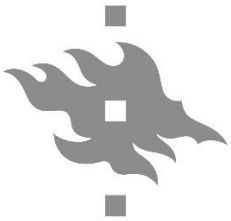

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Master's thesis in Geography

Geoinformatics

Spectral Footprints of Change: Mango Tree Distribution and Climate Change Mitigation Potential in Taita Hills, Kenya

Saana Järvinen

2025

Supervisors:

Prof. Petri Pellikka
Zhaozhi Luo

Master's Programme in Geography

Faculty of Science

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty	Osasto – Institution – Department
Faculty of Science	Department of Geosciences and Geography

Tekijä – Författare – Author

Saana Järvinen

Tutkielman otsikko – Avhandlings titel – Title of thesis

Spectral footprints of change: Mango tree distribution and climate change mitigation potential in Taita Hills, Kenya

Koulutusohjelma ja opintosuunta – Utbildningsprogram och studieinriktning – Programme and study track

Master's programme in geography, Geoinformatics

Tutkielman taso – Avhandlings nivå – Level of the thesis	Aika – Datum – Date	Sivumäärä – Sidoantal – Number of pages
Master's thesis, 30 credits	April, 2025	60 + 1 appendix

Tiivistelmä – Referat – Abstract

Trees play a significant role in mitigating climate change by storing carbon and releasing oxygen through photosynthesis. In Taita Hills, Kenya, *Mangifera indica*, commonly known as mango, is a key fruit-bearing tree species second only to avocado (*Persea americana*). Mango trees are predominantly cultivated by farmers in the lowlands, providing not only economic and food security but also shade to benefit both agricultural activities and residences. However, climate change is altering the region's vegetation dynamics. This shift is causing the highland climates of Taita Hills to become similar to those of the lowlands, creating new ecological niches suitable for mango cultivation at higher elevations.

Previous research has successfully utilized hyperspectral data for vegetation mapping, distinguishing mangoes from other species with high accuracy. However, these studies have been limited in scope and quantity, focusing only on small, localized areas or specific invasive species. This study aims to fill this gap in the literature and at the same time provide a deeper understanding of the patterns in mango tree distribution in mountainous terrain.

This study investigated the impact of climate change on the distribution and spreading of mango trees in Taita Hills. It followed the Positive-Unlabeled (PU) deep learning method for the one-class classification model training. The framework utilized airborne hyperspectral and LiDAR data. A spatiotemporal analysis was conducted for the study period from 2013 to 2022. Output of the classification model visualized a prediction of mango tree canopy pixels, and when compared between the years a change detection map was produced for the research area. In addition to classification, aboveground biomass (AGB) and aboveground carbon storage (AGCS) estimates were studied with allometric calculations using the trees' field measurements.

The results showed an increase in mango trees in the study area as well as a spreading pattern upwards in elevation. The classification model performed well in detecting mango trees within their canopy areas. The best F1 score was 0,8 for 2013 classification, and 0,86 for 2022. The model received an AUC score of 0,97 and 0,95 respectively for the years. The AGB and AGCS estimations for the measured mango trees showed plausible results, and a vague but linear negative connection between AGCS of mango trees and elevation was found. This thesis will provide new perspectives on tree distribution research using deep learning and hyperspectral remote sensing methods as well as its possibilities in climate change mitigation strategies.

Avainsanat – Nyckelord – Keywords

Taita Hills, Remote sensing, Hyperspectral imaging, Tree species mapping

Säilytyspaikka – Förvaringställe – Where deposited

University of Helsinki electronic theses library E-thesis/HELDA

Muita tietoja – Övriga uppgifter – Additional information

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty	Osasto – Institution – Department
Matemaattis-luonnontieteellinen tiedekunta	Geotieteiden ja maantieteen osasto

Tekijä – Författare – Author

Saana Järvinen

Tutkielman otsikko – Avhandlings titel – Title of thesis

Muutoksen spektriset jalanjäljet: Mangopuiden levinneisyys sekä ilmastonmuutoksen hillitsemispotentiaali Kenian Taitavuorilla

Koulutusohjelma ja opintosuunta – Utbildningsprogram och studieinriktning – Programme and study track

