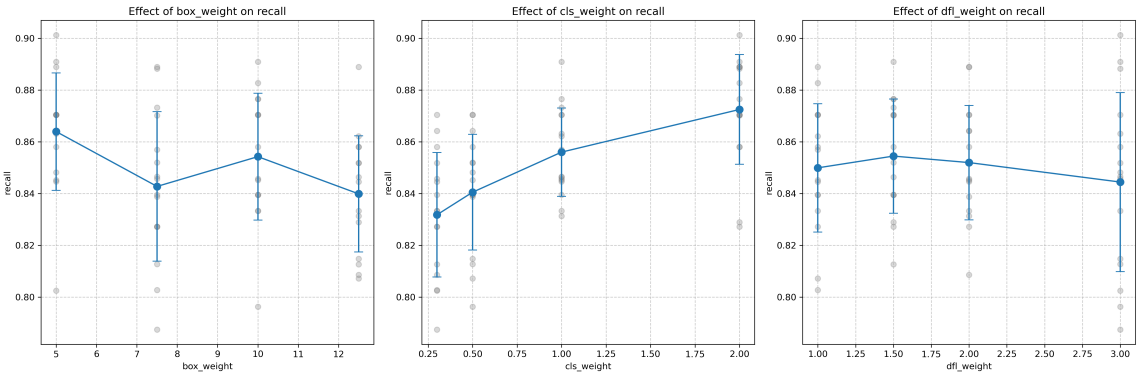

Figure 4.12: Response curves showing the effect of each hyperparameter on recall. Grey dots represent individual model configurations, while blue lines with error bars show the mean trend with confidence intervals.

4.3.2 Performance Trade-off Analysis

Understanding the trade-offs between different performance metrics is crucial for selecting optimal configurations in safety-critical applications like landmine detection. The comprehensive analysis of these trade-offs provides essential guidance for operational deployment decisions.

Pareto Frontier Analysis The trade-off between false negative rate and precision, critical in mine detection applications, is visualized in Fig. 4.13. This scatter plot reveals the Pareto frontier of model performance, where improvements in one metric necessarily come at the expense of the other. The identification of this frontier enables informed decision-making about optimal operating points based on specific operational requirements and risk tolerance.

We identified three notable configurations along this frontier, each representing different optimization priorities:

- **Best mAP**: A model achieving high overall performance with mAP ≈ 0.57, moderate false negative rate (≈ 0.11), and good precision (≈ 0.95). This configuration represents the best balance between all performance metrics and would be suitable

