ORIGINAL RESEARCH Open access

Machine Learning-Based Mapping of Sugarcane Burnt Areas in Smallholder Agricultural Systems

Grace Collins[1*], Oliver Evans[2], Noah Williams[1], Ethan Brooks[3]

Abstract

Satellite remote sensing approaches have demonstrated strong capability in accurately measuring burnt areas from various types of fires. Yet they often struggle to detect crop residue-burning zones, especially in smallholder agricultural settings, primarily due to limited access to high-frequency, fine-scale satellite imagery. In the present research, we employed Harmonized Landsat Sentinel-2 (HLS) data. We tested two machine learning classifiers (Support Vector Machines (SVM) and Artificial Neural Networks (ANN)) for mapping sugarcane burnt zones throughout the 2019–20 season in a smallholder-dominated area of Thailand. Findings revealed that both classifiers effectively identified the spatial distribution of sugarcane-burnt zones within the study area. The ANN achieved superior results compared to SVM at both the pixel and regional levels. At the pixel scale, ANN reached an accuracy of 93.4%, while the SVM's top Polynomial kernel attained 82.7%. Regionally, the ANN-derived average burnt area percentage (51.1%) aligned more closely with the official figure of 48.7% provided by Thailand's Office of Cane and Sugar Board (OCSB) than the SVM result of 62.9%. Total burnt area calculations from ANN and SVM (315 and 418 thousand ha, respectively) diverged further from the OCSB reported value (240 thousand ha) than the percentage estimates. Still, the classifiers' area calculations proved markedly more precise than those generated by MODIS burnt area products. In summary, the analysis showed that HLS data provided essential spectral details for constructing robust models that map burnt zones in smallholder environments with improved precision compared to existing global datasets. The ANN-based mapping procedure showed considerable promise for consistent surveillance of sugarcane-burnt areas and can aid the effective rollout of control policies in Thailand.

Keywords Neural networks, Support vector machines, Sentinel-1 SAR, Harmonized Landsat Sentinel-2 data, Land cover mapping

*Correspondence:
Grace Collins
grace.collins@sydney.edu.au
[1] Department of AI for Mango and Papaya Post-Harvest Ripening Prediction, Faculty of Agriculture, University of Sydney, Sydney, Australia
[2] Department of Machine Learning for Fruit Fly and Borer Infestation Models, Faculty of Agriculture, University of Queensland, Brisbane, Australia
[3] Department of Smart Ripening Chamber Control, Faculty of Agriculture, University of Melbourne, Melbourne, Australia

Introduction

In numerous top sugarcane-producing nations, such as Brazil, India, China, Thailand, and the United States of America, the practice of burning crop residues—whether before or after harvest—remains routine [1]. Emissions of air pollutants (e.g., particulate matter) from this activity are associated with major public health risks [2]. Beyond that, they alter the atmospheric composition and drive global climate shifts by disrupting the planet's radiation balance. Additionally, sugarcane residue burning harms soil health by reducing soil organic carbon and nutrients while promoting erosion, potentially leading to prolonged effects on agricultural output [3].

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

Given the substantial threats to human wellbeing and the environment, initiatives have emerged to curb sugarcane residue burning via policy interventions. For example, Thailand's government launched the Green Cane Purchasing policy from the 2019–20 season, mandating that mills restrict purchases of burnt cane to 30 % of overall volume, with a further reduction to 20% planned from the 2020–21 season [4]. In parallel, São Paulo State, Brazil's leading sugarcane hub, passed legislation (Law N◦11.241/2002) to end burning by 2031 progressively [5]. Besides these, multiple projects with mixed outcomes have aimed to improve emission inventories for crop residue burning to inform policy decisions and shape reduction plans [6].

Reliable identification of burnt zones is vital for gauging how effectively regulations limit burning practices. It is equally important to compute biomass emissions using bottom-up methods, which refine the emission inventories used in climate modeling to evaluate how releases influence climate patterns.

Satellite remote sensing has seen broad application in developing techniques to detect active fires and burnt areas [7, 8]. Methods focused on active fire detection generally use thermal anomalies (i.e., brightness-temperature contrasts between suspected fire locations and nearby cover) [9]. These brightness temperatures are derived from radiances measured in thermal infrared bands (e.g., 10–11 μm). Conversely, burnt area detection relies on variations in surface reflectance or spectral indices, such as the Burn Area Index (BAI) and Normalized Burn Ratio (NBR), to differentiate burnt from unburnt surfaces [10, 11].

Global operational datasets are available for active fires (MOD14/MYD14 and VNP14) and burnt areas (MCD64A1), derived from MODIS and VIIRS sensors [7]. Past analyses have frequently highlighted weak results from these datasets when applied to cropland-burnt zones, notably in smallholder contexts where plots are often less than 10 ha [6, 8, 12, 13]. Multiple issues explain this shortfall. Cropland fires tend to be brief, and satellite overpasses rarely coincide with burning periods, limiting the detection of active fire indicators. Moreover, the coarse resolution (≥ 250 m) of the instruments struggles to register spectral shifts caused by tiny fields and varied crop layouts in smallholder operations. Lastly, the underlying land cover datasets for these global products feature only general cropland categories and omit detailed crop-type information needed to pinpoint harvest schedules for specific crops.

Recent advances in satellite technology, particularly Harmonized Landsat and Sentinel-2 (HLS) data [14], merge observations from Landsat 8 and Sentinel-2 to provide frequent optical coverage, often at 4–8 day intervals. This timely, detailed imagery supports tracking of crop-specific alterations linked to harvest and burning, facilitating accurate delineation of burnt and unburnt zones from residue burning in smallholder systems.

Lately, machine learning techniques have gained traction for burnt area mapping. Unlike traditional rule-based systems, these methods can directly extract burnt-pixel traits from annotated training datasets. Earlier work has applied diverse machine learning methods, including random forests [15-17], support vector machines [18-20], and neural networks [17, 21-25] for this purpose. Nevertheless, few studies have examined the value of machine learning for mapping areas of agricultural land burned by fire, particularly in smallholder landscapes. As far as we know, no prior research has examined machine learning specifically for detecting burnt sugarcane residue areas.

In light of the demand for a reliable method to map burnt zones in sugarcane crop residue, this research evaluates two machine learning methods for identifying such zones in Thailand's smallholder systems during the 2019–2020 season using HLS satellite data. Initially, we produced a detailed 30 m land cover map of crop types, including sugarcane, derived from Sentinel-1 Synthetic Aperture Radar (SAR) imagery. We then isolated sugarcane pixels from this map to pull matching HLS observations across the season. Combined with ground records of burnt and non-burnt sites, these data informed the creation of mapping models using Support Vector Machines (SVM) and Neural Networks (NN). Implementation occurred in a key smallholder sugarcane zone in Thailand. We selected SVM and NN as they belong to different machine learning families. SVM functions as a standard classifier (akin to random forests) and acts as our reference to check if raw multispectral bands contain adequate detail for burnt area separation. Neural networks, representing a more sophisticated group, allowed us to measure gains in mapping precision beyond the SVM baseline.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

Materials and Methods

Study region

The study region lies in the central and northeastern sectors of Thailand **(Figure 1)**. It spans six provinces and is a leading sugarcane production zone, accounting for roughly 38% of national output. Sugarcane planting covers approximately 1.2 million ha. Around 90% of cultivation here is rainfed [26], with over 80% occurring on small farms of less than 10 ha [27]. The area hosts nine sugar mills. Additional prominent crops include rice, cassava, and maize.

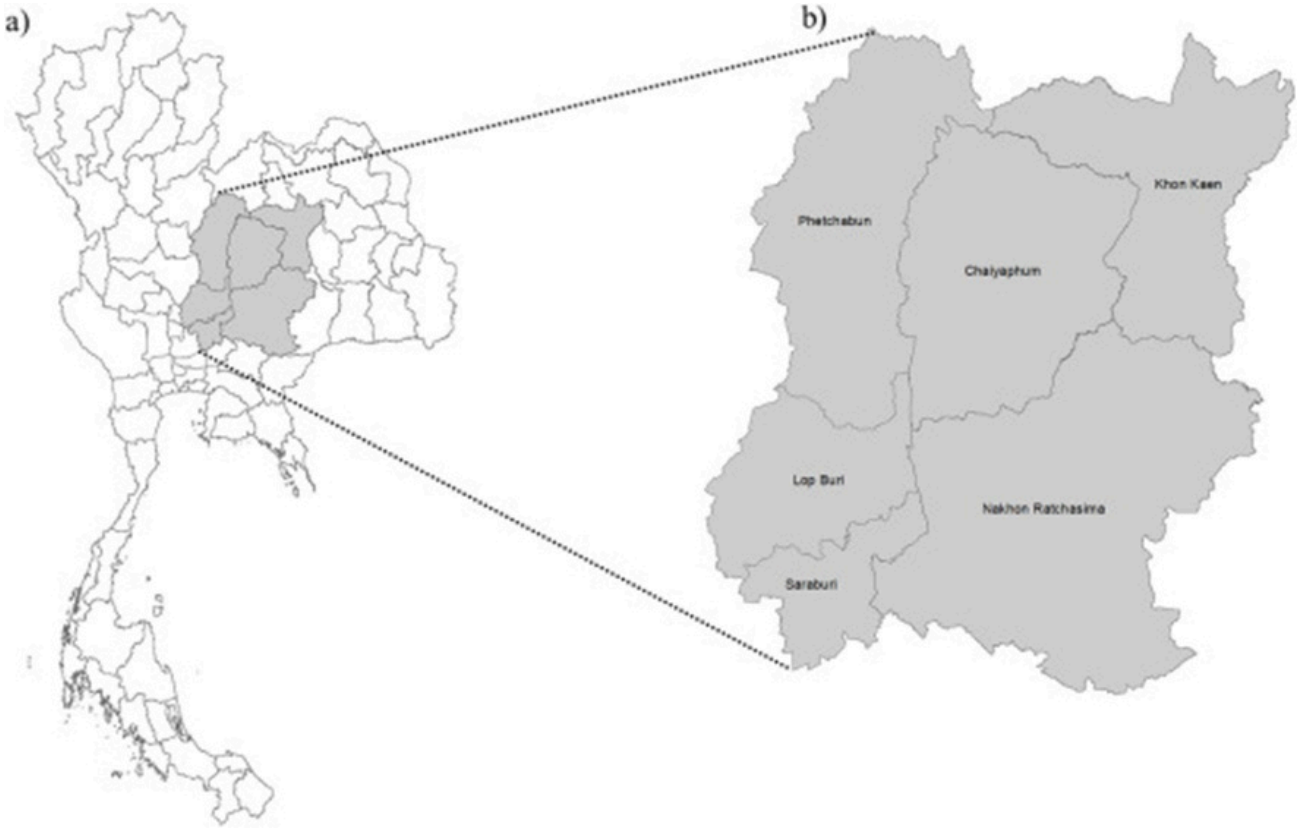

Figure 1. (a) Map of Thailand and (b) study region comprising six provinces.