Maantieteen maisteriohjelma, Geoinformatiikka

Tutkielman taso – Avhandlings nivå – Level of the thesis	Aika – Datum – Date	Sivumäärä – Sidoantal – Number of pages
Maisterintutkielma, 30 opintopistettä	Huhtikuu, 2025	60 + 1 liite

Tiivistelmä – Referat – Abstract

Puilla on merkittävä vaikutus ilmastonmuutoksen hillinnässä niiden hiilensidonnan sekä fotosynteesin kautta tapahtuvan hapentuoton kautta. Mango (*Mangifera indica*) on avokadon jälkeen yksi tärkeimmistä hedelmäpuista Taitavuorilla Keniassa. Maanviljelijät kasvattavat mangoja pääosin Taitan alankoalueilla, joilla puu tarjoaa ruoka- sekä taloudellisen turvan lisäksi myös varjoa hyödyntämään niin asumuksia kuin maataloustoimintaakin. Ilmastonmuutos kuitenkin muuttaa alueen kasviston dynamiikkaa, ja kasvuolosuhteet ylängöillä ovat alkaneet muistuttaa alankoalueille tyypillisiä olosuhteita. Ilmestyy uusia ekolokeroita, jolloin mangoa voi viljellä yhä korkeammilla alueilla.

Aiemmissa tutkimuksissa on menestyksekkäästi hyödynnetty hyperspektraalista dataa kasvikartoitukseen, joissa mangot on onnistuttu tarkasti erottamaan muista lajeista. Tällaiset tutkimukset ovat kuitenkin laajuudeltaan sekä volyymiltaan hyvin rajoittuneita keskittyen pieniin paikallisiin alueisiin, tai vain tiettyihin vieraslajeihin. Tämän tutkimuksen tavoitteena on täyttää kyseistä aukkoa tietokirjallisuudessa, sekä samalla luoda syvempää käsitystä mangopuiden levinneisyydestä vuoristoisessa maastossa.

Tässä tutkimuksessa selvitetään ilmastonmuutoksen vaikutusta mangopuiden levinneisyyteen sekä altitudisesti ylöspäin johtavaan leviämiseen Taitavuorilla. Tutkimuksessa hyödynnetään Positive-Unlabeled (PU) -oppimiseen pohjautuvaa metodia yksiluokkaisen luokittelumallin syväoppimiskoulutuksessa. Tutkimuksen taustalla on ITreeDet-prosessirakenne, joka hyödyntää kaukokartoituksen metodeja käyttäen hyperspektraalista ilmakuvadataa yhdessä LiDAR-datan kanssa. Luokittelumalli tuottaa ennusteen mangopuiden latvuspikselien sijainneista. Tutkimusjaksona toimivat vuodet 2013 ja 2022, ja muutoskartta koko tutkimusalueelle saatiin vertaamalla näiden kahden vuoden ennustetta toisiinsa. Luokittelumallin ennusteen lisäksi arvioitiin mangopuiden sitoman maanpäällisen biomassan sekä hiilipitoisuuden määrää sovittamalla puista tehdyt maastomittaukset mangopuille kehitettyihin allometrisiin yhtälöihin.

Tulokset osoittavat mangopuiden määrän lisääntyneen tutkimusalueella, sekä viittaavat puiden leviämiseen korkeussuunnassa ylöspäin. Luokittelumalli tunnisti hyvin mangopuut niiden latvusalueiltaan. Paras F1-luku oli 0,8 vuoden 2013 luokittelumallissa ja 0,86 vuoden 2022. Malli ylsi AUC-lukuihin 0,97 ja 0,95 vuosille tässä järjestyksessä. Maastossa mitattujen mangopuiden maanpäällisen biomassan sekä hiilipitoisuuden estimointi tuotti myös otollisia tuloksia. Suurpiirteinen, mutta suoraviivainen negatiivinen korrelaatio löytyi puiden ja korkeuden välille. Tämä tutkimus käyttää syväoppimista sekä hyperspektraalisen kaukokartoituksen metodeita tuottamaan uusia näkökulmia puiden levinneisyyden tutkimukseen, sekä sen hyödyntämismahdollisuuksiin ilmastonmuutoksen hillitsemisstrategioissa.

