MAPPING AND CLASSIFICATION OF INVASIVE TREE SPECIES CHINESE CELTIS IN RIPARIAN ECOSYSTEMS

Aranya Jha
Faculty ITC
University of Twente
Enschede, Netherlands
aranyajha.2000@gmail.com

Dr. Armando Apan
School of Surveying and Built Environment
University of Southern Queensland
Toowoomba, Australia
armando.apan@unisq.edu.au

Dr. Bikram Pratap Banerjee
School of Surveying and Built Environment
University of Southern Queensland
Toowoomba, Australia
bikram.banerjee@unisq.edu.au

***Abstract*—Chinese Celtis are tree species that have, over time, become a threat to the flora and faunal habitat of native Australian species in New South Wales and Queensland. Early detection has become essential for monitoring its spread and implementing targeted management strategies for the conservation of native ecosystems. Observing the distribution of this weed is made easier by data collected and processed using remote sensing methods. Furthermore, high-resolution images are favoured since they are more likely to differentiate Chinese Celtis from other native grasses and trees. In this research project, we investigate the use of UAV-acquired very high-resolution multispectral imagery to map the spatial distribution of invasive trees in a relatively small region along the Brisbane River. The spectral details of the vegetation were emphasised by calculating vegetation indices like NDVI and NDWI, while spatial patterns were quantified by extracting texture-based features. Several machine-learning models were then trained for classification using these features as input. AutoML tool in ArcGIS Pro was employed to automate model selection and tuning, helping streamline the workflow. Metrices such as confusion matrix, precision, recall and f1-score were used for model evaluation. Among the tested models, LightGBM (Light Gradient Boosting Machine) achieved the best classification performance of 96.64%. While UAVs are limited in large-scale coverage, their use in localised or high-priority areas provides an efficient and cost-effective option for detailed ecological monitoring. Future research could combine UAV imagery with images from other sensors, like LiDAR or hyperspectral, to improve classification accuracy in case of larger regions and address similar issues occurring in the environment.**

Keywords—Chinese Celtis, mapping invasive species, high-resolution UAV imagery, machine learning classification

I. INTRODUCTION

Chinese Celtis (Celtis sinensis) is an invasive tree species primarily distributed across Queensland and New South Wales, Australia, and is capable of reaching heights of up to 20 meters [1][2]. Native to East Asia, including China, Japan, Korea, and Taiwan [1]–[4], the species was initially introduced to Australia for ornamental purposes. However, it has since become widespread, with the earliest recorded occurrence in Queensland dating back to 1912 [4]. Due to morphological similarities, Chinese Celtis is frequently misidentified as Chinese Elm, Nettle Tree, Hackberry, or Native Silky Celtis [1][3][4].

The species exhibits deciduous or semi-deciduous characteristics, shedding its leaves between late winter and early spring, often coinciding with its flowering period. The fruits develop during spring and summer and reach maturity in autumn and early winter [2]. Chinese Celtis thrives in moist environments, favoring clay-rich soils along streams, rivers, and near rainforest margins [1][4]. Its rapid propagation is facilitated by seed dispersal through water and frugivorous birds and bats, enabling the formation of dense thickets in riparian zones [1]. This proliferation suppresses native ground vegetation, disrupts local ecosystems, and exacerbates ecological degradation.

The ecological impact of Chinese Celtis extends beyond competition with native species. It utilizes essential resources such as groundwater, nutrients, and sunlight, adversely affecting croplands. Furthermore, its dominance in riparian zones displaces native fauna and alters riverbank vegetation. The species also contributes to water contamination through the deposition of decaying leaves into nearby water bodies [3][4]. Management approaches include manual removal and chemical control, such as stem injection, foliar spraying, and basal barking. Young saplings can be effectively removed through manual extraction, whereas larger infestations require targeted chemical interventions [1][4].