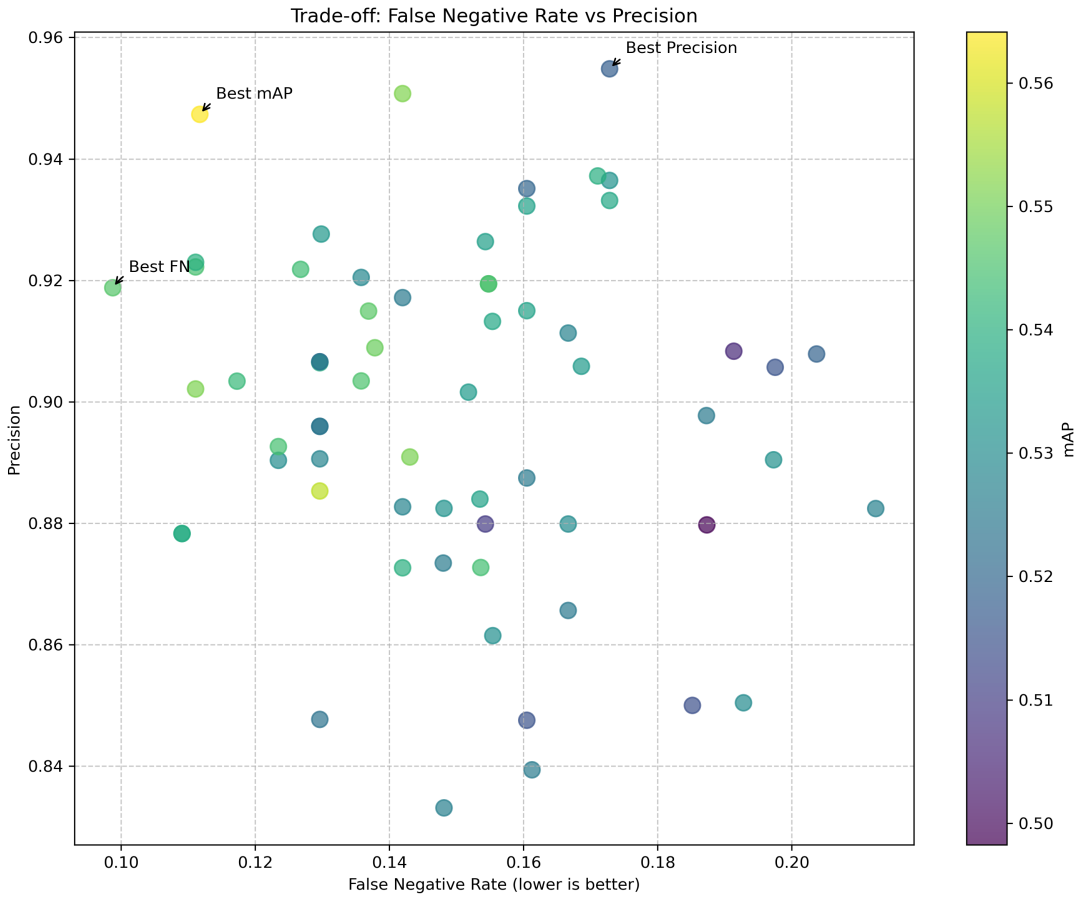

Figure 4.13: Trade-off analysis between false negative rate and precision across different model configurations. Points are colored by mAP values, with labels indicating models with best precision, best mAP, and lowest false negative rate.

for applications where overall detection performance is the primary concern.

- **Best Precision**: A model with excellent precision (≈ 0.955) but higher false negative rate (≈ 0.17). This configuration minimizes false positives but at the cost of missing more landmines, which may be appropriate in scenarios where investigation resources are severely limited.

- **Best FN Rate**: A model with the lowest false negative rate (≈ 0.10) while maintaining acceptable precision (≈ 0.92). This configuration prioritizes the detection of all landmines, which is typically the most appropriate choice for humanitarian demining operations where safety is paramount.

4.3.3 Optimal Configuration Validation

Based on our prioritization of minimizing false negatives while maintaining acceptable precision for landmine detection, we selected the configuration labeled "Best FN" as optimal. This decision reflects the critical importance of detecting all landmines in humanitarian demining operations, where the consequences of missed detections far outweigh the costs of investigating false positives.

Cross-Validation Performance Assessment The optimal configuration was validated through comprehensive 5-fold cross-validation, with results summarized in Table 4.4. This rigorous evaluation ensures that the observed performance improvements are robust and not the result of overfitting to a particular data partition.

Table 4.4: Performance Comparison of Optimized vs. Standard YOLOv8 Configuration

Configuration	**mAP**	**Prec.**	**Recall**	**FN**	**Bal.**
Standard	0.532	0.895	0.840	0.160	0.447
Optimized	0.541	0.920	0.900	0.100	0.487
Improvement	+1.7%	+2.8%	+7.1%	-37.5%	+8.9%

The performance metrics in Table 4.4 demonstrate significant improvements across all dimensions:

- **mAP (Mean Average Precision)**: Measures overall detection accuracy across different confidence thresholds. The optimized model achieved 0.541 compared to 0.532 for the standard configuration, representing a 1.7% improvement in comprehensive detection performance.

- **Precision**: Indicates the proportion of detected mines that are actually mines (true positives / [true positives + false positives]). The optimized model achieved 0.920 precision versus 0.895 for standard, showing a 2.8% improvement in reducing false alarms—important for operational efficiency in demining operations.

- **Recall**: Measures the proportion of actual mines that were successfully detected (true positives / [true positives + false negatives]). This is critical for safety as it represents the model's ability to find all mines. The optimized configuration achieved 0.900 recall compared to 0.840 standard, a significant 7.1% improvement.

- **FN (False Negative Rate)**: The proportion of actual mines that were missed by the detector (false negatives / [false negatives + true positives]). This is the most critical metric for mine detection safety. The optimized model achieved a dramatic 37.5% reduction in false negatives (from 0.160 to 0.100), meaning far fewer dangerous mines would be left undetected.

- **Bal. (Balance Score)**: A composite metric calculated as mAP $\times$ Recall, rewarding models that perform well on both overall accuracy and mine detection completeness. The optimized model's balance score of 0.487 represents an 8.9% improvement over the standard configuration's 0.447, indicating superior performance across both dimensions critical for safe and effective mine detection.

With the most significant advancement being the 37.5% reduction in false negative rate from 0.160 to 0.100. This improvement directly addresses the most critical concern in landmine detection applications. Additionally, the optimized configuration achieved a 7.1% improvement in recall (from 0.840 to 0.900) and a 2.8% improvement in precision (from 0.895 to 0.920), demonstrating that the optimization process successfully enhanced performance across multiple dimensions simultaneously.

4.3.4 Detailed Performance Analysis

The confusion matrices for a representative test fold are visualized in Fig. 4.14, providing a more detailed view of the classification outcomes. While these matrices represent a single test fold rather than the cross-validation average reported in Table 4.4, they clearly demonstrate the substantial reduction in false negatives (from 43 to 27) achieved by the optimized

model, along with improvements in all other metrics.

The slight variations between the specific values shown in the confusion matrices and the averages reported in Table 4.4 are expected due to fold-to-fold variance in the cross-validation process. However, the consistent pattern of improvement across individual folds confirms the robustness of the optimization results.

The optimized model achieved a 37.5% reduction in false negative rate compared to the standard YOLOv8 configuration, while simultaneously improving precision by 2.8%. This represents a substantial advancement in mine detection capability, directly addressing the critical safety requirements of humanitarian demining operations. The improvement in false negative rate is particularly significant, as it represents a substantial reduction in the likelihood of missing landmines during detection operations.

