



UNIVERSITI
TEKNOLOGI
MARA

ISETC2025

2ND INTERNATIONAL SCIENCE,
ENGINEERING AND TECHNOLOGY
COLLOQUIUM



THE POWER OF SCIENTIFIC INNOVATION

INTELLIGENT SOLUTIONS FOR A
SUSTAINABLE WORLD

ORGANIZED BY:
FACULTY OF APPLIED SCIENCES
UNIVERSITI TEKNOLOGI MARA
PERAK BRANCH, TAPAH CAMPUS
MALAYSIA

CO-ORGANIZED BY:



**2nd International Science, Engineering and Technology Colloquium
(ISETC 2025)**

Faculty of Applied Sciences,
Universiti Teknologi MARA,
Perak Branch, Tapah Campus,
35400 Tapah Road, Perak, Malaysia.

Colloquium date: 9th July 2025

Publication date: 9th September 2025

Proceedings of Extended Abstracts

EDITORS

Dr. Emi Norzehan Mohamad Mahbob

Dr. Saiyidah Nafisah Hashim

Ms. Lili Widarti Zainuddin

Ms. Nurul Nazwa Mohammad

Ms. Rozaina Saleh

e-ISBN: 978-629-97630-5-5

Copyright © 2025 ISETC

Published by Faculty of Applied Sciences, Universiti Teknologi MARA,
Perak Branch, Tapah Campus, 35400 Tapah Road, Perak, Malaysia.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the Faculty of Applied Sciences, Universiti Teknologi MARA, Perak Branch, Tapah Campus, 35400 Tapah Road, Perak, Malaysia.

PREFACE

Welcome to the 2nd International Science, Engineering and Technology Colloquium 2025 (ISETC 2025)!

It is with great pleasure that we present this year's colloquium, a distinguished gathering of scholars, researchers, and professionals dedicated to the advancement of applied sciences. Held virtually on 9th July 2025, ISETC 2025 serves as a platform to highlight cutting-edge research, foster meaningful networking, encourage collaborations, and explore innovative solutions to global challenges.

This year's theme, "The Power of Scientific Innovation: Intelligent Solutions for a Sustainable World" highlights our shared goal to solve global challenges through applied science. It celebrates innovation as a key to building a more sustainable and resilient future. We're proud to feature keynote speakers and presenters who bring fresh insights, bold ideas, and breakthroughs shaping tomorrow's world.

We would like to extend our heartfelt appreciation to the Faculty of Applied Sciences, Universiti Teknologi MARA, Perak Branch, Tapah Campus, for their outstanding support and assistance throughout the planning and execution of the event. Their dedication has been instrumental to the success of this colloquium. We would like to sincerely thank all participants and researchers for their valuable contributions. We also appreciate the reviewers for their careful evaluations and helpful feedback. Your hard work and support mean a lot to us.

As editors, we are proud to present this collection of proceedings. It is an honour to bring together the interesting and meaningful work of our speakers and authors. We hope the ideas shared will lead to new discoveries, encourage more research, and help solve global challenges through science and innovation.

Warm regards,

Dr. Emi Norzehan Mohamad Mahbob

Dr. Saiyidah Nafisah Hashim

Ms. Lili Widarti Zainuddin

Ms. Nurul Nazwa Mohammad

Ms. Rozaina Saleh

ISETC 2025 Proceedings Editors

SemanticForestMY: A HIGH-RESOLUTION REMOTE SENSING DATASET FOR FOREST AND NON-FOREST SEMANTIC SEGMENTATION

Imran Md Jelas^{1,2*}, Mohd Asyraf Zulkifley², Mardina Abdullah²

¹*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch Tapah Campus,
Perak, Malaysia*

²*Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Selangor, Malaysia*

**Corresponding author: imran499@uitm.edu.my*

Abstract: Accurate forest and non-forest segmentation remains a challenge in environmental monitoring due to limitations in ground truth quality and dataset standardization. This study introduces SemanticForestMY, a high-resolution remote sensing dataset developed from Google Earth Pro imagery across three key Malaysian ecosystems. A systematic annotation process combining automatic, assisted, and manual labeling techniques was employed to produce high-quality semantic segmentation masks. U-Net model evaluations achieved a validation accuracy of 93.10% and IoU of 91.36%, confirming dataset reliability. SemanticForestMY addresses existing dataset limitations and provides a solid foundation for future research in deep learning-based forest monitoring applications. Future work includes expanding the dataset with additional classes and temporal resolutions.