Given the invasive nature and ecological impact of Chinese Celtis, accurate mapping of its distribution is essential for developing effective management and control strategies. The species' dominance in riparian zones accelerates soil erosion, alters hydrological regimes, and creates significant ecological imbalances. Although the impacts of Chinese Celtis and associated management practices have been well-documented, there is a notable gap in the application of remote sensing techniques for its detection and monitoring in Australia.

Remote sensing has proven to be a valuable tool for monitoring invasive species. For example, the spread of Pinus caribaea in Sri Lanka was monitored over a 21-year period using Landsat imagery [5]. Similarly, UAV hyperspectral imagery with a spatial resolution of 2 cm was employed to map invasive noxious weed species in alpine grasslands [6]. Recent studies emphasize the importance of understanding species-specific ecology to design tailored remote sensing approaches and address the existing gap between researchers and practitioners,

particularly in the Pacific Islands region [7]. High-resolution UAV imagery and machine learning techniques have been effectively used to detect exotic conifers, such as Pinus sylvestris and P. ponderosa in New Zealand [8], and Hakea sericea in Portugal using object-oriented classification methods [9].

The coexistence of invasive species with native vegetation presents significant challenges for their accurate identification and monitoring. High-resolution UAV imagery, combined with advanced sensors such as multispectral and LiDAR, offers the potential to overcome these challenges [6][10]. Multispectral sensors facilitate vegetation analysis, while LiDAR provides critical information on canopy height and structure. Mapping workflows typically involve data acquisition, preprocessing, feature extraction, classification, and validation [5]. This study aims to map the distribution of Chinese Celtis in the Brisbane region using very high-resolution UAV multispectral imagery (7cm) and machine learning classification techniques to provide a solution for invasive species monitoring.

II. Materials and Method

A. Study Area and Data

The study area selected for this research is located near Brisbane, Queensland, Australia (Fig. 1), where the presence of Celtis sinensis has been documented. High-resolution aerial imagery was acquired using an unmanned aerial vehicle (UAV) equipped with a Micasense Altum sensor. The UAV operated in automated mode, with forward and side overlaps set at 85% to achieve comprehensive spatial coverage and minimize gaps in the dataset. Positional data, including altitude, latitude, and longitude, were recorded through an integrated global positioning system (GPS). The Micasense Altum sensor collected multispectral imagery across six distinct spectral bands: blue (475 nm), green (560 nm), red (668 nm), red-edge (717 nm), near-infrared (842 nm), and thermal. This multispectral configuration facilitated the extraction of spectral information critical for vegetation classification and species-level differentiation. To enhance geometric precision and radiometric fidelity, ground control points (GCPs) were strategically deployed within the study area. Five permanent GCPs were established, with their exact coordinates measured using a high-precision global navigation satellite system-real-time kinematic (GNSS-RTK) receiver. These GCPs were incorporated during image processing to perform geometric corrections and radiometric calibration, thereby ensuring the accuracy and reliability of the UAV-derived data.

B. UAV Data Preprocessing

Following the acquisition of raw imagery using the Micasense Altum sensor, data preprocessing was conducted using Pix4D Mapper. The preprocessing workflow included radiometric and geometric corrections to ensure data accuracy and consistency. Pix4D Mapper employs the structure-from-motion (SfM) technique to process the calibrated images captured during the UAV flight mission. This approach enabled the generation of a reflectance orthomosaic, a digital surface model (DSM), and a digital terrain model (DTM). Camera parameters were fine-tuned by utilizing target points, which were matched and triangulated across overlapping images.