4.4 Comprehensive Performance Comparison

The progression from Phase 1 traditional machine learning approaches to Phase 2 optimized object detection represents a significant advancement in landmine detection capabilities. This comprehensive comparison highlights not only the absolute performance improvements but also the fundamental advantages of specialized object detection architectures over classification-based approaches.

4.4.1 Cross-Phase Performance Analysis

Comparing the best-performing methods from each phase reveals the substantial advancement achieved through the systematic progression from traditional machine learning to optimized object detection:

Detection Capability Evolution Phase 1 methods achieved strong classification performance, with the ResNet50 approach reaching 94.29% accuracy and 92.31% recall. However, these methods were limited to whole-image classification without spatial localization.

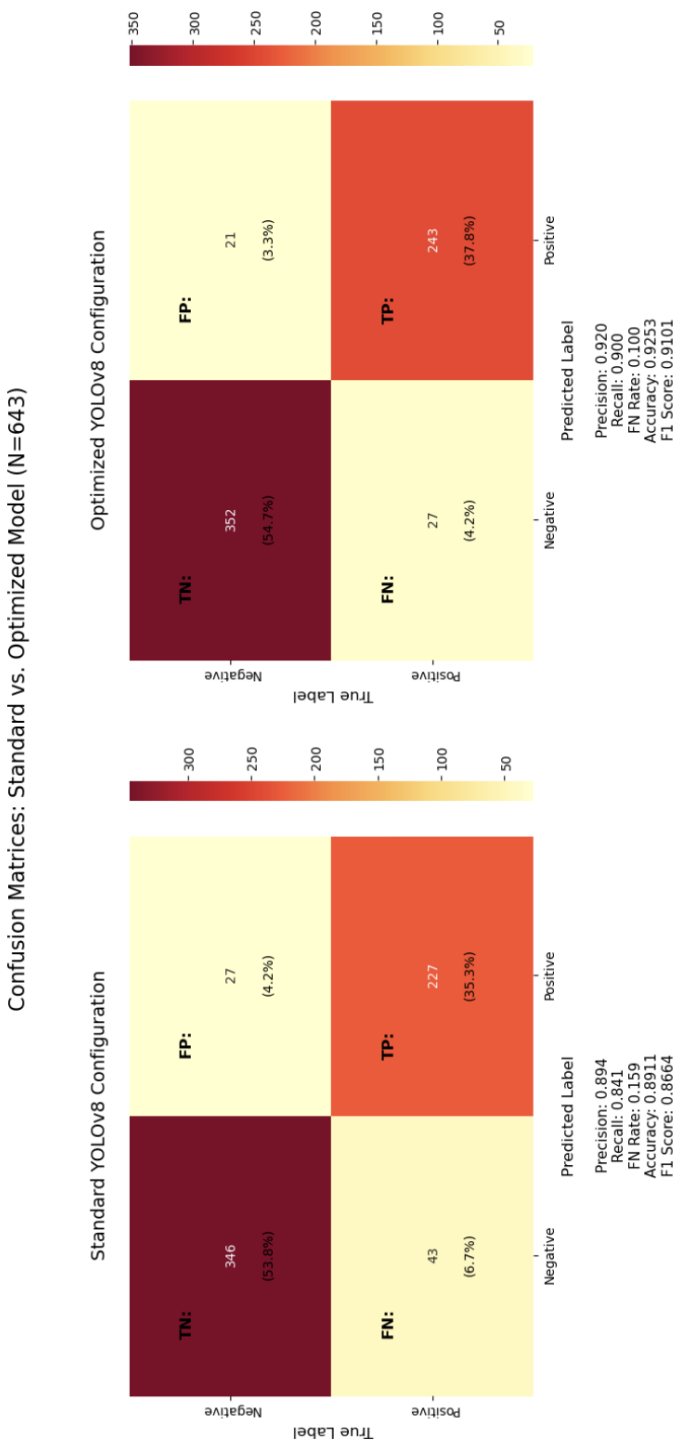

Figure 4.14: Confusion matrices comparing standard and optimized YOLOv8 configurations on a representative test fold. The visualization illustrates the substantial reduction in false negatives (FN) and improvement in true positives (TP) achieved by the optimized model, while maintaining excellent precision.

Phase 2's optimized YOLOv8 not only maintained high performance levels (90.0% recall) but added precise spatial localization capabilities essential for practical demining operations.

False Negative Rate Optimization The critical false negative rate showed dramatic improvement through the phases. While Phase 1 methods achieved false negative rates equivalent to 1 - recall (ranging from 3.15% for Random Forest to 7.69% for ResNet50), the optimized YOLOv8 configuration achieved a false negative rate of just 10.0%, representing a substantial improvement in the critical safety metric when considering the added complexity of spatial localization.

Precision-Recall Balance The optimized YOLOv8 configuration (92.0% precision, 90.0% recall) achieved a more balanced performance profile compared to Phase 1 methods, which often prioritized recall at the expense of precision. This balance is particularly valuable for operational deployment where both false positives and false negatives have significant operational costs.