Keywords: Remote sensing, forest segmentation, deep learning, semantic segmentation, high-resolution dataset

INTRODUCTION

Accurate and consistent semantic segmentation of forest and non-forest areas plays a critical role in environmental monitoring, biodiversity conservation, and climate change mitigation. Despite the availability of various remote sensing datasets, challenges persist due to limited resolution, inconsistent ground truth quality, and non-standardized partitioning methods. In response, SemanticForestMY introduces a high-resolution, systematically constructed dataset tailored specifically for binary semantic segmentation tasks in forest monitoring. Building upon recent advancements in automatic and assisted ground truth annotation (Md Jelas et al., 2023; Md Jelas, Zulkifley, & Abdullah, 2024), this dataset provides a robust foundation for developing and benchmarking deep learning models in remote sensing applications.

MATERIALS AND METHODS

The SemanticForestMY dataset was constructed to address the critical need for a high-resolution, high-quality forest and non-forest segmentation benchmark using remote sensing imagery. Satellite images were acquired from Google Earth Pro for three ecologically significant regions in Malaysia: Endau Rompin (2°30'42.20"N, 103°32'56.39"E), Royal Belum (5°47'9.93"N, 101°30'51.44"E), and Tasik Chini (3°22'32.82"N, 102°36'11.18"E). For each location, imagery from the years 1990, 2000, and 2010 were collected. These images, stored in PNG format with maximum resolutions of 8192 × 4883 pixels, cover large areas with a spatial extent of approximately 137,500 meters.

Ground truth annotations were generated through a semi-automated process that began with green channel extraction, followed by Gaussian blur and thresholding using both Otsu's and global methods. Subsequently, OpenCV's SimpleBlobDetector was applied to filter noise. Manual refinement was conducted to correct edge inconsistencies and small artifacts. This hybrid approach enhanced mask quality while preserving annotation efficiency.

Each annotated image was divided into 224 × 224 pixel tiles, yielding 756 tiles per area and time point. To ensure robust training and reproducibility, the dataset was partitioned using a carefully designed random_state assignment strategy that restricted repeated values and enforced statistical spacing. The dataset was split into 4,761 training tiles,

675 validation tiles, and 1,368 testing tiles using the train_test_split function. No data augmentation techniques were applied to maintain the originality of the data distribution.

To validate segmentation effectiveness, a baseline U-Net model was trained with input size $224 \times 224 \times 3$, using binary cross-entropy loss and the Adam optimizer (learning rate = 0.001). The model was trained for 100 epochs with a batch size of 16. Performance was evaluated using standard metrics including accuracy, precision, recall, F1-score, and intersection over union (IoU). This setup demonstrates that SemanticForestMY provides a rigorous, reliable benchmark for forest segmentation research. The overall dataset construction process is visualized in Figure 1.