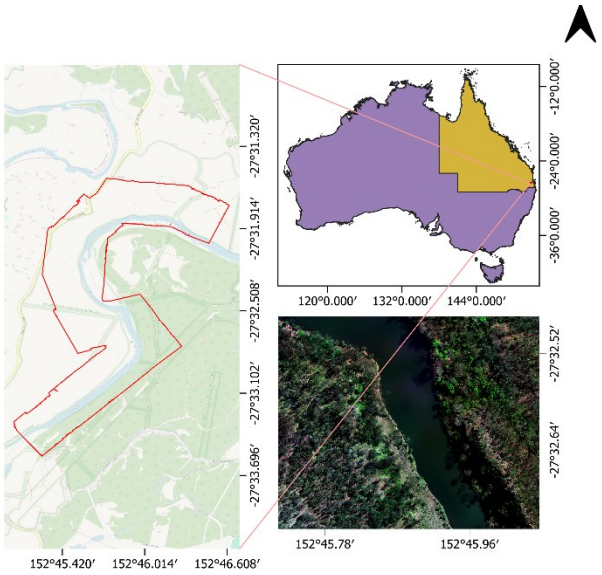

Fig. 1. Region of Interest - near Brisbane, Queensland

A bundle adjustment algorithm was applied to optimize image alignment, creating a sparse point cloud by linking generalized keypoints across the image set. Ground control points (GCPs) collected during the aerial survey were integrated into this process to enhance geometric precision. These GCPs ensured accurate registration of the imagery to ground coordinates. To construct the reflectance orthomosaic, overlapping features in the images were matched using computed keypoints, ensuring seamless integration of individual frames. The orthomosaic was subsequently exported in GeoTIFF (.tif) format for further analysis. The bundle block adjustment was refined to improve the orientation and alignment of the densified point cloud, resulting in the generation of DSM and DTM products, which were also exported in rasterized .tif formats. Finally, the multispectral bands captured by the sensor were mosaiced into a unified multispectral image stack. This stack formed the basis for feature extraction and subsequent analytical processes, ensuring that the data were ready for classification and modeling tasks.

C. Machine Learning Classification

The classification process in this study involved a structured workflow comprising training sample creation, feature extraction, and model training to identify and map Celtis sinensis within the region of interest using high-resolution UAV multispectral imagery. Training samples were created to represent five distinct land cover classes observed within the study area: Chinese Celtis, bare ground, water bodies, other native vegetation, and sedge grass. High-resolution imagery guided the delineation process to ensure that the samples captured the variability within each class and provided consistent representation across all categories. Initially, 212 polygons representing five land cover classes were manually created but this sample size was insufficient for robust model training. This issue was addressed by generating approximately 173,000 points withing the polygon. The dataset was then divided using a 70:30 ratio, allocating 70% for training and 30% for validation purposes.

To enhance classification accuracy, feature extraction was performed to derive key spectral and textural attributes. A total of eleven vegetation indices and eight textural metrics were calculated from the multispectral imagery. Vegetation indices, including the Normalized Difference Vegetation Index (NDVI), Soil-Adjusted Vegetation Index (SAVI), and others, were computed to highlight vegetation health, density, and biophysical properties, thereby aiding in the differentiation of Celtis sinensis from native vegetation [11], [12] and other land cover types. These indices provided critical spectral information for species-level classification (Table I).

Table I. List of radiometric indices used

	Indices	Equation
Vegetation	NDVI (Normalized Difference Vegetation Index)	$\frac{(NIR - RED)}{(NIR + RED)}$
	RVI (Ratio Vegetation Index)	$\frac{(NIR)}{(RED)}$
	SAVI (Soil-Adjusted Vegetation Index)	$\frac{1.5 * (NIR - RED)}{(NIR + RED + 0.5}$
	GEMI (Global Environmental Monitoring Index)	$\eta * (1 - 0.25 * \eta) - \frac{RED - 0.125}{1 - RED}$
	IPVI (Infrared Percentage Vegetation Index)	$\frac{NIR}{NIR + RED}$
	LAIFromNDVILog (Leaf Area Index)	a*log(NDVI)
Water	NDWI Normalized Difference Water Index)	$\frac{GREEN - NIR}{GREEN + NIR}$
Soil	RI (Redness Index)	$\frac{RED^2}{BLUE * GREEN}$
	CI (Colour Index)	$\frac{RED - BLUE}{RED + BLUE}$
	BI (Brightness Index)	$\frac{\sqrt{RED^2 + GREEN^2}}{2}$