Avainsanat – Nyckelord – Keywords

Taitavuoret, Kaukokartoitus, Hyperspektraalinen data, Puulajien kartoitus

Säilytyspaikka – Förvaringställe – Where deposited

University of Helsinki electronic theses library E-thesis/HELDA

Muita tietoja – Övriga uppgifter – Additional information

Table of contents

List of Abbreviations

AGB - Aboveground Biomass

AGCS - Aboveground Carbon Storage

ALS - Airborne Laser Scanning

AP - Average Precision

AUC - Area Under (the ROC) Curve

CD - (tree) Collar Diameter

CHM - Canopy Height Model

CNN - Convolutional Neural Network

CSC - Center for Scientific Computing

DBH - Diameter at Breast height

DGNSS - Differential Global Navigation Satellite System

DPB - Diameter of Primary Branches

DTM - Digital Terrain Model

FACE - Free-air CO2 enrichment

HRS - Hyperspectral Remote Sensing

IVI - Importance Value Index

km - kilometer

LAI - Leaf Area Index

LiDAR - Light Detection and Ranging

m - meter

m a.s.l. - meters above sea level

MNF - Minimum Noise Fraction

nm - nanometer

NDMI - Normalized Difference Moisture Index

NDVI - Normalized Difference Vegetation Index

OA - Overall Accuracy

OCC - One-Class Classification or One-Class Classifier

PN - Positive and Negative (dataset)

PR - Precision–Recall

PU - Positive and Unlabeled (dataset/learning)

puAUC - AUC for positive and unlabeled samples

QGIS - Quantum GIS

RGB - Red, Green and Blue

ROC - Receiver Operating Characteristic

SNR - Signal-to-Noise-Ratio

UAV - Unmanned Aerial Vehicle

UHI - Urban Heat Islands

1. Introduction

Africa is facing the effects of climate change stronger and faster than the global average. Even though there is no Africa-wide climate effect, the whole continent is adversely affected, and many regions of Africa will suffer from more frequent and intense droughts and floods (Collier, Conway, & Venables, 2008). In Kenya the likelihood, duration and frequency of droughts will likely decrease (Haile et al., 2020). Therefore, it is important to be able to model the phenomena that influence climate change mitigation in the region.

Trees and their distribution play a significant role in climate change. While carbon dioxide acts as the prime factor of climate change, the plantae in turn sequester greenhouse gases. As trees store such a large amount of biomass, their role in storing carbon and releasing oxygen through photosynthesis is notable.

Remote sensing is a collection of methods that can be used to map trees. Remote sensing means acquiring electromagnetic radiation information of a geographical region remotely. In practice this is usually done with aerial or satellite imagery by measuring the radiance of visible and infrared bands. Therein, hyperspectral data is characterized by a large number of narrow and contiguous spectral bands. Hyperspectral data is thus well-suited for tree species mapping, as it enables the distinction of subtle spectral differences between species.

This thesis has a focus on the agroforestry of the mountainous Kenyan region of Taita Hills. Mango trees are an essential species that are cultivated nearly throughout the whole area, with significant impact on the local environment. Recent observations show that the trees are becoming more common in the highlands of Taita Hills, where they were rare before (Mjomba, 2023). The shift is caused by the warming climate, when the highlands are beginning to resemble growing habitats similar to the region's lowlands. Due to this phenomenon, mango tree distribution is a good indicator of climate change on local vegetation.

ITreeDet is a classification model based on hyperspectral imagery that can be used to track the distribution of tree species. The model is used in this thesis to detect changes in mango tree distribution between the years 2013 and 2022. As a part of the study, a field trip to Kenya and Taita Hills was conducted, during which tree locations were mapped for classifier source

data. In addition, mango tree measurements such as tree height were taken of the visited trees, as well as ground temperature measurements under and outside the mango tree canopy, along with interviews from the mango tree cultivators. The aims of this study are both to monitor the evolution of the distribution of mango trees in Taita Hills, and to examine the suitability of one-class classification method for mango tree mapping purposes. Moreover, the mango trees' contribution to aboveground biomass and carbon content is estimated along with the canopy's cooling effect to its surroundings. This thesis studies the interactive role of agroforestry of mango trees in climate change mitigation strategies.