Brief description of the classifiers

Support vector machines (SVMs)

Support Vector Machines constitute a family of supervised learning classifiers rooted in statistical learning theory [28-30]. In a typical binary classification scenario, the training set comprises N samples denoted as x1, x2, …, xN, each paired with labels y1, y2, …, yN, where yi belongs to the set {−1, 1} for i = 1 to N. A basic linear SVM identifies the optimal separating hyperplane W by maximizing the margin, the distance from the hyperplane to the nearest positive and negative data points. The equation representing this Linear SVM hyperplane is given by:

$$W.X-b=0$$

In this formulation, b indicates the bias or offset from the origin, W controls the direction of the hyperplane, and X refers to any point located on the hyperplane itself.

When the data is linearly separable, the SVM selects the hyperplane that achieves the widest possible margin — the space between two parallel supporting hyperplanes that bound the positive and negative samples. As depicted in **Figure**

© 2025 The Author(s). | Published by Agro Intelligence Research Press

2, the data points touching these parallel boundaries are known as support vectors. Consequently, the primary aim of SVM is to locate the hyperplane that maximizes the separation between the boundaries of the positive and negative support vectors.

- W.Xi - b ≥ 1 for all points X belonging to the positive class yi = 1 (equality holds for support vectors)
- W.Xi - b ≤ −1 for all points X belonging to the negative class yi = −1 (equality holds for support vectors)
- Combining the above two, yi(W.xi - b) ≤ 1 for all training samples 1≤ i ≤ N.

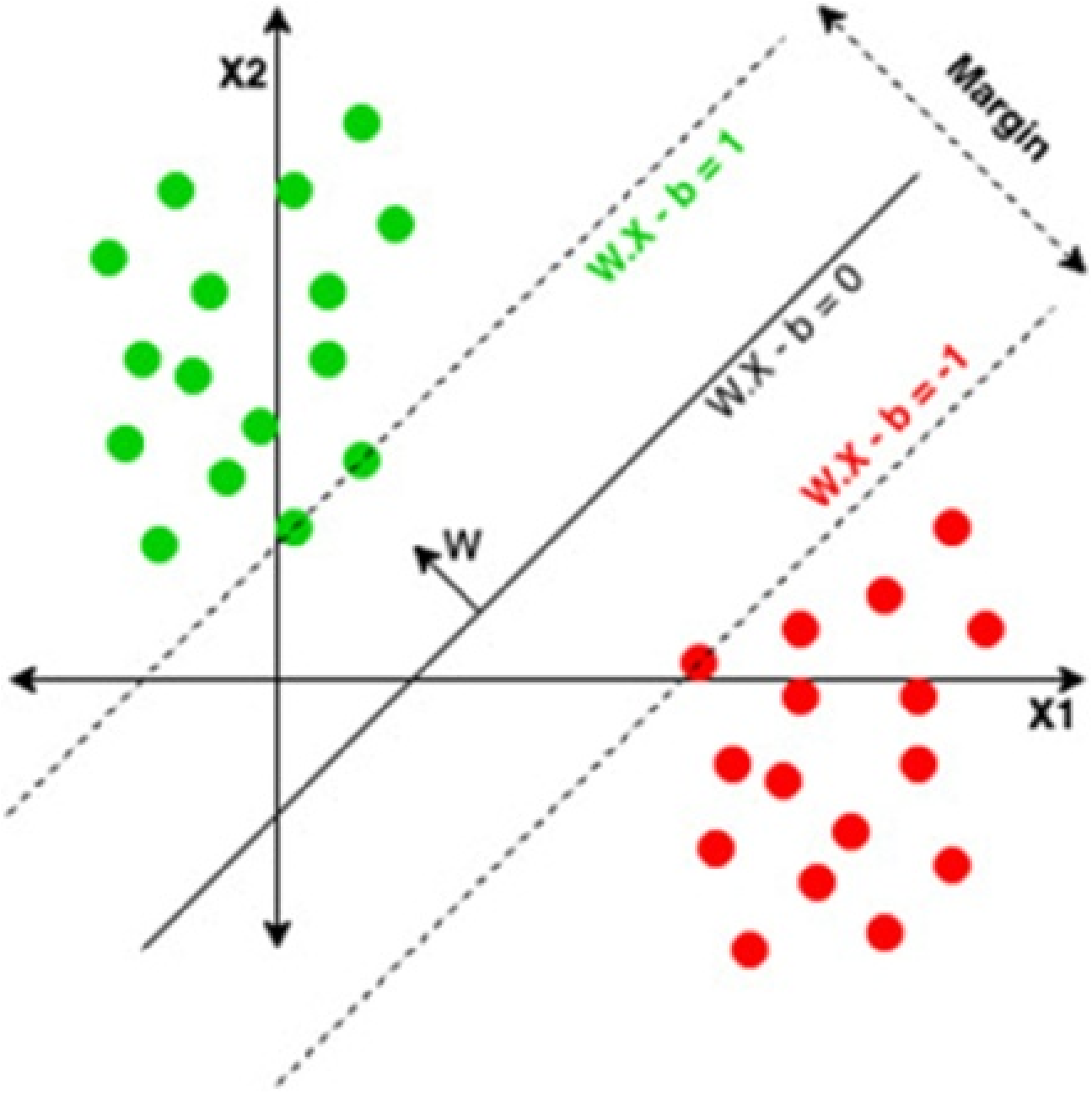

Figure 2. An illustration of a typical Linear SVM classifier for binary classification.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

From a geometric perspective, the margin width between the two supporting hyperplanes is 2/||W||. Therefore, the optimization task is to minimize ||W||. Overall, the optimization problem for a simple linear SVM can be expressed as follows:

$$\text{Min } ||W|| \quad (1)$$

$$\text{Such that } yi^*(W.xi - b) >= 1 \text{ for } 1 <= i <= N$$

For cases involving complex, non-linearly separable data distributions, the same principle can be applied by first projecting the data into a higher-dimensional feature space where linear separation becomes feasible. This mapping is performed using kernel tricks, including the RBF, polynomial, and sigmoid kernels.

In this research, SVM models with multiple kernel types served as the baseline for evaluating burnt-area classification accuracy and gauging the inherent information in the dataset. These results were subsequently compared with Neural Networks to assess the relative strengths and modeling capacities of both techniques in burnt-area delineation.

Neural networks

Neural Networks (NN) offer a more advanced and powerful class of supervised classifiers than SVMs. They build on Multilayer perceptron networks [31, 32] and draw inspiration from biological brain functions, emulating signaling processes among neurons. These models consist of an input layer, multiple hidden layers, and a final output layer. The incorporation of non-linear activation functions into the hidden layers enables neural networks to process complex, non-linear relationships effectively. Training is driven by the back-propagation algorithm, which iteratively refines the network's parameters to better align with the training data.

The core principles of Neural Networks have delivered remarkable progress across numerous domains, including Computer Vision [33], Natural Language Processing [34], speech processing [35], and many others. Remote sensing research has also increasingly adopted these powerful tools, for example, in preseason crop mapping applications [36]. The structure of the NN employed in this study is displayed in **Figure 3**, while comprehensive model specifications are provided in Section 5.1 of the implementation details.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