Textural features were extracted using the Haralick texture analysis method, which quantifies the spatial arrangement of pixel intensities within the imagery. Metrics such as contrast, entropy, correlation, and homogeneity were computed to capture fine-scale textural differences between the land cover classes. These textural metrics complemented the spectral data by providing insights into surface patterns and heterogeneity, improving the model's ability to discriminate between classes with similar spectral characteristics. The final models were assessed based on their performance in identifying Celtis sinensis and other land cover classes, ensuring minimal misclassification and high reliability. The extracted features served as inputs for training machine learning models aimed at classifying land cover types with a focus on detecting Celtis sinensis. The training process incorporated automated machine learning (AutoML) techniques to streamline model selection, hyperparameter optimization, and validation. This automated process enabled the systematic evaluation of various algorithms, including decision trees, gradient boosting methods like XGBoost and LightGBM, and a final ensemble model that combined predictions of multiple base models tested during the AutoML process, to identify the most effective classifier for the dataset.

The AutoML framework streamlined the machine learning workflow by automating key steps such as model comparison and hyperparameter tuning. This approach not only reduced manual intervention but also optimized the models to achieve high classification accuracy while maintaining computational scalability. By leveraging advanced feature selection and optimization techniques, the workflow ensured that the models were fine-tuned to effectively exploit both spectral and textural attributes.

III. Results and Discussion

The detection and classification of Celtis sinensis and other land cover types using high-resolution multispectral UAV imagery yielded highly accurate results. Model performance was rigorously evaluated using multiple metrics, including precision-recall curves, receiver operating characteristic (ROC) curves, confusion matrices, and key statistical measures such as precision, recall, and F1-score. These evaluations provided a comprehensive assessment of the models' predictive accuracy and robustness.

A. Results

The AutoML-generated ensemble model achieved a log loss of 0.093 with an efficient training time of 0.9 seconds. Among individual models, LightGBM demonstrated the lowest logloss (0.093) and was identified as the best individual model by the AutoML framework despite a longer training time (59.24 seconds). In contrast, the decision tree model demonstrated inferior predictive performance, recording a log loss of 0.404. The learning curves for both training and testing datasets indicated steady convergence for the ensemble and LightGBM models, suggesting well-optimized solutions with minimal overfitting.

The precision-recall curve (Fig. 2) highlights the model's strong capability to distinguish between land cover classes, with an overall Area Under the Curve (AUC) value of 0.995. Class-wise AUC values ranged from 0.988 to 1.000, underscoring the model's ability to generalize across different categories. Similarly, the ROC analysis (Fig. 3) further corroborated the model's accuracy, with all class-specific AUC values reaching 1.000.

The confusion matrix (Fig. 4) demonstrated the high effectiveness of the LightGBM model in classifying Celtis sinensis, with minimal instances of misclassification. The precision and recall values for the target class were both 0.9506, resulting in an F1-score of 0.9506 (Table II). This balanced performance indicates that the model minimized both false positives and false negatives, ensuring reliable identification of the invasive species. These results affirm the robustness of the methodology and highlight the potential of integrating spectral and textural features with advanced machine-learning techniques for ecological monitoring.