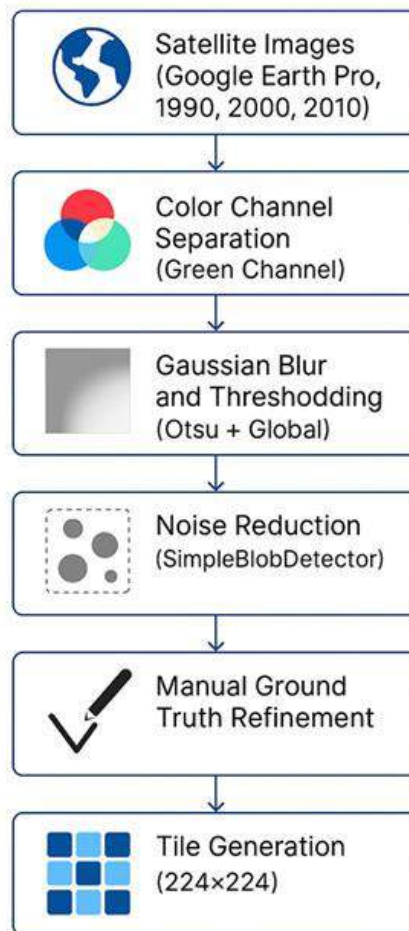


Figure 1. SemanticForestMY dataset construction pipeline

Additionally, an example illustrating a satellite image and its corresponding ground truth mask is presented in Figure 2, demonstrating the annotation quality and segmentation clarity achieved in the SemanticForestMY dataset.