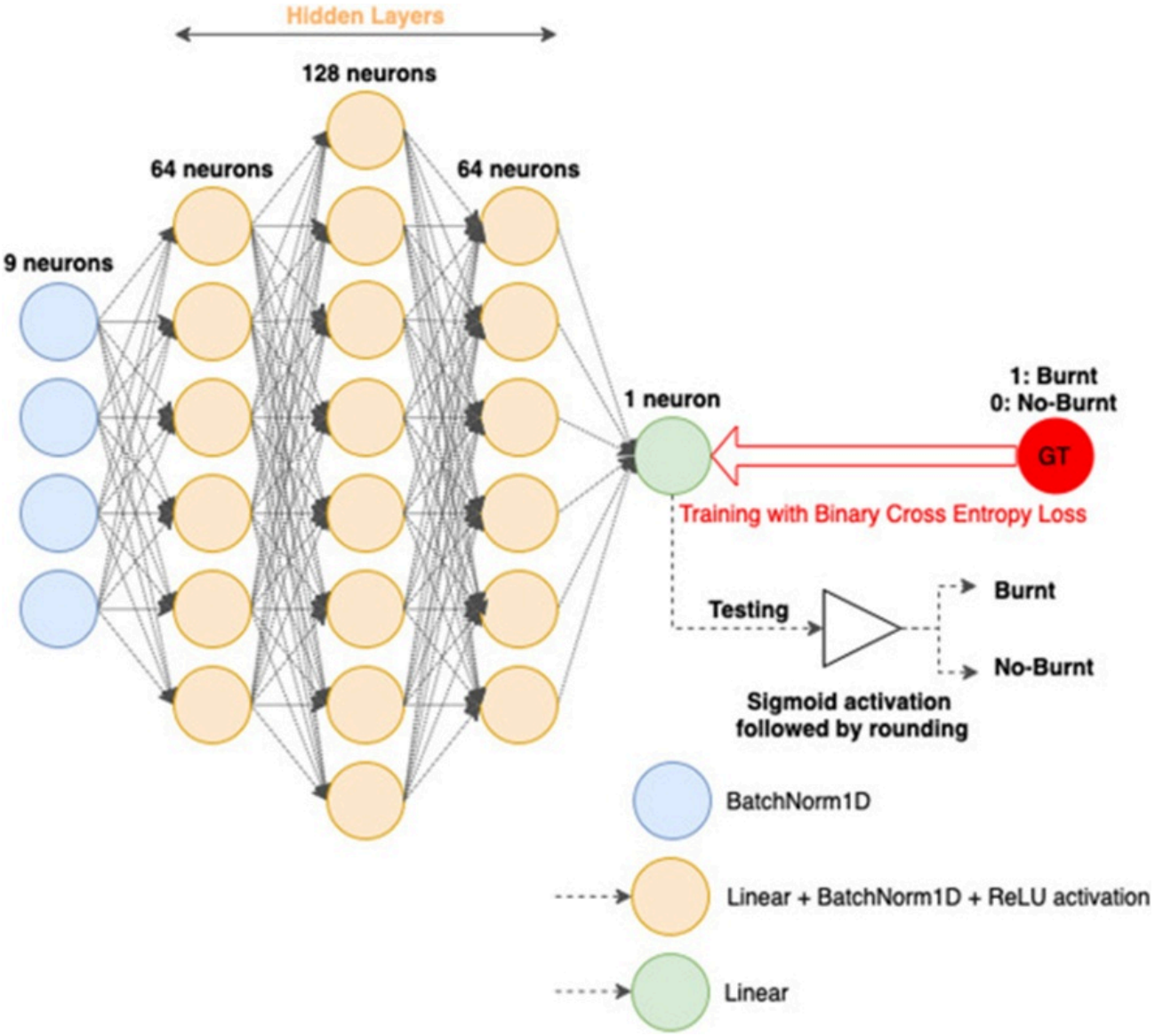

Figure 3. Architecture of the Neural Network (NN) used for burnt area mapping. The NN model was designed with 9 nodes, representing the reflectance of 9 spectral bands in the Sentinel-2 data, and 3 hidden layers with 64, 128, and 64 nodes, respectively, with ReLU non-linear activations, and an output layer with one node for our binary classification prediction.

Data sources and preparation

Developing classifiers and generating maps of sugarcane burnt areas required the following datasets: 1) a sugarcane mask for pinpointing sugarcane pixels, 2) ground-based field observations for model training and validation, and 3) satellite data serving as input features. **Figure 4** outlines the workflow for data preparation and classifier deployment. In Step 1, a 30 m resolution sugarcane mask was derived from the 2019 land cover map generated from Sentinel-1 SAR imagery using a Long Short-Term Memory (LSTM) deep neural network. In Step 2, harmonized Landsat 8 and Sentinel 2 (HLS) data [14] covering nine spectral bands were retrieved for harvested non-green sugarcane pixels identified via the mask during the harvesting window. In Step 3, HLS data from burnt and unburnt sugarcane fields — identified through windshield surveys — were collected for training, validation, and testing both NN and SVM models. In Step 4, the trained models were applied across all sugarcane pixels using HLS inputs to produce burnt and non-burnt area maps for the 2019–20 season. Further details on data sources and preparation are elaborated in the sections that follow.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

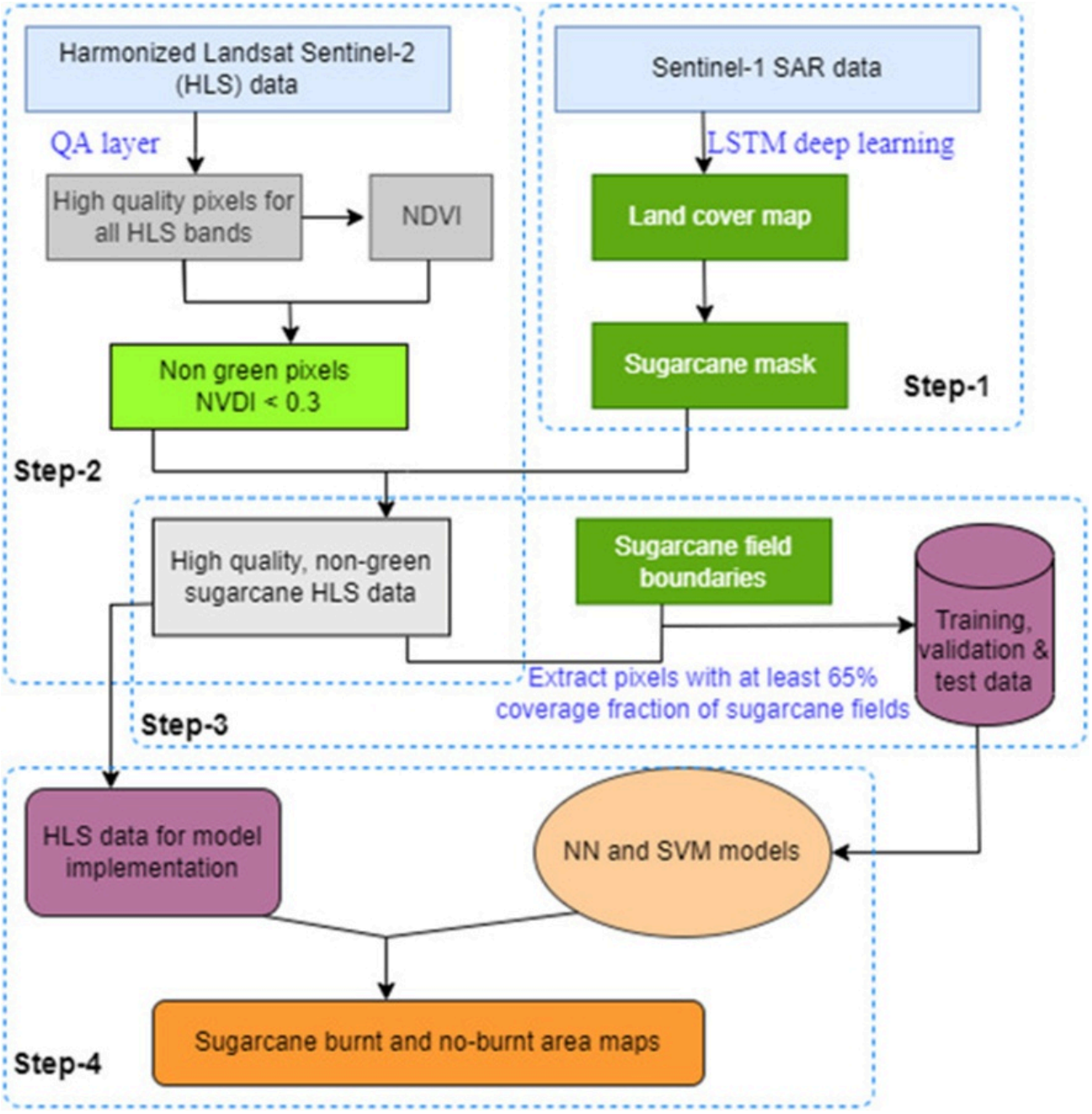

Figure 4. Schematic illustrating steps for HLS data preparation for model development and implementation.

Sentinel-1 SAR satellite data

Creating accurate post-fire burnt-area maps depends on a reliable crop-specific mask to locate sugarcane fields. To develop a detailed land cover map encompassing sugarcane, Sentinel-1 Synthetic Aperture Radar (SAR) data was utilized. Sentinel-1 consists of two satellites in a near-polar sun-synchronous orbit, delivering a 6-day repeat cycle and supporting various acquisition modes, with the interferometric wide swath (IW) mode selected for this land cover task. A complete stack of 53 images from 2019 was downloaded from the Alaska Satellite Facility for the study area. Ground-range-detected products underwent preprocessing with the SNAP Sentinel-1 toolbox, including orbit correction, thermal noise removal, radiometric calibration for incidence-angle effects, terrain correction using NASA SRTM 30 m data [37], and speckle filtering to mitigate interference noise. The processed data were converted to the decibel scale, yielding VV and VH polarization bands with sigma-nought backscatter coefficients [8].

© 2025 The Author(s). | Published by Agro Intelligence Research Press

Sugarcane mask

The sugarcane mask was constructed from the 2019 land cover map of the study area **(Figure 4)**. Initially, a comprehensive land cover classification covering sugarcane grown in the 2019–2020 season was generated using sigma-nought backscatter from Sentinel-1 VV and VH bands combined with elevation information. Elevation values were sourced from NASA SRTM data [37]. Model development relied on the Long Short-Term Memory (LSTM) deep neural network implemented in PyTorch [36].

Input sequences were generated by compositing 7-day periods from available Sentinel-1 images, producing 53 composites, each containing three channels (VV, VH, and elevation). Training labels for multiple land cover classes were assigned to 1700 field polygons, which, when intersected with Sentinel-1 data, yielded 50,450 labeled pixels split into training (70%), validation (20%), and testing (10%) sets.