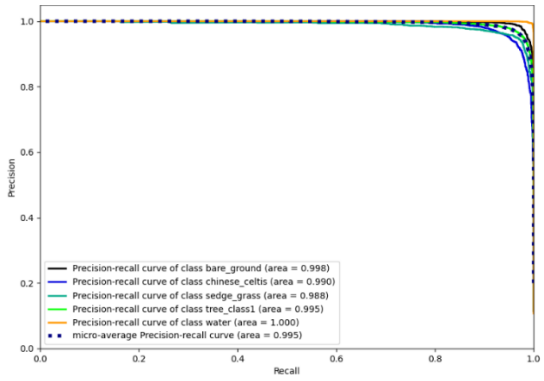

Fig. 2. Precision-Recall curve for the five landcover classes

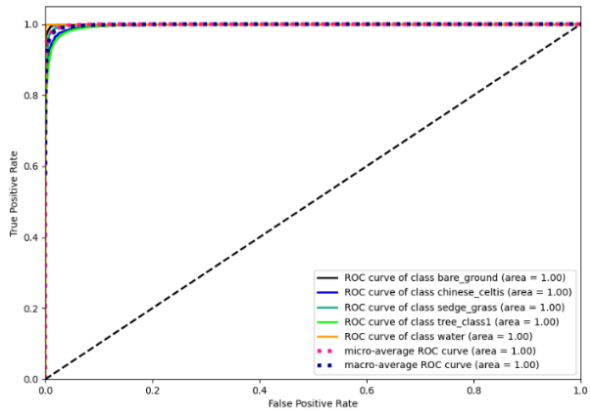

Fig. 3. ROC Curve for the five landcover classes

Table II. Metric details for LightGBM model

	Bare Ground	Chinese Celtis	Sedge Grass	Tree class 1	Water	Accuracy
Precision	0.975	0.951	0.947	0.970	0.991	0.966
Recall	0.976	0.951	0.969	0.962	0.997	0.966
F1-Score	0.975	0.950	0.958	0.966	0.994	0.966

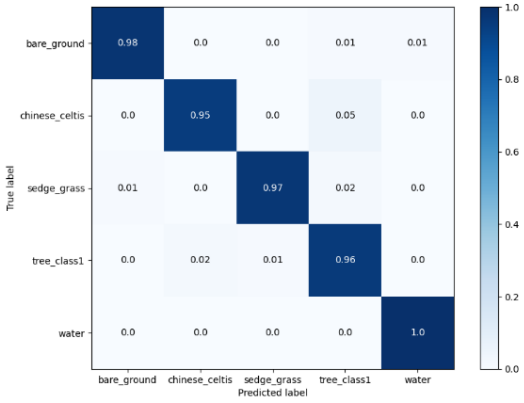

Fig. 4. Normalized Confusion Matrix

B. Discussion

The application of high-resolution multispectral UAV imagery, integrated with advanced feature extraction techniques and automated machine learning, demonstrated substantial efficacy in mapping invasive species such as Celtis sinensis. The models developed through the AutoML framework were evaluated using a suite of statistical and graphical performance measures, including ROC and precision-recall curves, log loss, and confusion matrices. These metrics collectively confirmed the strong predictive capability of the models, particularly the LightGBM classifier and the AutoML-generated ensemble model. Amongst all the tested models, the LightGBM algorithm performed the best overall achieving 96% classification accuracy.

The high AUC values across all classes (ranging from 0.988 to 1.000) and the precision-recall AUC of 0.995 underscore the model's ability to generalize well across land cover types. The ROC curves which are based on predicted class probabilities, illustrate consistent performance across classification thresholds, further validating the model's robustness. The confusion matrix showed that misclassification of Celtis sinensis was minimal, with a strong F1-score of 0.9506, suggesting that both precision and recall were well-balanced. However, minor misclassifications were observed, particularly between Celtis sinensis and Tree Class 1, likely attributable to overlapping spectral signatures between these vegetation types, emphasizing the need for incorporating additional data modalities or refined feature sets to enhance class separability

While the study included both radiometric indices and texture features as input variables, the individual contribution of each feature type was not separately evaluated. Future work could incorporate feature importance analysis or stepwise feature elimination to isolate and evaluate the impact of specific input variables.