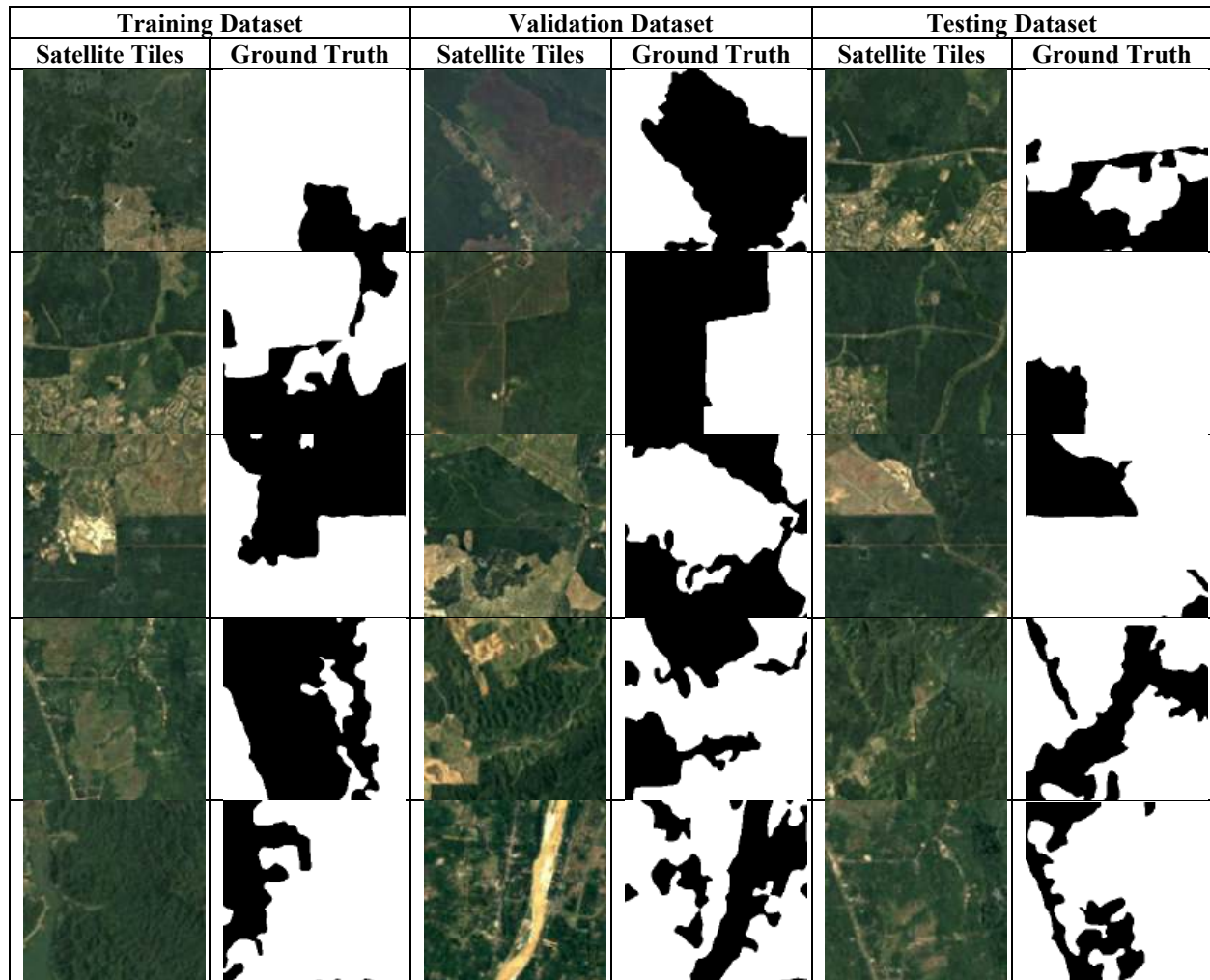


Figure 2. Representative examples of satellite images and corresponding ground truth masks from the SemanticForestMY dataset.

RESULTS AND DISCUSSION

To assess the segmentation quality and reliability of the SemanticForestMY dataset, a U-Net model was implemented and evaluated using the complete training and testing partitions. The outcomes of this baseline evaluation are presented in Table 1, summarizing key performance metrics achieved during the training and validation processes.

Table 1. U-Net segmentation performance on SemanticForestMY dataset

Metric	Accuracy	Loss	Precision	Recall	F1 Score	IoU	Runtime
Training (%)	92.94	17.97	94.13	96.38	95.24	90.92	2h 55m 7s
Validation (%)	93.10	17.56	94.02	97.02	95.48	91.36	

Note: Validation performance indicates model generalizability. Results are based on 4,761 training tiles and 675 validation tiles from three spatio-temporal regions.

The U-Net model attained a validation accuracy of 93.10%, a validation F1-Score of 95.48%, and an Intersection over Union (IoU) of 91.36%, reflecting the dataset's ability to support high-fidelity semantic segmentation tasks. The consistency observed between training and validation performances with only minor variations in loss values suggests that the SemanticForestMY dataset possesses strong generalization properties and is free from significant overfitting tendencies.

Several key factors contribute to these favorable outcomes. The use of high-resolution satellite imagery from Google Earth Pro ensured a rich spatial detail that enabled models to learn subtle distinctions between forest and non-forest areas. Additionally, the semi-automated annotation pipeline, incorporating Gaussian blur smoothing, Otsu and global thresholding, and noise filtering through SimpleBlobDetector, provided high-quality and consistent ground truth masks. The strict dataset partitioning strategy based on non-repetitive, statistically diverse random_state values further supported unbiased model evaluation, preserving the integrity of training, validation, and testing splits.

Comparative evaluations against prior datasets in deforestation monitoring (Md Jelas, Zulkifley, Abdullah, et al., 2024) highlight the superiority of SemanticForestMY in achieving cleaner spatial segmentation and stronger performance stability. Notably, the high recall rate of 97.02% indicates that the model was particularly effective at correctly identifying forest regions, minimizing false negatives, which is critical in deforestation and conservation studies.

Interestingly, despite the absence of common data augmentation techniques such as rotation, flipping, or scaling, the U-Net model was still able to achieve high performance. This outcome underscores the diversity and richness of the collected satellite imagery across different temporal snapshots and geographic locations, which inherently contributed to model robustness without the need for synthetic sample expansion.

Moving forward, future investigations could explore the application of more advanced semantic segmentation architectures, including DeepLabV3+, ResU-Net, or Transformer-based models, to potentially extract even finer structural features from the dataset. Expanding SemanticForestMY to include additional land cover classes, seasonal variations, and implementing augmentation pipelines may also further enhance model robustness and broaden its applicability to more complex environmental monitoring tasks.

CONCLUSIONS

The development of the SemanticForestMY dataset demonstrates that systematically acquired and carefully annotated high-resolution satellite imagery can significantly improve the performance of semantic segmentation models for forest and non-forest classification tasks. The strong validation metrics achieved by the U-Net model including over 93% validation accuracy and 91% IoU indicate that the principles of consistent data acquisition, hybrid annotation methods, and controlled partitioning contribute to highly generalizable models.