A four-layer LSTM architecture was trained using the training and validation data. Model performance was evaluated on the held-out test set. Upon successful training, the model was applied across the entire study region to extract sugarcane pixels and generate the final mask. This sugarcane mask achieved an overall accuracy of approximately 91% **(Figure 5)**.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

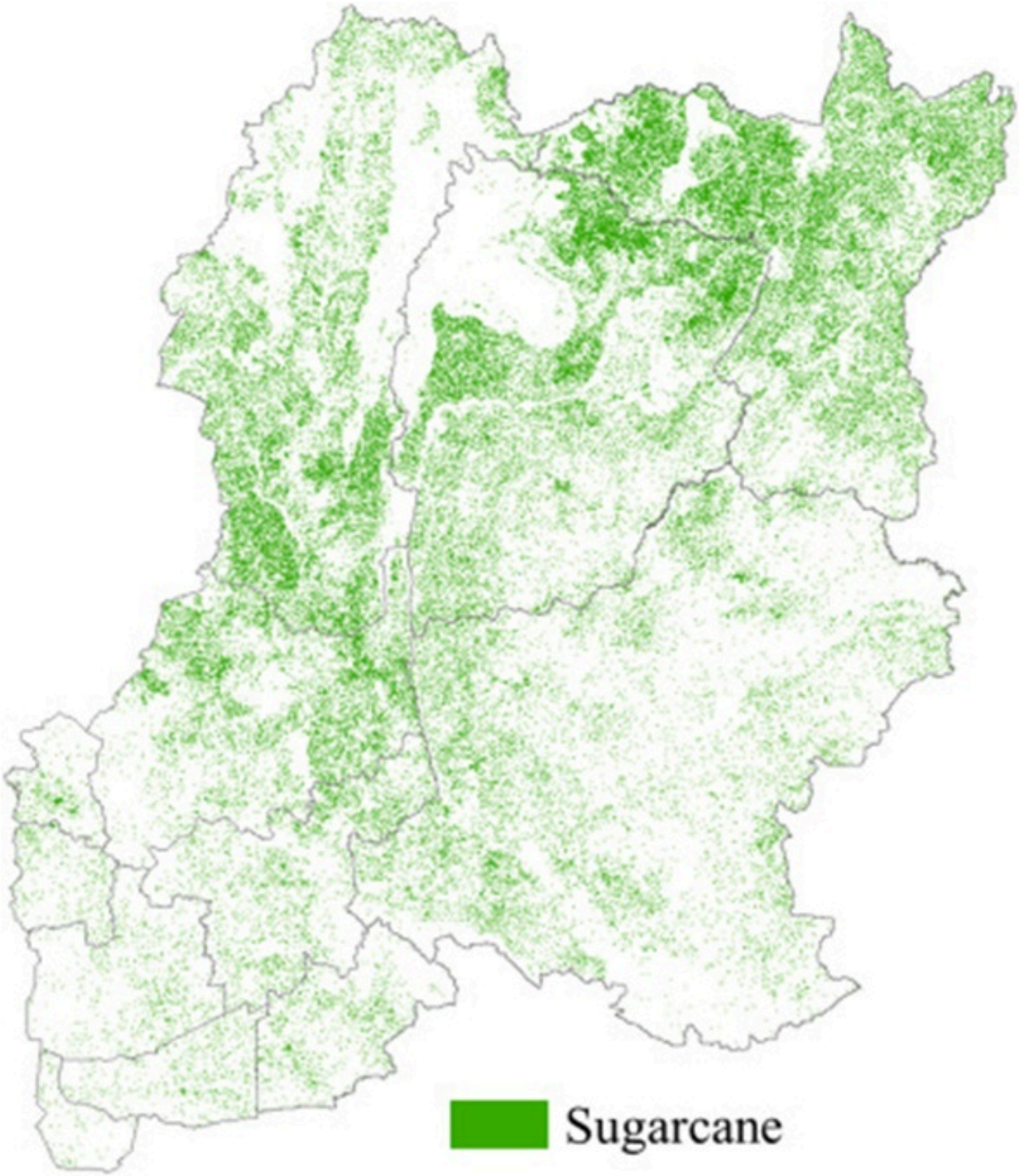

Figure 5. Sugarcane mask (30 m) produced using the 2019 land cover map developed using the LSTM deep neural network algorithm and Sentinel-1 SAR data.

Harmonized Landsat Sentinel (HLS) satellite data

The developed classification approaches relied on harmonized Landsat 8 and Sentinel 2 (HLS) surface reflectance from nine spectral bands as input features [14]. The HLS product delivers consistent, radiometrically aligned imagery originating from Landsat 8's Operational Land Imager (OLI) and Sentinel-2A/B's MultiSpectral Instrument (MSI). Merging

© 2025 The Author(s). | Published by Agro Intelligence Research Press

these sources results in high-frequency (1–4 days) time series at 30 m spatial resolution. All imagery underwent standardized processing including atmospheric correction, cloud and shadow masking, precise geolocation, illumination and view angle normalization, and cross-sensor spectral band harmonization [38].

A total of 29 HLS scenes were obtained from December 1, 2019, to March 31, 2020, covering the sugarcane residue burning and harvesting window. The accompanying quality assurance (QA) layer was used to exclude pixels affected by clouds, cloud shadows, or snow.

Preparation of HLS data

Surface reflectance values from HLS for sugarcane pixels were essential for both model training and large-scale burnt area mapping. The sugarcane mask was first reprojected to match the HLS data coordinate system (WGS 84 / UTM zone 47 N). The mask and HLS imagery were spatially aligned to extract reflectance values for all 9 bands at each sugarcane pixel location. Following harvest — whether burned or unburned — fields should be devoid of green vegetation. Any remaining green pixels suggest either unharvested mature cane or regrowth of the subsequent crop. To filter these out, the Normalized Difference Vegetation Index (NDVI) was calculated, and pixels with NDVI values above 0.3 were removed as green vegetation. The resulting non-green HLS observations collected during the burning and harvesting phases served as the basis for model development and operational mapping.

Training, validation, and test data

Windshield surveys were performed during the harvesting phase of the 2019–20 season to gather residue management details from 480 sugarcane fields spread throughout the study region **(Figure 6)**. These fields, when intersected with the sugarcane mask, yielded a total of 22,000 sample pixels at 30-m resolution. Allocation included 81% for classifier training, 9% for the validation set to identify the top-performing model, and 10% for the test set to measure final performance. Sampling focused on pixels where field coverage reached at least 65% to confirm that HLS data accurately represented the spectral signatures of actual sugarcane fields.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

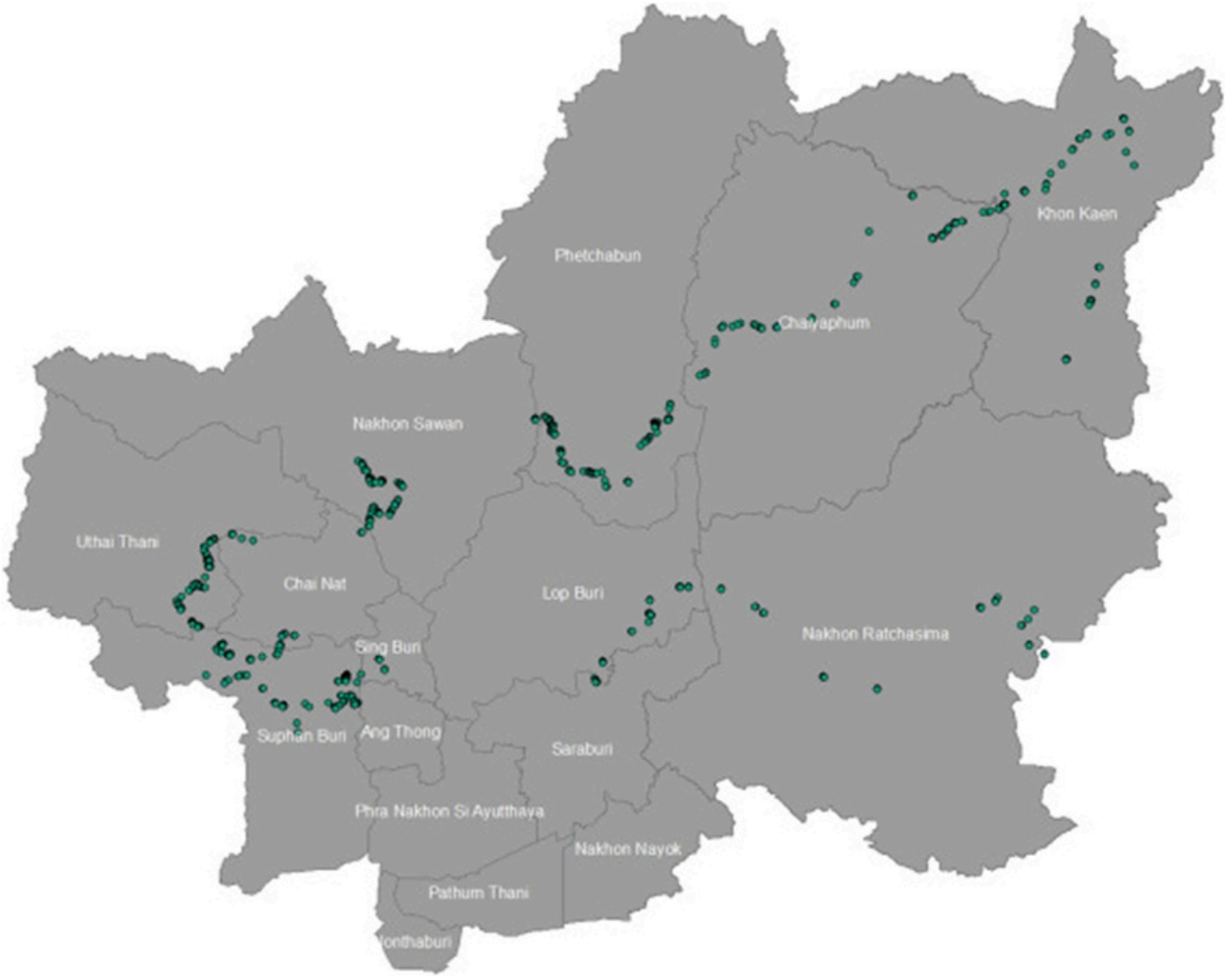

Figure 6. Green-filled circles represent locations of burnt and non-burnt sugarcane fields. These field coordinates were collected using windshield surveys.