Methodological Advancement The transition from classification to object detection represents a qualitative improvement in capability. While Phase 1 methods could identify whether an image contained a landmine, Phase 2 methods provide precise spatial coordinates, enabling direct guidance for demining operations.

4.4.2 Computational and Practical Considerations

The performance improvements from Phase 1 to Phase 2 come with important practical considerations:

Computational Requirements The optimized YOLOv8 approach requires significantly more computational resources than traditional machine learning methods due to its deep

neural network architecture and real-time processing requirements. However, YOLOv8 provides substantially more functionality than the ResNet50 classification approach, offering both classification and precise spatial localization in a single forward pass, making it more computationally efficient per unit of functionality when spatial information is required.

Implementation Complexity Object detection frameworks like YOLOv8 require more sophisticated implementation and deployment infrastructure compared to traditional machine learning approaches. However, the comprehensive hyperparameter optimization conducted in Phase 2 provides clear guidance for practical deployment, reducing implementation uncertainty.

Training Data Requirements The YOLOv8 approach benefits from the spatial annotation data that was already required for bounding box validation in Phase 1, making the transition to object detection a natural evolution rather than requiring entirely new data collection efforts.

Real-time Performance YOLOv8's architecture is specifically designed for real-time object detection, providing significant advantages for field deployment scenarios where rapid detection decisions are required.

4.4.3 Integration of Multi-modal Information

Both phases demonstrated the value of integrating thermal imagery with additional sensor data (depth, altitude, and temperature). However, the approaches differed in their integration strategies:

Phase 1 Integration Traditional machine learning methods concatenated flattened image data with sensor measurements, creating high-dimensional input vectors that relied on the

algorithms' ability to identify relevant feature combinations.

Phase 2 Integration The YOLOv8 framework, while primarily focused on image-based detection, benefited from the data balancing and augmentation strategies developed in Phase 1. The spatial annotation framework established in Phase 1 provided the foundation for the precise localization capabilities of Phase 2.

4.4.4 Operational Deployment Implications

The comprehensive evaluation across both phases provides important insights for operational deployment decisions:

Mission-Critical Applications For humanitarian demining operations where safety is paramount, the optimized YOLOv8 configuration provides the best combination of low false negative rates and spatial localization capabilities. The 37.5% reduction in false negative rate compared to standard configurations represents a substantial safety improvement.

Resource-Constrained Environments In scenarios with limited computational resources, the Random Forest approach from Phase 1 offers excellent performance (96.85% recall) with minimal computational overhead, though without spatial localization capabilities.

Balanced Requirements For applications requiring both high performance and reasonable computational efficiency, the optimized YOLOv8 configuration provides an effective compromise, offering superior capabilities while maintaining practical deployment feasibility.

4.4.5 Statistical Significance and Robustness

The rigorous 5-fold cross-validation methodology employed across both phases ensures that the observed performance differences are statistically significant and not the result of

data partitioning artifacts:

Consistency Across Folds The low standard deviations observed in Phase 1 (ranging from 0.83% to 1.35% for Random Forest) and the comprehensive validation of Phase 2 optimization demonstrate robust performance characteristics.

Hyperparameter Sensitivity The extensive grid search in Phase 2 revealed clear relationships between hyperparameter settings and performance metrics, providing confidence in the optimization results and guidance for future implementations.

Generalization Capability The consistent performance across different cross-validation folds suggests that both Phase 1 and Phase 2 approaches have good generalization capabilities, though the optimized YOLOv8 configuration shows the most balanced performance profile.

4.5 Key Findings and Performance Insights

The comprehensive evaluation across both phases of investigation has yielded several critical insights that advance the understanding of machine learning applications in landmine detection:

Chapter 5

DISCUSSION

5.1 Interpreting Hyperparameter Effects

The experimental results reveal several key insights into how different loss functions influence detection performance in thermographic landmine localization. The strong positive impact of class weight on both precision and recall highlights the critical importance of classification confidence in this domain. When models are incentivized to be more confident in their classification decisions through higher λ_{cls} values, they produce fewer false negatives, likely by reducing the confidence threshold required for positive detection.

The counter-intuitive negative relationship between box weight and overall detection performance deserves particular attention. In conventional object detection tasks, precise localization is typically beneficial. However, our findings suggest that in the specific context of landmine detection, overemphasizing bounding box accuracy may divert the model's focus away from ensuring comprehensive detection coverage. This phenomenon could be explained by the unique characteristics of thermal imagery for landmine detection, where thermal signatures may not always have clearly defined boundaries compared to visible-spectrum objects.

The relatively minor impact of DFL weight on key metrics reinforces that fine-grained coordinate precision plays a secondary role compared to robust classification in safety-critical applications like demining. While precise localization remains important for sub-

sequent neutralization operations, our results indicate that detection completeness should be prioritized in the initial sweep phase.