1.1. Agroforestry and climate change in Taita Hills

"Agroforestry is the intentional integration of trees and shrubs into crop and animal farming systems to create environmental, economic, and social benefits."
(USDA, 2011)

While the term agroforestry is relatively new, the practice itself is old. Agroforestry provides various ecosystem services as well as enhances biodiversity. The goal is to achieve both economic and environmental benefits for farmers, nature and society. (Pantera, Mosquera-Losada, Herzog, & Den Herder, 2021)

Agroforestry is an important bywork, and integrating trees into agricultural systems offers multiple benefits. CO_2 acts as the main component of the greenhouse effect, which has been studied to be the root cause of climate change (Mitchell, 1989). Considering the significant amounts of carbon that is stored, trees can be seen as climate change mitigators. Furthermore, trees provide shade, which has a cooling effect on the surrounding areas. The shade provision is an important factor, especially if the tree is tall and the leaf area dense (Manickathan, Defraeye, Allegrini, Derome, & Carmeliet, 2018). Trees also reduce evaporation of the soil, which helps to conserve water and control erosion. Because of their roots, the soil is better kept intact meanwhile the trees also act as a windshield. Flowering trees attract pollinators, notably so if they are not adapted to wind pollination (Kumar, 2016). Falling leaves can be used as litter or food for the livestock and tree pruning produces firewood and timber, which are also available for sale. In multi-species farms the litter and roots of the trees improve both physical

and chemical properties of the soil, which minimizes the often seen decline of soil fertility (Albrecht & Kandji, 2003). Selecting fruit- or nut-bearing species allows harvests to provide food as well as economic safety.

Taita Hills is a topographically complex mountainous area in South-Eastern Kenya with high biodiversity both in flora and fauna. Agroforestry is present throughout the whole area of Taita Hills, and especially towards the west side according to the land cover map of 2020 (Abera, Vuorinne, Munyao, Pellikka, & Heiskanen, 2022). Climate change produces threats in the Taita Hills area, as agriculture remains to be the main source of income for most of the inhabitants. In Taita Hills, land cover change is prominent (Figure 1.). Taita-Taveta county has been through severe droughts, and when the crops have failed, agroforestry has saved the economy of many small farmers. The exotic species are popular, because they grow faster and stronger than the indigenous trees. Because of their soft and straight trunk, they are commonly used as building material and usually produce more than the local ones. The most significant species for agroforestry in Taita Hills, in order, are silver oak (*Grevillea robusta),* mango (*Mangifera indica),* avocado (*Persea americana)* and eucalyptus (*Eucalyptus spp.)* (Amara, Adhikari, Mwamodenyi, Pellikka, & Heiskanen, 2023). The landscape characteristics of Taita Hills can be studied from Figure 2.

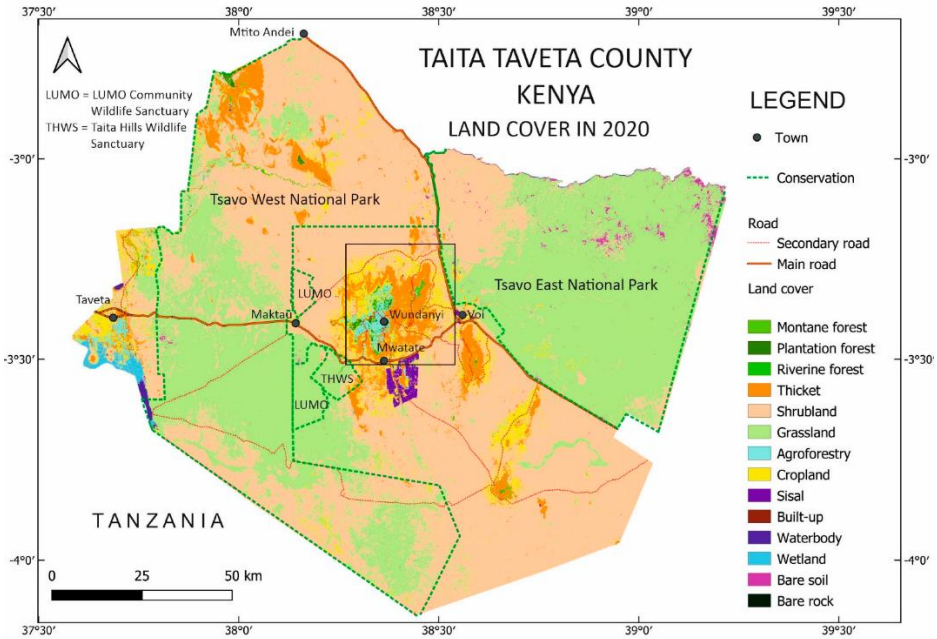

Figure 1. Land cover map of Taita Hills (Abera et al., 2022).