However, several challenges were observed. First, the dataset currently focuses solely on binary classification (forest vs non-forest), which may limit its applicability for more complex land cover studies involving multiple classes. Second, the absence of data augmentation, while preserving data originality, may expose models to slight overfitting in more diverse real-world conditions.

Theoretical implications of this work suggest that datasets constructed with rigorous ground truth strategies and statistically balanced partitioning yield stronger model generalization even without heavy regularization. Practically, SemanticForestMY provides an immediately usable resource for training and benchmarking machine learning models targeting tropical forest monitoring and conservation efforts.

In conclusion, SemanticForestMY addresses a notable gap in the remote sensing community by offering a reproducible, scalable, and high-quality dataset for forest segmentation. Future recommendations include expanding the dataset to multi-class annotations (e.g., distinguishing forest types, water bodies, urban areas), incorporating seasonal variations, applying systematic data augmentation strategies, and testing advanced architectures such as DeepLabV3+, HRNet, and Swin-Transformer-based models to further improve segmentation robustness.

REFERENCES

- Md Jelas, I., Zulkifley, M. A., & Abdullah, M. (2023). Automated Ground Truth Annotation for Forest and Non-Forest Classification in Satellite Remote Sensing Images. *2023 4th International Conference on Artificial Intelligence and Data Sciences: Discovering Technological Advancement in Artificial Intelligence and Data Science, AiDAS 2023 - Proceedings*, 331–336. <https://doi.org/10.1109/AiDAS60501.2023.10284683>
- Md Jelas, I., Zulkifley, M. A., & Abdullah, M. (2024). Forest Segmentation: Spatio-Temporal Ground Truth Labelling via Assisted Annotation. *2024 IEEE 8th International Conference on Signal and Image Processing Applications, ICSIPA 2024*. <https://doi.org/10.1109/ICSIPA62061.2024.10686968>
- Md Jelas, I., Zulkifley, M. A., Abdullah, M., & Spraggon, M. (2024). Deforestation detection using deep learning-based semantic segmentation techniques: a systematic review. In *Frontiers in Forests and Global Change* (Vol. 7). Frontiers Media SA. <https://doi.org/10.3389/ffgc.2024.1300060>



Surat kami : 700-KPK (PRP.UP.1/20/1)
Tarikh : 20 Januari 2023



Prof. Madya Dr. Nur Hisham Ibrahim
Rektor
Universiti Teknologi MARA
Cawangan Perak

Tuan,

**PERMOHONAN KELULUSAN MEMUAT NAIK PENERBITAN UiTM CAWANGAN PERAK
MELALUI REPOSITORI INSTITUSI UiTM (IR)**

Perkara di atas adalah dirujuk.

2. Adalah dimaklumkan bahawa pihak kami ingin memohon kelulusan tuan untuk mengimbas (*digitize*) dan memuat naik semua jenis penerbitan di bawah UiTM Cawangan Perak melalui Repositori Institusi UiTM, PTAR.

3. Tujuan permohonan ini adalah bagi membolehkan akses yang lebih meluas oleh pengguna perpustakaan terhadap semua maklumat yang terkandung di dalam penerbitan melalui laman Web PTAR UiTM Cawangan Perak.

Kelulusan daripada pihak tuan dalam perkara ini amat dihargai.

Sekian, terima kasih.

"BERKHIDMAT UNTUK NEGARA"

Saya yang menjalankan amanah,

Setuju.

27.1.2023

SITI BASRIYAH SHAIK BAHARUDIN
Timbalan Ketua Pustakawan

PROF. MADYA DR. NUR HISHAM IBRAHIM
REKTOR
UNIVERSITI TEKNOLOGI MARA
CAWANGAN PERAK
KAMPUS SERI ISKANDAR

nar