Development and implementation of the models

NN-based method

A NN approach was constructed, trained, and optimized within the PyTorch Python framework. According to the setup shown in **Figure 3**, the network included 9 input nodes to capture reflectance from 9 Sentinel-2 spectral bands, along with 3 hidden layers with 64, 128, and 64 nodes, respectively. ReLU served as the non-linear activation function, and a single output node handled the binary prediction task. BatchNorm layers were added after each hidden layer but omitted from the final output. The complete structure followed this sequence: [Linear(9, 64), BatchNorm, ReLU], [Linear(64, 128), BatchNorm, ReLU], [Linear(128, 64), BatchNorm, ReLU], [Linear(64, 1)], with Linear(M, N) indicating a fully connected layer between M and N nodes. Model training used BCEWithLogitsLoss from PyTorch with the Adam optimizer at a learning rate of 0.001, running for 5000 epochs.

SVM-based method

The SVM approach was developed using the scikit-learn library in Python. Available kernel options — Linear, RBF, polynomial, and sigmoid — were all tested. Input consisted of raw band values, with separate SVM models trained on each kernel using the combined {train+val} dataset. Final testing performance was measured by applying the models to

© 2025 The Author(s). | Published by Agro Intelligence Research Press

the separate test set. Default settings were maintained due to the exploratory nature of the SVM evaluation: the regularization parameter C remained at 1.0, and gamma was internally scaled based on the training sample variance.

Implementation of models

Trained versions of both NN and SVM were deployed on the prepared HLS time series covering every sugarcane pixel in the full study region. This generated maps that distinguished burnt from non-burnt sugarcane pixels across the harvesting timeframe from December 01, 2019, to March 31, 2020.

Accuracy assessment

Pixel-level evaluation

Assessment of NN and SVM performance was conducted at the pixel and regional scales. The test dataset supported pixel-level checks, while OCSB provincial sugarcane production reports informed regional analysis based on harvesting methods.

At the pixel scale, SVM classification used the maximum-margin hyperplane defined by parameters W and b. Each test pixel received a label based on its relation to the hyperplane, and accuracy was the count of correct predictions divided by the total number of test pixels.

NN pixel-level checks occurred in two phases. During training, validation was monitored at intervals across all 5000 epochs by feeding those pixels into the evolving model. Outputs were transformed via sigmoid activation to the 0–1 range, then thresholded at 0.5 to label pixels as burnt (1) or non-burnt (0). The highest-performing validation model was chosen and applied to test pixels, with results summarized through the confusion matrix **(Figure 7c)**.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

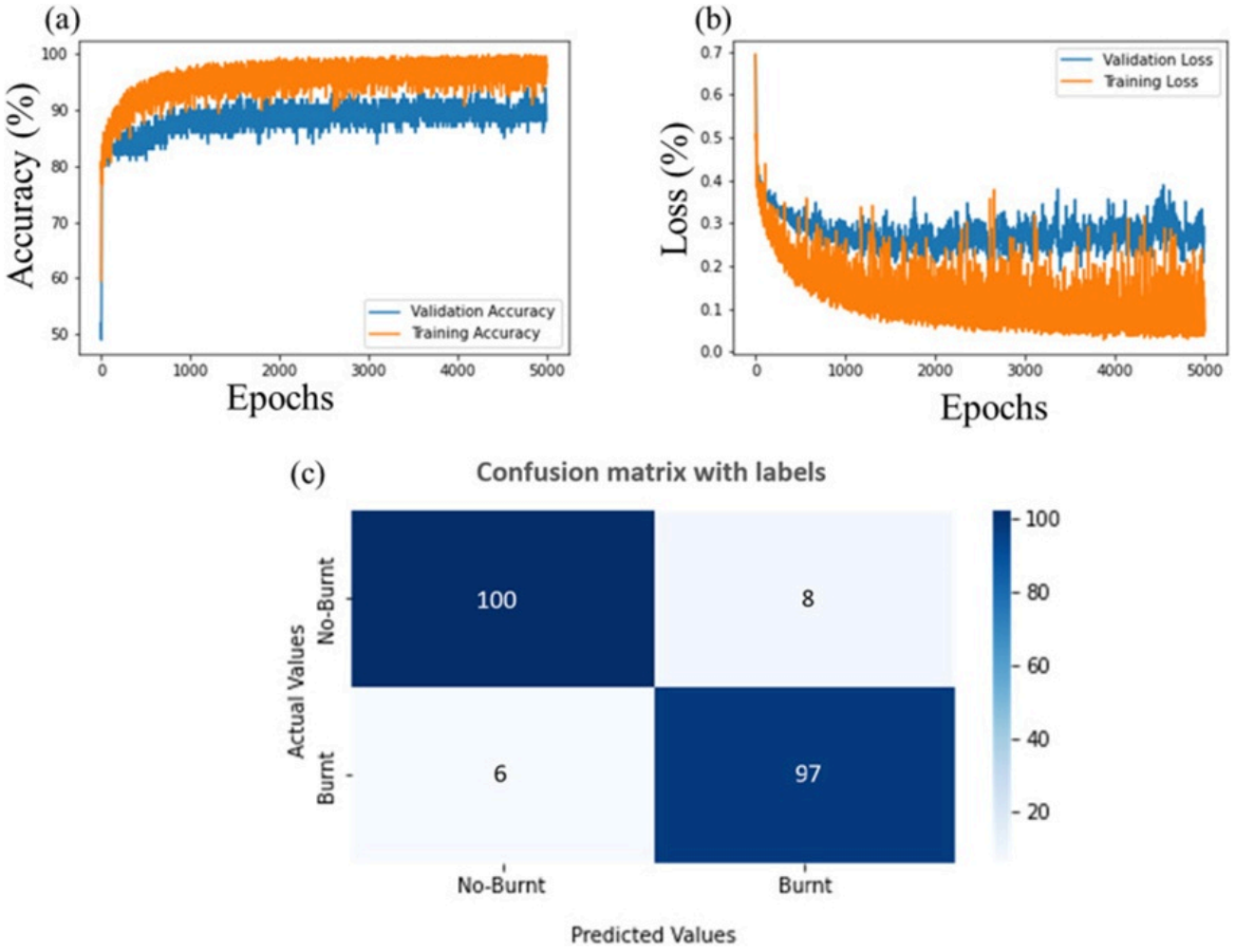

Figure 7. (a) Training and Validation Accuracies, (b) Training and Validation Losses, and (c) Confusion Matrix on the test data.

McNemar's test [39] was used to determine whether differences in performance between classifiers were statistically significant. Under a chi-squared (χ^2) distribution, the test examines the null hypothesis of equivalent classifier effectiveness.

$$\chi 2 = (f_n - f_m)^2 / (f_n + f_m) \quad (2)$$

Where fn counts cases misclassified by the first classifier but correctly handled by the second, and fm counts cases correctly handled by the first but missed by the second.

Burnt area detection quality was further analyzed via three indicators from Kontoes *et al.* [40]: detected area efficiency (DAE), skipped area rate (SAR, representing omission error), and false area rate (FAR, representing commission error). Calculations followed these standard formulae:

$$\text{Detection Area Efficiency (DAE)} = \frac{\text{DBA}}{\text{DBA+SBA}} \quad (3)$$

© 2025 The Author(s). | Published by Agro Intelligence Research Press

$$\text{False Area Rate (Commission Error)} = \frac{\text{FBA}}{\text{DBA+FBA}} \quad (4)$$

$$\text{Skipped Area Rate (Ommission Error)} = \frac{\text{SBA}}{\text{DBA+SBA}} \quad (5)$$

Regional scale evaluation

Comparison against OCSB reports (Office of Cane and Sugar Board [41]) for the 2019–20 season enabled regional evaluation of both NN and SVM outputs. OCSB records detail the procured sugarcane volumes, separated into fresh (unburnt) and burnt categories, per processing unit. With 1–3 units typically present per province and cultivation occurring nearby, provincial aggregates were formed for production by practice. Provincial percent burnt area and total burnt hectares were derived from these figures. Parallel aggregation of NN- and SVM-classified pixels by practice and province within the study area enabled direct comparison, yielding absolute percentage error values.

Results and Discussion

Pixel-level performance

Performance of the NN classifier

Evaluation included two stages: ongoing checks during model training and final checks after full deployment. Trends in accuracy and loss for both the training and validation sets are shown in **Figure 7**. Training loss declined progressively with advancing epochs. In contrast, accuracy approached 100%, indicating successful fitting to the training examples. Validation patterns mirrored this behavior, suggesting strong generalization potential across broader areas. The maximum validation accuracy reached 93% at epoch 3393. Application of this selected model to the test set yielded 93.43% accuracy.

Performance of the SVM classifier

Most kernels achieved solid results above 80% accuracy, except for the sigmoid kernel. Polynomial and RBF kernels stood out, delivering comparable results (Polynomial = 81.2% & RBF = 79.8%). The leading result belonged to the polynomial kernel at 81.2 %, ahead of the linear kernel at 80.2 %. Sigmoid delivered only 28.6 %, a notably weaker outcome. Polynomial and RBF kernels improve separation by lifting data into higher dimensions, outperforming direct-space linear and sigmoid kernels. Analogous patterns emerged in prior land cover studies [42, 43].

Overall, SVM models trailed the NN approach, as evidenced by McNemar's test results and burn area detection indicators.

The 2×2 contingency table **(Table 1)** covering 213 fields showed NN correctly identifying 147 fields that SVM had missed, while SVM succeeded on 9 fields where NN failed. Agreement occurred on 52 correct and 5 incorrect cases. McNemar's test yielded a chi-square value of 122, well above the 3.84 critical threshold (alpha = 0.05). This outcome rejected equal performance and affirmed NN's clear advantage.

Table 1. Contingency table for the predictions from SVM and NN models on the test dataset.

Empty cell	**SVM (Correct)**	**SVM (Incorrect)**
NN (Correct)	52	147
NN (Incorrect)	9	5

© 2025 The Author(s). | Published by Agro Intelligence Research Press

Table 2 highlights NN's higher detected area efficiency, along with lower rates of skipped and false areas, compared to SVM.

Table 2. Ratio of detected, skipped, and false areas with respect to the ground-truth area for NN and SVM predictions on the test dataset.