Figure 2. Landscape of Taita Hills (Saana Järvinen, January 2024).

1.2. Mango trees

Mango, *Mangifera indica*, is an exotic tree to Kenya, yet it is found in almost all farms practicing agroforestry across the country (McMullin et al., 2019). The trees are originally native from an area bordering Myanmar, Bangladesh and India, and their largest numbers today are growing in the Malay peninsula (Litz, 2009). Because of its origins, the mango tree has adapted best to the tropical monsoon climate, with a clear dry season followed by rains (Litz, 2009). Mangoes thrive in climates with low relative humidity and little to no rainfall during flowering, fruit setting and harvesting (Makhmale, Bhutada, Yadav, & Yadav, 2016). Nowadays mango is cultivated all around the tropics in over 103 counties (Jahurul et al., 2015), and it is the fifth most cultivated fruit in the world (Normand, Lauri, & Legave, 2015). The popularity may be caused by their ease of maintenance and longevity, as they may live up to a hundred years (Tharanathan, Yashoda, & Prabha, 2006). With a juvenile period of around 3 to 7 years, they are also relatively fast growing (Litz, 2009). A study conducted in Zimbabwe found that mango tree cultivation is independent of socioeconomic status (Musvoto & Campbell, 1995). Mangoes also provide great nutrition containing amino acids, carbohydrates, organic acids, protein and vitamins (Litz, 2009).

In Taita Hills mango trees are a prominent species in the agroforestry field. Mangoes are one of the most common fruit bearing tree species in Taita Hills, most popular being avocado (Piiroinen, Heiskanen, Maeda, Viinikka, & Pellikka, 2017). Mango trees also bring shade for

both land cultivating and residencies, and in the Taita Hills area they are often placed near buildings (Piiroinen, Heiskanen, Mõttus, & Pellikka, 2015). The trees are mostly planted and cultivated by farmers, especially in the lowlands of Mwatate (Abera et al., 2022). These specific areas are known to be called “mango valleys” by the locals. This study, however, concentrates on the mango trees in the areas of higher elevation. A small mango tree in a Taita Hills farm is portrayed in Figure 3.

Figure 3. A small mango tree midst vegetation in the highlands of Taita Hills (Saana Järvinen, 26th January 2024).

1.2.1. Mango trees and climate change

Sudden or unusual climatic changes can influence mango trees in various ways. Although the mango tree thrives in warm to hot temperatures and is fairly adapted to dry environments and water stress (Normand et al., 2015), extreme events of heat, drought or evaporative demand may damage its production capacity (Makhmale et al., 2016). A mature mango tree can prevail temperature extremes from 0°C to 48°C (Makhmale et al., 2016) and survive without irrigation for over eight months, but increasing temperatures may alter floral induction (Normand et al., 2015). The tree may grow without a relevant pause for the flowering to start if the temperatures,

rainfall and humidity levels are too high, all while changes in rainfall patterns may lead to low pollinator activities, poor fruit set and low quality in ripe mango fruits (Makhmale et al., 2016). While the drying climate can increase the fruit quality and temperature rise in cooler areas might be positive, both might provoke damage to photosynthesis (Normand et al., 2015). Mango trees are also facing threats of pests, which may be encouraged by unseasonal rains linked with climate change (Makhmale et al., 2016)(Dianda et al., 2023).