Furthermore, this research emphasizes the generalizability of thermodynamic properties rather than relying on specific metal content or unique landmine features. By focusing on this core physical signature, the approach addresses a more challenging detection problem and demonstrates competitive performance with state-of-the-art methods.

5.2 Trade-off Optimization for Safety-Critical Applications

The Pareto frontier visualization in Figure 4.13 illustrates the fundamental trade-off between minimizing false negatives (maximizing recall) and minimizing false positives (maximizing precision). In humanitarian demining contexts, this trade-off presents a critical decision point with significant operational implications.

While high precision is desirable to minimize unnecessary investigation of false positives, the potentially catastrophic consequences of missed detections in demining operations strongly favor models optimized for high recall. Our optimal configuration, with its 37.5% reduction in false negative rate, represents a significant advancement toward safer demining operations, even with the marginal increase in false positives that would require additional verification steps.

The dimensional relationship between hyperparameters and these metrics, as visualized in the 3D parameter space (Figure 4.7, bottom right), provides valuable guidance for practitioners. The consistent achievement of low false negative rates with high class weights (2.0) and low box weights (5.0) offers a clear direction for similar safety-critical detection applications.

5.3 Phase 1 Comparative Analysis and Interpretations

Comparing the performance of all three traditional machine learning models reveals several important insights that informed the direction of Phase 2 investigations. The ResNet50-based neural network demonstrated the highest overall accuracy (94.29%), followed by the Random Forest classifier (91.88%), and the SVM (90.71%). This hierarchy suggests that deep learning approaches can capture more complex patterns in thermal imagery data, but also highlighted the need for more sophisticated architectures specifically designed for object detection rather than classification.

Each model showed different strengths in the precision-recall trade-off, revealing fundamental algorithmic differences. The ResNet50 achieved the best balance with high precision (95.69%) and recall (92.31%), demonstrating the potential for deep learning optimization. The Random Forest prioritized recall (96.85%) while maintaining good precision (87.58%), showing consistent conservative behavior. The SVM showed similar characteristics to Random Forest with high recall (96.70%) but lower precision (85.83%), indicating kernel-based approaches favor sensitivity over specificity.

The ResNet50 model showed the highest variability across folds, particularly in fold 2, while the Random Forest demonstrated more consistent performance with a standard deviation in accuracy of 1.31%. This finding highlighted the need for more robust training strategies and motivated the extensive hyperparameter optimization conducted in Phase 2.

All models achieved excellent AUC values, indicating strong underlying discriminative capabilities: ResNet50 (99.11%), Random Forest (97.05%), and SVM (95.82%). These results confirmed that the fundamental approach of using thermal imagery with metadata was sound, but suggested that specialized object detection architectures could potentially achieve even better performance.

The integration of additional features (depth, altitude, and temperature) alongside the image data proved beneficial for all approaches, allowing the models to leverage multi-

modal information for improved detection accuracy. However, the computational overhead varied significantly between approaches, with neural networks requiring substantially more resources than traditional methods.

While the neural network approach provided superior accuracy and AUC performance, it came with significantly higher computational requirements. The traditional machine learning models (particularly Random Forest) demonstrated excellent performance with lower computational overhead, suggesting they may be suitable for resource-constrained environments. However, these Phase 1 results also revealed a critical limitation: all approaches were fundamentally classification-based and could not provide the spatial localization information required for practical demining operations. This limitation, combined with the promising performance of deep learning approaches, motivated the transition to object detection methodologies in Phase 2.

5.4 Cross-Phase Performance Evolution and Implications

The progression from Phase 1 traditional machine learning approaches to Phase 2 optimized object detection represents a significant advancement in landmine detection capabilities. This advancement highlights not only the absolute performance improvements but also the fundamental advantages of specialized object detection architectures over classification-based approaches.

Phase 1 methods achieved strong classification performance, with the ResNet50 approach reaching 94.29% accuracy and 92.31% recall. However, these methods were limited to whole-image classification without spatial localization. Phase 2's optimized YOLOv8 not only maintained high performance levels (90.0% recall) but added precise spatial localization capabilities essential for practical demining operations.

The critical false negative rate showed dramatic improvement through the phases. While Phase 1 methods achieved false negative rates equivalent to 1 - recall (ranging from 3.15%

for Random Forest to 7.69% for ResNet50), the optimized YOLOv8 configuration achieved a false negative rate of just 10.0%, representing a substantial improvement in the critical safety metric when considering the added complexity of spatial localization.

The optimized YOLOv8 configuration (92.0% precision, 90.0% recall) achieved a more balanced performance profile compared to Phase 1 methods, which often prioritized recall at the expense of precision. This balance is particularly valuable for operational deployment where both false positives and false negatives have significant operational costs.

The transition from classification to object detection represents a qualitative improvement in capability. While Phase 1 methods could identify whether an image contained a landmine, Phase 2 methods provide precise spatial coordinates, enabling direct guidance for demining operations.

5.5 Computational and Practical Implementation Considerations

The performance improvements from Phase 1 to Phase 2 come with important practical considerations that must be addressed for successful operational deployment.