IV. Conclusion

Celtis sinensis, an invasive tree species, poses a growing threat to ecosystems in eastern Australia. In this study, we combined vegetation indices, texture-based metrics, and machine learning techniques to map its distribution with high accuracy. The use of automated model selection and parameter tuning reduced the manual workload, making the approach more accessible for users without a deep background in machine learning. Although the use of UAV imagery limits large-scale application, the workflow can serve as a foundation for future efforts that combine UAV and satellite imagery (LiDAR or hyperspectral sensors) to enable regional-scale monitoring. This streamlined process supports faster and more affordable ecological monitoring. The findings underline how remote sensing and data-driven methods can work together to improve conservation efforts, offering a practical tool for detecting and managing invasive species.

References

[1] T. R. Armstrong and S. L. Keegan, "Celtis sinensis and its control," *Proc. 11th Aust. Weeds Conf.*, vol. 30, pp. 504–505, 1996.

[2] R. Ensbey, "Chinese celtis : identification and control," vol. 922, no. January, pp. 1–4, 2010.

[3] C. J. O'Brien, V. Mellor, and V. J. Galea, "Controlling Woody Weed Chinese Elm (Celtis sinensis Pers.) with Stem-Implanted Herbicide

Capsules," *Plants*, vol. 11, no. 3. 2022. doi: 10.3390/plants11030444.

[4] N. S. W. Department of Primary Industries, "Chinese Celtis," Weeds. Accessed: Jan. 14, 2025. [Online]. Available: https://weeds.dpi.nsw.gov.au/Weeds/ChineseCeltis

[5] W. D. K. V. Nandasena, L. Brabyn, and S. Serrao-Neumann, "Monitoring invasive pines using remote sensing: a case study from Sri Lanka," *Environ. Monit. Assess.*, vol. 195, no. 2, 2023, doi: 10.1007/s10661-023-10919-1.

[6] F. Xing, R. An, X. Guo, and X. Shen, "Mapping invasive noxious weed species in the alpine grassland ecosystems using very high spatial resolution UAV hyperspectral imagery and a novel deep learning model," *GIScience Remote Sens.*, vol. 61, no. 1, 2024, doi: 10.1080/15481603.2024.2327146.

[7] C. M. H. Chan, C. J. Owers, S. Fuller, M. W. Hayward, D. Moverley, and A. S. Griffin, "Capacity and capability of remote sensing to inform invasive plant species management in the Pacific Islands region," *Conserv. Biol.*, no. September 2023, pp. 1–19, 2024, doi: 10.1111/cobi.14344.

[8] J. P. Dash, M. S. Watt, T. S. H. Paul, J. Morgenroth, and G. D. Pearse, "Early detection of invasive exotic trees using UAV and manned aircraft multispectral and LiDAR Data," *Remote Sens.*, vol. 11, no. 15, pp. 1–21, 2019, doi: 10.3390/rs11151812.

[9] F. Alvarez-Taboada, C. Paredes, and J. Julián-Pelaz, "Mapping of the invasive species Hakea sericea using Unmanned Aerial Vehicle (UAV) and worldview-2 imagery and an object-oriented approach," *Remote Sens.*, vol. 9, no. 9, 2017, doi: 10.3390/rs9090913.

[10] K. K. Singh, T. D. Surasinghe, and A. E. Frazier, "Systematic review and best practices for drone remote sensing of invasive plants," *Methods Ecol. Evol.*, vol. 15, no. 6, pp. 998–1015, 2024, doi: 10.1111/2041-210X.14330.

[11] P. A. Dmitriev *et al.*, "Identification of species of the genus Acer L. using vegetation indices calculated from the hyperspectral images of leaves," *Remote Sens. Appl. Soc. Environ.*, vol. 25, p. 100679, 2022, doi: https://doi.org/10.1016/j.rsase.2021.100679.

[12] W. C. Lyons, G. Buddhi R., C. Richard, A. Shikha, G. Maheteme, and K. and Andries, "Evaluating spectral properties of invasive plant species in Kentucky recreation areas," *Int. J. Remote Sens.*, vol. 45, no. 18, pp. 6470–6494, Sep. 2024, doi: 10.1080/01431161.2024.2391095.