Method	Area detection accuracy ↑	Missed area percentage (omission error) ↓	Incorrect area percentage (commission error) ↓
SVM	43.6	56.31	67.6
NN	94.1	5.8	7.6

SVM's reduced effectiveness is likely due to the size of the training sample and the number of input features. Earlier comparisons [44] found that SVMs were competitive or superior with small sample sizes, yet NN gain dominance as sample sizes grow. Li *et al.* [45] observed NN superiority for active fire detection using around 20,000 samples. The present study employed roughly 18,000 training samples, a level that favored NN. The quantity of input variables also matters for SVM, with more variables typically improving its performance. Raczko and Zagajewski [46] noted that NN offer advantages when variable counts remain suboptimal for SVMs. The choice of nine spectral bands here may have limited the SVM's ability to match NN capabilities.

Comparison of regional-scale performance

Maps showing burnt areas from the NN and SVM classifiers, including their geographic layouts, are presented in **Figures 8a and 8b**. Despite the SVM map detecting a much greater quantity of burnt pixels **(Figure 8c)**, the general distribution patterns across both outputs stayed similar. For example, provinces in the southwest had higher burnt-pixel counts than those in the northeast. According to **Figures 8a and 8b**, Lop Buri province, recorded 71.10% of the burnt area (SVM estimate = 82.80%), and Sara Buri province recorded 55.93% (SVM estimate = 68.72%), both situated in the southwest. Meanwhile, Chaiyaphum and Khon Kaen provinces in the northwest registered lower burning levels at 42.21% (SVM estimate = 56.01%) and 41.04% (SVM estimate = 46.46%), respectively.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

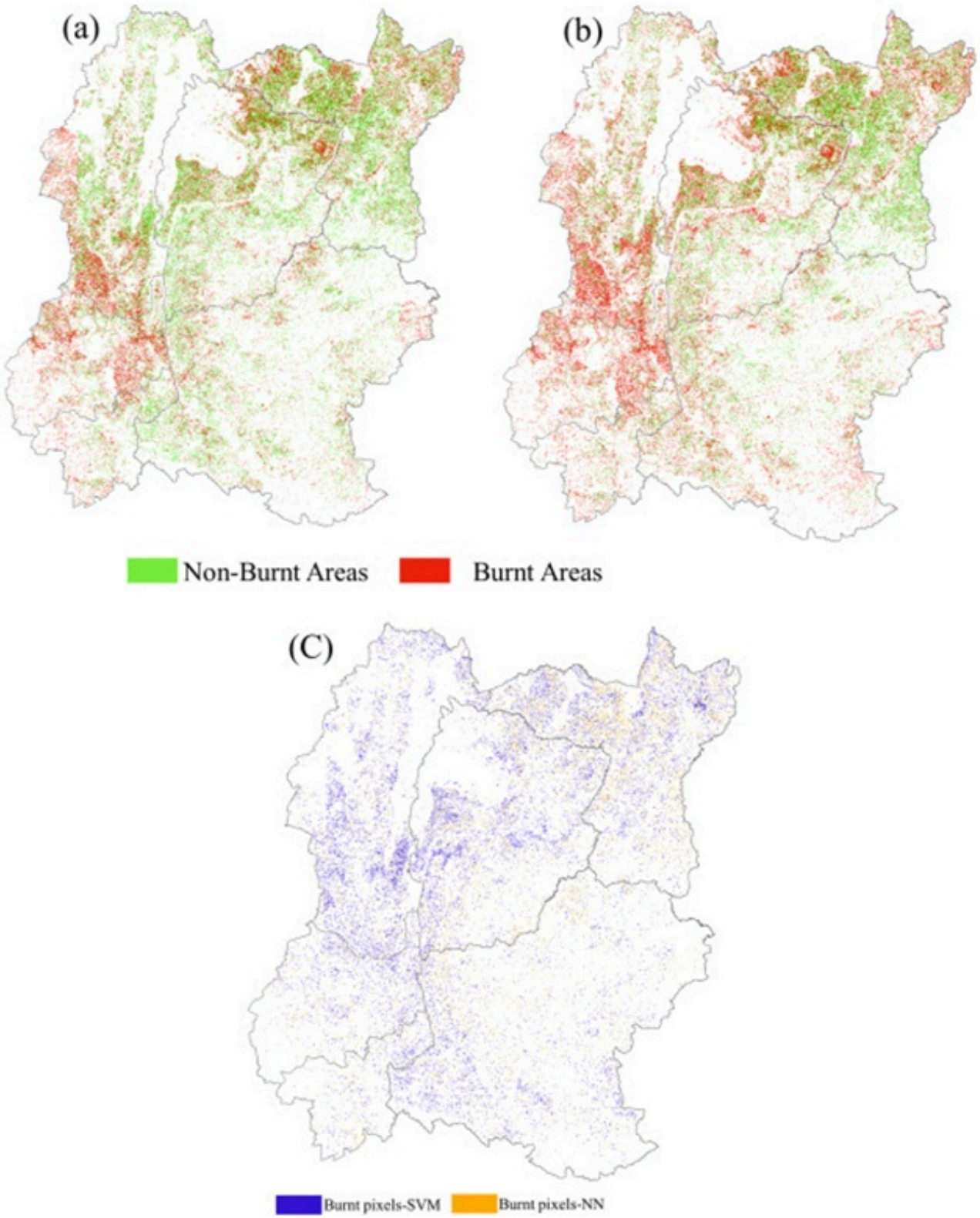

Figure 8. Comparison of predicted burnt (red colored pixels) and no-burnt (green colored pixels) maps of the neural network (NN) model (a) and the support vector machines with polynomial kernel function (SVM) model (b). Difference between SVM and NN predicted burnt pixels, and (c) Burnt -SVM represents pixels that the SVM model predicted as

© 2025 The Author(s). | Published by Agro Intelligence Research Press

burnt, but the NN model classified as non-burnt. At the same time, Burnt-NN indicates pixels that the NN model classified as burnt, but the SVM model predicted as non-burnt.

When NN and SVM percentage estimates were benchmarked against OCSB official statistics, the NN figures demonstrated tighter correspondence to the reported values than the SVM figures did **(Figure 9)**. For the full study region during the 2019–20 growing season, the mean percent burnt area came to 51.08% under NN and 62.90% under SVM. These surpassed the OCSB baseline of 48.71% by 4.86% and 29.13%, respectively. Provincial-level absolute percentage errors fluctuated substantially, ranging from 0.19% to 19.59% for NN estimates and from 13.34% to 74.28% for SVM estimates. The greatest errors for both methods occurred in Chaiyaphum province. This outcome probably relates to the unusually small burnt area percentage reported by OCSB for that location, which featured just one sugarcane processing facility and a reported burning rate of 29% for 2019–20 — markedly below the overall regional mean.

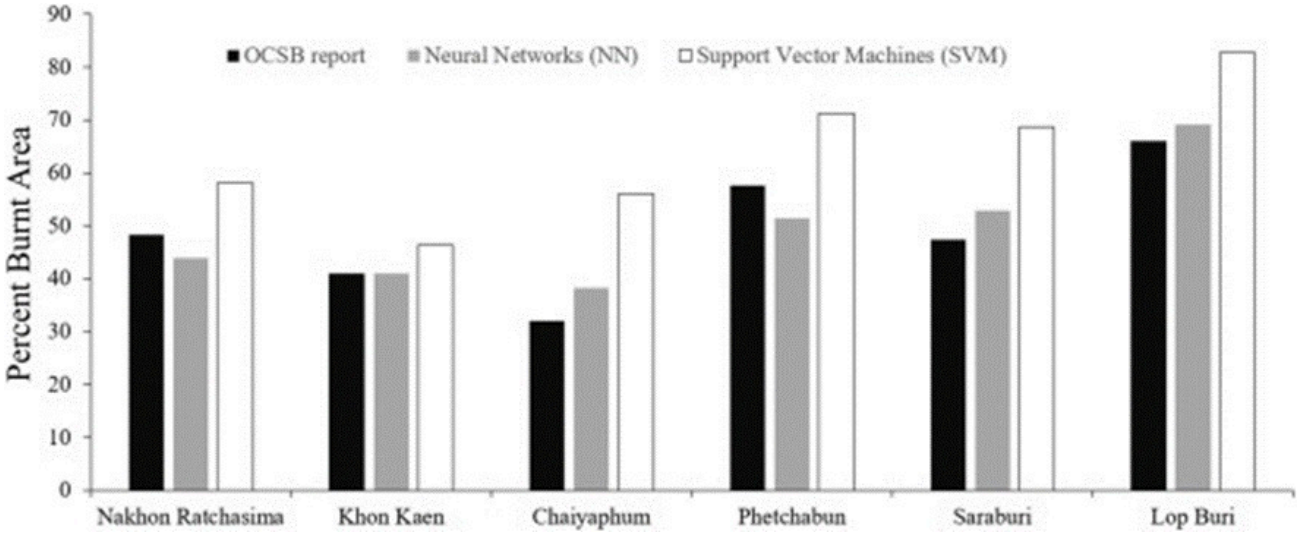

Figure 9. Comparison of aggregated percent burnt areas by province, determined based on neural network (NN) and support vector machines (SVM) with the Office of Cane and Sugar Board (OCSB) of Thailand.