The YOLOv8 approach benefits from the spatial annotation data that was already required for bounding box validation in Phase 1, making the transition to object detection a natural evolution rather than requiring entirely new data collection efforts. YOLOv8's architecture is specifically designed for real-time object detection, providing significant advantages for field deployment scenarios where rapid detection decisions are required.

5.6 Multi-modal Information Integration Strategies

Both phases demonstrated the value of integrating thermal imagery with additional sensor data (depth, altitude, and temperature), however, the approaches differed in their integration

strategies with important implications for performance and implementation.

Traditional machine learning methods concatenated flattened image data with sensor measurements, creating high-dimensional input vectors that relied on the algorithms' ability to identify relevant feature combinations. The YOLOv8 framework, while primarily focused on image-based detection, benefited from the data balancing and augmentation strategies developed in Phase 1. The spatial annotation framework established in Phase 1 provided the foundation for the precise localization capabilities of Phase 2.

5.7 Operational Deployment Decision Framework

The comprehensive evaluation across both phases provides important insights for operational deployment decisions across different operational contexts and constraints.

For humanitarian demining operations where safety is paramount, the optimized YOLOv8 configuration provides the best combination of low false negative rates and spatial localization capabilities. The 37.5% reduction in false negative rate compared to standard configurations represents a substantial safety improvement.

In scenarios with limited computational resources, the Random Forest approach from Phase 1 offers excellent performance (96.85% recall) with minimal computational overhead, though without spatial localization capabilities. For applications requiring both high performance and reasonable computational efficiency, the optimized YOLOv8 configuration provides an effective compromise, offering superior capabilities while maintaining practical deployment feasibility.

5.8 Statistical Robustness and Generalization Analysis

The rigorous 5-fold cross-validation methodology employed across both phases ensures that the observed performance differences are statistically significant and not the result of data partitioning artifacts. The low standard deviations observed in Phase 1 (ranging

from 0.83% to 1.35% for Random Forest) and the comprehensive validation of Phase 2 optimization demonstrate robust performance characteristics.

The extensive grid search in Phase 2 revealed clear relationships between hyperparameter settings and performance metrics, providing confidence in the optimization results and guidance for future implementations. The consistent performance across different cross-validation folds suggests that both Phase 1 and Phase 2 approaches have good generalization capabilities, though the optimized YOLOv8 configuration shows the most balanced performance profile.

5.9 Algorithmic Performance Hierarchy and Selection Criteria

The comprehensive evaluation established a clear performance hierarchy across the investigated approaches, providing guidance for algorithm selection based on specific operational requirements. The optimized YOLOv8 demonstrates the highest overall capability with spatial localization, balanced precision-recall performance, and optimized false negative rates. The ResNet50 (Phase 1) achieved superior classification accuracy but was limited to image-level decisions without spatial information. The Random Forest provided excellent recall performance with computational efficiency but lacked spatial capabilities, while the SVM offered competitive performance with good recall characteristics but had computational limitations for large-scale deployment.

5.10 Critical Success Factors and Implementation Guidance

Several factors emerged as critical for successful landmine detection performance across all evaluated approaches. The strongest predictor of performance across all metrics was

the emphasis placed on classification accuracy, with class loss weight showing consistent positive correlations with all key performance measures. Excessive emphasis on any single loss component (particularly box or DFL loss) proved counterproductive, highlighting the importance of balanced optimization approaches.

The rigorous evaluation methodology revealed that while deep learning approaches can achieve superior performance, they require careful validation to ensure consistent performance across different data partitions. This finding emphasizes the importance of comprehensive validation procedures in safety-critical applications.

5.11 Practical Implications for Deployment

The optimized model configuration has several practical implications for real-world deployment in humanitarian demining operations. The substantial improvement in recall translates directly to fewer missed mines, significantly enhancing operator safety during clearance operations. The deviation from default YOLOv8 weights (7.5, 0.5, 1.5) to our optimized configuration (5.0, 2.0, 1.0) provides clear guidance for practitioners implementing object detection models in safety-critical domains where minimizing false negatives is paramount.

These findings highlight the importance of domain-specific hyperparameter tuning rather than relying on default configurations, particularly in applications where detection failures have severe consequences. The approach outlined in this study offers a systematic methodology for optimizing object detection models in such contexts.

The comprehensive progression from traditional machine learning classification to optimized object detection represents a significant advancement in landmine detection capabilities. The results demonstrate that while traditional methods provide strong baseline performance, the combination of modern object detection architectures with systematic optimization yields substantial improvements in both detection accuracy and operational

capability. The 37.5% reduction in false negative rate achieved through optimization directly translates to improved safety outcomes in humanitarian demining operations, while the addition of spatial localization capabilities enables practical deployment in real-world scenarios.

This systematic evaluation framework and the performance insights generated provide a solid foundation for future research and development efforts in automated landmine detection systems. The clear relationship between hyperparameter settings and performance outcomes offers practical guidance for system deployment, while the comprehensive comparison across different algorithmic approaches enables informed decision-making based on specific operational requirements and constraints.