However, the tree needs environmental triggers to flower. Mango flowering is controlled by water stress in the tropic and temperature change in subtropics (Cho, Yoon, & An, 2017; Normand et al., 2015; Prates, Züge, Leonel, Souza, & Ávila, 2021). Mangoes flower during warm weather after a period of drought (Singh, 1960) in climatic conditions as in Taita Hills lowlands. Mango seed dispersal is primarily controlled by humans. However, similar to other fruit trees, mango seeds can also be dispersed by frugivores, after which the tree may survive in the wild on its own (Schupp, 1993), given that pollination is successful. In Taita Hills, the distribution of mango trees is principally reliant on cultivators, but the success of growth, flowering and fruiting depends on the climate.

Changing climate conditions are often considered as a threat to agriculture, yet for some areas the swift can be seen as positive progress. Rising temperatures since 1998 (Mjomba, 2023) are altering the region's ecosystem dynamics, which causes that the growth environments in the highlands are beginning to resemble the same conditions as in the lowlands. This creates new ecological niches. Due to these changes in the climate, new growth environments are appearing likewise for mango trees in Taita Hills. As a result, the mango trees seem to be succeeding and producing flowers in areas that they weren't before, while they are still growing well in their old habitats. The climate conditions in the highlands were not suited for mangoes before, but now it seems that they are beginning to grow more widely and to produce fruit also in areas they were not before (Mjomba, 2023).

1.3. Remote sensing in tree mapping

"Remote sensing is an advanced technology used for obtaining information about a target through the analysis of data acquired by a sensor from the target at a distance." (Pu, 2017)

Remote sensing has been proven to be an efficient method for tree mapping purposes. Remote sensing methods can be applied for multiple different causes ranging everywhere from individual tree detection in urban settings using tree inventory (Liu, Coops, Aven, & Pang, 2017), (Wallace et al., 2021) to mangrove forest detection with Normalized Difference Vegetation Index (NDVI) and Normalized Difference Moisture Index (NDMI) (X. Zhang et al., 2017). Multi-class classification methods are ideal for monitoring many different ecological indices (Bourel & Segura, 2018) and they are especially common in tree species classification methods (Fassnacht et al., 2016), if the species of interest are many. For instance, multispectral data is adequate to plant health accession (Xiao & McPherson, 2005), as it provides enough spectral bands for NDVI estimations. However, for plant detection purposes more spectral bands and a narrowed bandwidth are needed.

1.3.1. Hyperspectral remote sensing

Hyperspectral remote sensing (HRS) can be used on a wide variety of surfaces and materials of different kinds of physical and chemical properties. Hyperspectral imaging has the ability to detect more wavelength areas in electromagnetic spectrums compared to, for example, a film camera or digital camera with only 3-4 wavelength ranges, while also having narrower bandwidths to distinguish more nuances in the reflectance (Pu, 2017). In short, hyperspectral imaging has better spectral resolution. As a technique that employs passive sensors to capture reflected solar radiation from vegetation, HRS plays a crucial role in remote plant detection (Pu, 2017). Each plant and tree species reflects light slightly differently, and the reflectance can be studied with a signature curve for the species of interest by the different wavelengths it reflects (Roy, 1989). These unique spectral reflectance profiles can be used to detect species with high accuracy.

Hyperspectral data has been frequently used for vegetation mapping in previous research. HRS has received good results in for example Leaf Area Index (LAI) detection in shrub thicket canopies (Brantley, Zinnert, & Young, 2011), as well as in tropical forests with the help of reference data (Ferreira, Zortea, Zanotta, Shimabukuro, & De Souza Filho, 2016). Hyperspectral imagery is often used together with Canopy Height Model (CHM) based on Light Detection And Ranging (LiDAR) for tree mapping and forest structure studies (Alonzo, McFadden, Nowak, & Roberts, 2016; Dian, Pang, Dong, & Li, 2016; C. Zhang & Qiu, 2012). LAI and carbon storage are often simultaneously calculated from the LiDAR data. Airborne

HRS data has received accurate results for invasive tree species mapping in Taita Hills (Piiroinen, 2018).