Discrepancies in total areas

Percent burnt area results aligned fairly well with OCSB documentation, above all for the NN technique, yet calculations of overall burnt extent revealed bigger variations irrespective of method **(Figure 10)**. In the study region for the 2019–20 season, NN produced an approximate total burnt area of 315 thousand ha, while SVM reached 418 thousand ha, against the OCSB total of 249 thousand ha. The corresponding mean absolute percentage errors were 26.99% for NN and 59.02% for SVM when measured against OCSB data. The close match in percentages, paired with divergence in absolute areas, pointed to uncertainty drivers lying outside the classifiers. Among these, the sugarcane mask stood out, as its derived sugarcane coverage surpassed the OCSB sugarcane area total by 23%.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

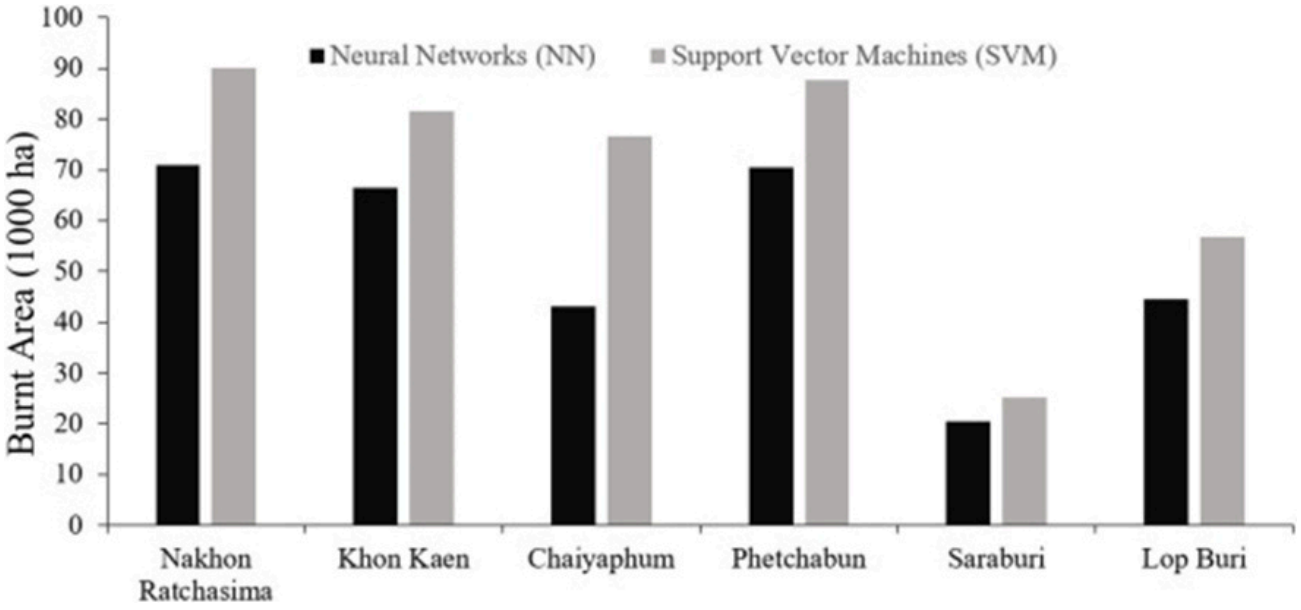

Figure 10. Comparison of aggregated total burnt areas by province, determined based on neural network (NN) and support vector machines (SVM) with the Office of Cane and Sugar Board (OCSB) of Thailand.

Comparisons of global products

Evaluating the effectiveness of SVM and NN classifiers in burnt area mapping involved contrasting their outputs with two prominent global burnt area datasets: a) the MODIS Terra and Aqua combined MCD64A1 Version 6 (500 m) product, and b) the MODIS FIRECCI51 (~250 m) product. MCD64A1 delivers monthly layers of MODIS surface reflectance at 500 m resolution, merged with 1 km-resolution active fire detections [7]. Similarly, FIRECCI51 generates monthly outputs from MODIS Near Infrared Red reflectance at 250 m resolution supplemented by MODIS active fire information [47]. The intersection of these global layers with the local sugarcane mask isolated relevant burnt sugarcane pixels to calculate regional burnt totals **(Figure 11)**.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

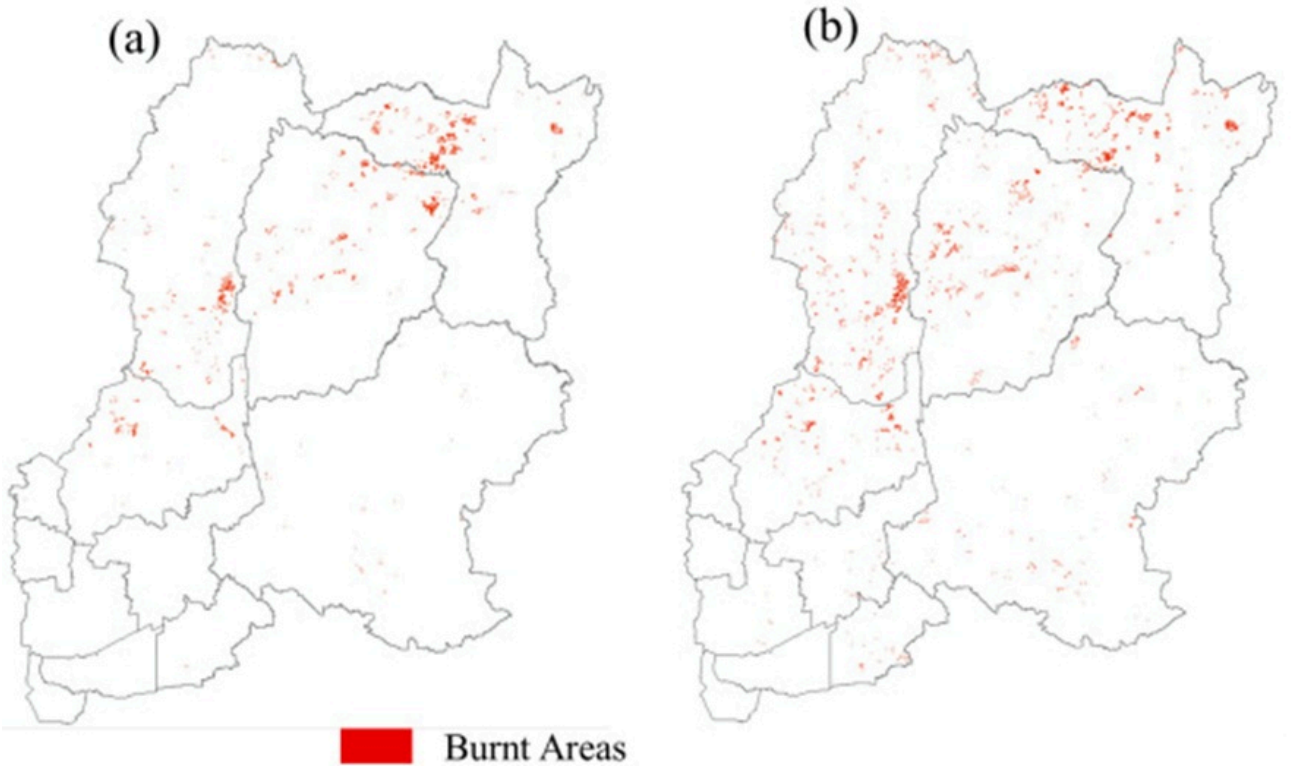

Figure 11. Burnt areas extracted for Sugarcane pixels from two global products: (a) MCD64A1.006 (500 m) and (b) FIRECCI51 (~250 m).

During the 2019–20 season, MCD64A1 indicated 46 thousand ha of burnt sugarcane area, and FIRECCI51 indicated 67 thousand ha within the study boundaries. Both totals were substantially below the OCSB-reported 249 thousand ha, falling short by 82% and 73%, respectively. Overcounting of sugarcane pixels in the mask may have led to minor additional burnt detections. Excluding such inflation would further reduce global estimates. Thus, while NN and SVM tended to overestimate burnt extents, the global products strongly underestimated them relative to OCSB figures. Still, the developed machine learning classifiers aligned better overall with official records.

Conclusion

Due to the demand for reliable tools to detect crop residue burning and the scarcity of machine learning research on the topic, this investigation tested two approaches — support vector machines across varied kernel types, together with artificial neural networks — to map sugarcane-burnt zones in smallholder-dominated districts of Central and Northeast Thailand. The chief difficulty in crafting trustworthy detection systems for residue burning under smallholder conditions lies in the scarcity of crop-specific satellite imagery that can capture the distinct spectral signatures of burnt surfaces. This gap was bridged by creating precise sugarcane extent maps from dense Sentinel-1 SAR time series analyzed via Long Short-Term Memory (LSTM) deep neural network techniques. Furthermore, harmonized Landsat 8 and Sentinel 2 (HLS) observations at a high temporal frequency (4–8 days) provided critical spectral details specific to sugarcane throughout the brief burning-harvesting interval, facilitating the construction of robust SVM and NN models.

At the pixel scale, both SVM and NN achieved high accuracy, exceeding 80% for burnt sugarcane identification. Polynomial and Radial Basis Function kernels ranked among the SVM options, surpassing the linear and sigmoid versions. Overall, NN achieved higher accuracy than SVM by over 10%. Regional analysis indicated that NN and SVM outputs both reflected provincial differences in burning intensity, despite variations in the exact burnt-pixel counts. SVM percent burnt area reached 62.90%, showing a larger deviation from the OCSB (48.71%), whereas NN at 51.08% differed

© 2025 The Author(s). | Published by Agro Intelligence Research Press

by less than 5%. Greater mismatches appeared in total burnt area figures, largely traceable to sugarcane mask inaccuracies that inflated coverage. In contrast, MODIS global products MCD64A1 and FIRECCI51 heavily underestimated burnt extents, recording 82% and 73% shortfalls relative to OCSB. The machine learning solutions presented here — especially NN — outperformed global alternatives in reproducing spatial patterns and area estimates.

Regional model result uncertainties may arise from sugarcane classification errors and the broad harvesting timeframe adopted. Upcoming efforts will focus on enhancing the sugarcane mask using more extensive ground-truth data for land-cover training and on generating fine-scale harvesting schedules from satellite and field inputs. With machine learning progressing swiftly, innovative architectures such as transformers and established pre-trained networks like DeepLab3+ and U-Net are proving valuable for tasks ranging from fruit ripeness evaluation to crop ailment detection [48-50]. These will be investigated to differentiate between burnt and non-burnt samples. Additionally, testing of the models on additional crops, such as rice, and across other key smallholder regions is scheduled.