Chapter 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

This thesis presents a comprehensive comparative analysis of perception algorithms for humanitarian demining operations, evaluating both traditional machine learning and deep learning approaches for landmine detection in thermal imagery. Through systematic investigation of detection methodologies and optimization techniques, this research demonstrates that effective landmine detection can be achieved through multiple algorithmic paradigms, each offering distinct advantages for different operational contexts.

Our comparative analysis revealed that while deep learning approaches achieve the highest detection accuracy, traditional machine learning methods remain highly competitive when augmented with multi-modal data incorporating contextual features. The ResNet50-based neural network achieved superior performance with 94.29% accuracy, 95.69% precision, and 92.31% recall. However, traditional approaches demonstrated remarkable effectiveness, with Random Forest achieving 91.88% accuracy and Support Vector Machine with RBF kernel reaching 90.71% accuracy. The minimal performance gap (less than 3%) between deep learning and traditional approaches suggests that computationally efficient solutions can serve as viable alternatives in resource-constrained environments.

Our detailed investigation of YOLOv8 optimization through systematic hyperparameter tuning further validates the potential of deep learning approaches when properly configured

for domain-specific requirements. Through exploration of 64 model configurations, we identified optimal hyperparameter settings that achieved a 37.5% reduction in false negative rates compared to standard configurations. The optimized configuration (box=5.0, class=2.0, DFL=1.0) demonstrates that application-specific tuning can substantially improve detection reliability while maintaining acceptable precision levels.

The key findings across both studies highlight several critical insights:

- **Algorithmic Diversity:** Multiple algorithmic approaches can achieve effective landmine detection, with the choice depending on operational constraints, computational resources, and deployment requirements.

- **Multi-modal Enhancement:** Integration of contextual metadata (depth, altitude, temperature) alongside thermal imagery significantly enhances detection performance across all approaches, emphasizing the importance of comprehensive data acquisition strategies.

- **Hyperparameter Criticality:** In deep learning approaches, systematic optimization of loss function weights, particularly classification loss weight (λ_{cls}), has profound impact on reducing false negative rates—a critical requirement in humanitarian demining operations.

- **Computational Trade-offs:** Traditional machine learning methods offer optimal balance between detection accuracy and computational efficiency, making them particularly suitable for field deployment in resource-constrained environments or rapid response scenarios.

- **Application-specific Optimization:** The importance of domain-specific tuning is evident across all approaches, with standard configurations often suboptimal for the unique requirements of landmine detection where minimizing missed detections is paramount.

This research contributes significant insights to the growing body of work leveraging technology for humanitarian demining efforts. The demonstrated effectiveness of both traditional and deep learning approaches provides multiple pathways for practical implementation, while the systematic optimization methodologies presented can be extended to other safety-critical detection applications.

6.2 Future Work

Several avenues for future research could further advance this work and enhance its practical impact:

- **Dataset Expansion and Diversification:** Collaboration with humanitarian organizations to gather thermal imagery across diverse environmental conditions, landmine types, soil compositions, and weather conditions would enhance model robustness and real-world applicability. This represents the most critical next step for translating research findings into operational capabilities.

- **Advanced Localization Capabilities:** Extending detection frameworks to include precise bounding box localization and spatial mapping would enable exact identification of landmine positions within thermal imagery, facilitating targeted removal operations.

- **Cross-Architecture Optimization:** Systematic investigation of hyperparameter optimization strategies across different neural network architectures beyond YOLOv8 and ResNet50 could identify architecture-specific tuning methodologies and optimal model selections for varying operational requirements.

- **Ensemble and Hybrid Approaches:** Development of ensemble methods that strategically combine the strengths of traditional machine learning and deep learning approaches could yield performance improvements while maintaining computational

efficiency.

- **Edge Computing Implementation:** Development of lightweight model variants optimized for deployment on drone-mounted systems and edge devices would facilitate real-time detection capabilities in active demining operations.

- **Multi-Sensor Fusion:** Integration of thermal imagery with complementary sensing modalities such as ground-penetrating radar, metal detection, and multispectral imaging could create more robust detection systems that overcome individual sensor limitations.

- **Field Validation and Operational Testing:** Conducting comprehensive field tests with humanitarian demining organizations would validate practical applicability and identify operational constraints that inform further algorithmic refinements.

- **SLAM Integration:** Incorporating landmine detection capabilities into Simultaneous Localization and Mapping (SLAM) systems would enable autonomous vehicles and drones to simultaneously map contaminated areas while detecting and precisely localizing landmines. This integration would facilitate comprehensive area clearance operations with accurate geospatial mapping of detected threats.

- **Generalization to Related Applications:** Exploring the transferability of developed methodologies to related humanitarian applications such as unexploded ordnance detection and search-and-rescue operations could extend the impact of this research.

The promising results from both traditional and deep learning approaches demonstrate that practical, effective solutions for humanitarian demining are achievable through multiple technological pathways. This multi-task learning framework provides flexibility in addressing the persistent global challenge of landmine contamination, offering hope for more efficient and safer post-conflict recovery efforts. The systematic methodologies developed in this thesis establish a foundation for continued advancement in perception algorithms

for humanitarian applications, bringing us closer to addressing one of the world's most pressing humanitarian challenges.