Acknowledgements

This research is partly funded by the National Aeronautics and Space Administration (NASA), United States NASA-LCLUC Program to the project titled ‘‘Agricultural Land Use Change in Central and Northeast Thailand: Effects on Biomass Emissions, Soil Quality and Rural Livelihoods’’ (Grant number 80NSSC18K0722) and the U.S. Department of Agriculture, Agricultural Research Service. USDA is an equal opportunity provider and employer.

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 15 Oct 2023 Revised: 20 Dec 2023 Accepted: 05 Jan 2024
Published online: 25 May 2025

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Prato D, Huertas J. Determination of the area affected by agricultural burning. Atmosphere (Basel). 2019;10(6):312. https://doi.org/10.3390/atmos10060312.

2. Shyamsundar P, Springer N, Tallis H, Polasky S, Jat M, Sidhu H, et al. Fields on fire: Alternatives to crop residue burning in India. Science. 2019;365:536. https://doi.org/10.1126/science.aax1728.

3. Anderson-Teixeira K, Davis S, Masters M, DeLucia E. Changes in soil organic carbon under biofuel crops. GCB Bioenergy. 2009;1:75-96.

4. Kumar I, Bandaru V, Yampracha S, Sun L, Fungtammasan B. Limiting rice and sugarcane residue burning in Thailand: Current status, challenges and strategies. J Environ Manag. 2020;276:111228. https://doi.org/10.1016/j.jenvman.2020.111228.

5. Valente F, Laurini M. Pre-harvest sugarcane burning: A statistical analysis of the environmental impacts of a regulatory change in the energy sector. Clean Eng Technol. 2021;4:100255. https://doi.org/10.1016/j.clet.2021.100255.

6. Hall J, Argueta F, Giglio L. Validation of MCD64A1 and FireCCI51 cropland burned area mapping in Ukraine. Int J Appl Earth Obs Geoinf. 2021;102:102437. https://doi.org/10.1016/j.jag.2021.102437.

7. Giglio L, Schroeder W, Justice CO. The collection 6 MODIS active fire detection algorithm and fire products. Remote Sens Environ. 2016;178:31-41. https://doi.org/10.1016/j.rse.2016.02.054.

8. Lasko K, Vadrevu K, Tran V, Ellicott E, Nguyen T, Bui H, et al. Satellites may underestimate rice residue and associated burning emissions in Vietnam. Environ Res Lett. 2017;12.

9. Justice C, Giglio L, Korontzi S, Owens J, Morisette J, Roy D, et al. The MODIS fire products. Remote Sens Environ. 2002;83:244-62.

10. Escuin S, Navarro R, Fernandez P. Fire severity assessment by using NBR (Normalized Burn Ratio) and NDVI (Normalized Difference Vegetation Index) derived from LANDSAT TM/ETM images. Int J Remote Sens. 2008;29:1053-73.

11. Miller J, Thode A. Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR). Remote Sens Environ. 2007;109:66-80.

12. Deshpande M, Pillai D, Jain M. Detecting and quantifying residue burning in smallholder systems: An integrated approach using Sentinel-2 data. Int J Appl Earth Obs Geoinf. 2022;108.

13. Hall J, Loboda T, Giglio L, McCarty G. A MODIS-based burned area assessment for Russian croplands: Mapping requirements and challenges. Remote Sens Environ. 2016;184:506-21.

14. Claverie M, Ju J, Masek J, Dungan J, Vermote E, Roger J, et al. The harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens Environ. 2018;219:145-61.

15. Carreiras J, Quegan S, Tansey K, Page S. Sentinel-1 observation frequency significantly increases burnt area detectability in tropical SE Asia. Environ Res Lett. 2020;15.

16. Gaveau D, Descals A, Salim M, Sheil D, Sloan S. Refined burned-area mapping protocol using Sentinel-2 data increases estimate of 2019 Indonesian burning. Earth Syst Sci Data. 2021;13:5353-68.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

17. Ramo R, Garcia M, Rodriguez D, Chuvieco E. A data mining approach for global burned area mapping. Int J Appl Earth Obs Geoinf. 2018;73:39-51.

18. Cao X, Chen J, Matsushita B, Imura H, Wang L. An automatic method for burn scar mapping using support vector machines. Int J Remote Sens. 2009;30:577-94.

19. Dragozi E, Gitas I, Stavrakoudis D, Theocharis J. Burned area mapping using support vector machines and the FuzCoC feature selection method on VHR IKONOS imagery. Remote Sens (Basel). 2014;6:12005-36.

20. Petropoulos G, Kontoes C, Keramitsoglou I. Burnt area delineation from a uni-temporal perspective based on Landsat TM imagery classification using Support Vector Machines. Int J Appl Earth Obs Geoinf. 2011;13:70-80.

21. Brivio P, Maggi M, Binaghi E, Gallo I. Mapping burned surfaces in sub-Saharan Africa based on multi-temporal neural classification. Int J Remote Sens. 2003;24:4003-18.

22. Gimeno M, San-Miguel-Ayanz J. Evaluation of RADARSAT-1 data for identification of burnt areas in Southern Europe. Remote Sens Environ. 2004;92:370-5.

23. Gomez I, Martin M. Prototyping an artificial neural network for burned area mapping on a regional scale in Mediterranean areas using MODIS images. Int J Appl Earth Obs Geoinf. 2011;13:741-52.

24. Petropoulos G, Vadrevu K, Xanthopoulos G, Karantounias G, Scholze M. A comparison of spectral angle mapper and artificial neural network classifiers combined with Landsat TM imagery analysis for obtaining burnt area mapping. Sensors. 2010;10:1967-85.

25. Pinto M, Libonati R, Trigo R, Trigo I, DaCamara C. A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images. ISPRS J Photogramm Remote Sens. 2020;160:260-74.

26. Chaibandit K, Konyai S, Slack DC. Evaluation of the water footprint of sugarcane in eastern Thailand. Eng J. 2017;21(5):193-201.

27. Athipanyakul T, Choonhawong K, Potchanasin C. The challenge for Thai sugarcane farmers. FFTC Agricultural Policy Platform. 2020. https://ap.fftc.org.tw/article/1840 (accessed on Oct 23, 2022).

28. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory. 1992;5:144-52.

29. Vapnik V. The nature of statistical learning theory. New York: Springer-Verlag; 1995.

30. Vapnik V. Statistical learning theory. New York: Wiley; 1998.

31. Haykin S. Neural networks: A comprehensive foundation. PTR, Prentice Hall; 1994.

32. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychol Rev. 1958;65(6):386.

33. Cox D, Dean T. Neural networks and neuroscience-inspired computer vision. Curr Biol. 2014;24:R921-9.

34. Kaushik S, Kumar M, Arabnia H. Natural language translation system using neural networks. Proceedings of the International Conference on Artificial Intelligence. 2001;1073-80.

35. Dhanjal S. Artificial neural networks in speech processing: Problems and challenges. 2001 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing. 2001. p. 510-13.

© 2025 The Author(s). | Published by Agro Intelligence Research Press

36. Yaramasu R, Bandaru V, Pnvr K. Pre-season crop type mapping using deep neural networks. Comput Electron Agric. 2020;176.

37. NASA Shuttle Radar Topography Mission (SRTM). Shuttle Radar Topography Mission (SRTM) Global. Distributed by OpenTopography; 2013. Accessed 2023 Mar 21. https://doi.org/10.5069/G9445JDF.

38. Bolton D, Gray J, Melaas E, Moon M, Eklundh L, Friedl M. Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. Remote Sens Environ. 2020;240.

39. De Leeuw J, Jia H, Yang L, Liu X, Schmidt K, Skidmore A. Comparing accuracy assessments to infer superiority of image classification methods. Int J Remote Sens. 2006;27:223-32.

40. Kontoes C, Poilve H, Florsch G, Keramitsoglou I, Paralikidis S. A comparative analysis of a fixed thresholding vs. a classification tree approach for operational burn scar detection and mapping. Int J Appl Earth Obs Geoinf. 2009;11:299-316.

41. Office of Cane and Sugar Board (OCSB). Report of the comparison of sugarcane and sugar production efficiency of sugar factories in Thailand in the production year 2019/20. Bangkok, Thailand: Office of Cane and Sugar Board; 2020. http://www.sugarzone.in.th. Accessed 2022 Oct 12.

42. Huang C, Davis L, Townshend J. An assessment of support vector machines for land cover classification. Int J Remote Sens. 2002;23:725-49.

43. Yekkehkhany B, Safari A, Homayouni S, Hasanlou M, Karimipour F, Abbaspour R. A comparison study of different kernel functions for SVM-based classification of multi-temporal polarimetry SAR data. 1st ISPRS International Conference on Geospatial Information Research. 2014;40:281-5.

44. Shao Y, Lunetta R. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. ISPRS J Photogramm Remote Sens. 2012;70:78-87.

45. Li Y, Feng Z, Chen S, Zhao Z, Wang F. Application of the artificial neural network and support vector machines in forest fire prediction in the Guangxi Autonomous Region, China. Discrete Dyn Nat Soc. 2020.

46. Raczko E, Zagajewski B. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. Eur J Remote Sens. 2017;50:144-54.

47. Chuvieco E, Pettinari ML, Lizundia-Loiola J, Storm T, Padilla Parellada M. ESA Fire Climate Change Initiative (Fire_cci): MODIS Fire_cci burned area pixel product, version 5.1. Centre for Environmental Data Analysis; 2018. https://doi.org/10.5285/58f00d8814064b79a0c49662ad3af537.

48. Bose K, Shubham K, Tiwari V, Patel KS. Insect image semantic segmentation and identification using UNET and DeepLab V3+. Singapore: Springer Nature Singapore; 2023. p. 703-11.

49. Cai C, Tan J, Zhang P, Ye Y, Zhang J. Determining strawberries' varying maturity levels by utilizing image segmentation methods of improved DeepLabV3+. Agronomy. 2022;12.

50. Li K, Zhao W, Peng R, Ye T. Multi-branch self-learning vision transformer (MSViT) for crop type mapping with optical-SAR time-series. Comput Electron Agric. 2022;203:107497.

© 2025 The Author(s). | Published by Agro Intelligence Research Press