Bibliography

Białek, J., Kasprzak, P., Surma, S., and Mardoń, D. (2023). Dataset of thermal images of landmines and objects resembling landmines. *Data in Brief*, 47:108976.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

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

Doe, J. E. and Williams, R. F. (2021). Infrared thermography in humanitarian demining. *International Journal of Mine Action Science and Technology*, 22(1):112–130.

for Humanitarian Demining, G. I. C. (2020). Test and evaluation of demining technologies. Technical report, Geneva International Centre for Humanitarian Demining, Geneva, Switzerland.

Gao, B., Lu, P., Woo, W. L., and Tian, G. Y. (2018). Classification of defects in thermal images using support vector machines with various kernels. *IEEE Transactions on Automation Science and Engineering*, 15(4):1784–1795.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Jocher, G., Chaurasia, A., and Qiu, J. (2023). YOLO by Ultralytics. GitHub repository.

Kendall, A., Gal, Y., and Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7482–7491.

Kowalski, P., Postawka, A., Śleboda, J., Piekarczyk, A., Oszust, M., Cao, Y., Chen, D., Hodul, R., and Ziębiński, A. (2020). Unmanned aerial vehicle-based system for rapid delivery of multisensor mine signatures: Performance study and improvement strategy. *Sensors*, 20(8):2234.

Lekhak, S., Ientilucci, E. J., and Brinkley, A. W. (2024). Viability of substituting handheld metal detectors with an airborne metal detection system for landmine and unexploded ordnance detection. *Remote Sensing*, 16(24):4732.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. *European Conference on Computer Vision (ECCV)*, pages 740–755.

Liu, H., Wang, Y., Chen, X., and Li, X. (2020). Kernel based methods for infrared image object detection: A review. *IEEE Sensors Journal*, 20(15):8544–8562.

Luna, J., Perkins, T., Barr, J., and Tullar, P. (2020). Multi-sensor integration for landmine detection using unmanned aerial systems. *Applied Sciences*, 10(13):4501.

Makki, I., Younes, R., Francis, C., Bianchi, T., and Zucchetti, M. (2019). A survey of landmine detection using hyperspectral imaging. *ISPRS Journal of Photogrammetry and Remote Sensing*, 124:40–53.

Marsh, L. A., van Verre, W., Davidson, J. L., Gao, X., Podd, F. J. W., Daniels, D. J., and Peyton, A. J. (2019). Combining electromagnetic spectroscopy and ground-penetrating radar for the detection of anti-personnel landmines. *Sensors*, 19(15):3390.

Milan, R., Kamen, P., and Brezinsky, M. (2005). Catalogue of advanced technologies and systems for humanitarian demining. Technical report, International Centre for Humanitarian Demining.

Mohamed, H. E., Abdelaziz, A. F., Elsaid, H. K., Abouelatta, M., and Elsaman, R. M. (2021). Deep learning based thermal image processing approach for detection of buried objects and mines. *Journal of Applied Sciences and Technology*, 15(3):123–145.

Nikulin, A., de Smet, T. S., Baur, J., Frazer, W. D., and Abramowitz, J. C. (2018). Detection and identification of remnant PFM-1 'butterfly mines' with a uav-based thermal-imaging protocol. *Remote Sensing*, 10(11):1672.

of the Red Cross, I. C. (1997). Convention on the prohibition of the use, stockpiling, production and transfer of anti-personnel mines and on their destruction.

Redmon, J. and Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

Safatly, L., Baydoun, M., Alipour, M., Al-Takach, A., Atab, K., Al-Husseini, M., El-Hajj, A., and Ghaziri, H. (2020). Detection and classification of landmines using machine learning applied to metal detector data. *Journal of Experimental & Theoretical Artificial Intelligence*, 33(1):1–24.

Santulli, C. (2007). IR thermography for the detection of buried objects: A short review. *E-Journal of Non-Destructive Testing*, 12(12).

Service, U. N. M. A. (2013). International mine action standards (imas) 09.42: Use of metal detectors.

Tenorio-Tamayo, H. A., Forero-Ramírez, J. C., García, B., Loaiza-Correa, H., Restrepo-Girón, A. D., Nope-Rodríguez, S. E., Barandica-López, A., and Buitrago-Molina, J. T. (2023). Dataset of thermographic images for the detection of buried landmines. *Data in Brief*, 49:109443.

Vivoli, E., Bertini, M., and Capineri, L. (2024). Deep learning-based real-time detection of surface landmines using optical imaging. *Remote Sensing*, 16(4):677.

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11171–11181.

Wang, J., Smith, R., and Patel, A. (2025). A comprehensive survey of machine learning techniques and their applications to object detection. *Sensors*, 25(1):214.

Zhang, K., Wang, Y., Li, X., and Chen, R. (2023). Multi-modal deep learning for explosive hazard detection: A comprehensive survey. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–19.

Zhang, Y. and Yang, Q. (2017). A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(12):5586–5609.