1.3.2. Remote sensing of mango trees

Mango trees in Taita Hills have been mapped in low-quantity on a small site with the help of hyperspectral data as a part of a crop research, which showed that the eight studied species could be differentiated from each other with high overall accuracy (Piiroinen et al., 2015). Invasive tree species *Eucalyptus spp*. and *Acacia mearnsii* have been previously mapped with one-class classification method in Taita Hills (Piiroinen et al., 2018), (Zhao et al., 2022). Mangoes have been studied in orchard level with ground based hyperspectral imagery for yield (Gutiérrez, Wendel, & Underwood, 2019), (Sarron, Malézieux, Sané, & Faye, 2018) and maturity estimations (Wendel, Underwood, & Walsh, 2018). However, mango trees have not been mapped with airborne hyperspectral data on a wider scope, nor has the spreading of mango trees been studied before in Taita Hills.

1.4. Classification methods

The function of the classification paradigm is to separate an unfamiliar data object into predefined categories. The simplest example of this is a binary classification, where the data is divided into two different classes. However, in cases where there is only one target of interest, the classification of other objects in the dataset is indifferent. This is called One-Class Classification (OCC). In OCC, the target class or so-called positive class is well known and appropriately characterized in the training data, while the negative classes have few to no samples of them (Tax & Duin, 2001). OCC algorithms can be divided into three categories based on the availability of training data:

1. Learning with positive samples only. The main idea is to model the distribution of the positive class, identifying other objects as anomalies (Tax, 1999).
2. Learning with positive examples and some amount of poorly sampled negative examples or artificially generated outliers (Tax & Duin, 2001).
3. Learning with positive and unlabeled data, or positive-unlabeled (PU)-learning. Unlike binary classification, this method does not require a strictly defined negative class (Nigam, Mccallum, Thrun, & Mitchell, 2000).

The unlabeled data in the third category refers to samples that are randomly selected from all tree pixels. In contrast to positive-negative (PN) learning, in PU-learning the negative class is estimated by the marginal distribution and consists of the combination of positive class and unlabeled data. The PU-method is more efficient, as the model can be trained with additional data to enhance the differences between positive and unlabeled data (Khan & Madden, 2014).

The classification model used in this study is a PU-learning based OCC method called ITreeDet. The framework uses hyperspectral imagery together with a LiDAR-derived tree mask as input data source, which is then processed through a Convolutional Neural Network (CNN) to extract robust spectral-spatial features. The training is performed via a case-control setup, where labeled positive samples are independently drawn from the true positive distribution, while unlabeled samples are randomly drawn from the overall distribution of tree species. The framework utilized a loss function called absNegative to guide parameter updates. The output produces a pixel-level classification of species' presence or absence, which can be used to map the distribution of the target of interest in the study area.

1.5. Objectives

This thesis investigates the effect of climate change on the mango tree distribution in the Taita Hills area. The main focus of this study is on detecting mango trees in Taita Hills, as well as studying the possible spreading over the years. It also analyses the change in growing patterns during the decade between two datasets from 2013 and 2022. With up to 8 years of reaching maturity (Litz, 2009), the growth and emergence of new mango trees can be detected within this time period. The study concentrates on the areas particularly in the Taita Hills highlands due to the change happening in the growing habitats caused by the warming climatic conditions. Hyperspectral airborne remote sensing data is utilized together with LiDAR-derived tree mask to map the distribution and its change in response to climate change in Taita Hills. In addition, the thesis aims to estimate the aboveground biomass and carbon contents of the mango trees for 2022 data as well as to study the tree canopy's cooling effect on its surroundings.

The research questions for this thesis are the following:

1. How has the distribution of mango trees changed between the years 2013 and 2022 in Taita Hills? Are there more trees in the highlands than before?
2. How do mango trees contribute to climate change mitigation regarding carbon storing and cooling their surroundings?
3. Is one-class classification feasible for mango tree mapping and change detection using hyperspectral and LiDAR data?

2. Material

2.1. Study area: Taita Hills

The research was conducted on a 10 x 10 km area in Taita Hills (03°20'S, 38°15'E) of Taita-Taveta county, Kenya (Figure 4). The elevation of the study area varies from 1100 to over 2200 meters above sea level (m a.s.l.), its highest peak being in Vuria (Petri Pellikka, Ylhäisi, & Clark, 2004). The mountainous terrain of Taita Hills is surrounded by the lowlands of dry savannas and shrublands. The area has a bimodal rainfall pattern with long rains from March to May and short rains towards the end of the year. As a result of the orographic rainfall, where an elevation change forces the cloud to rise, cool and precipitate, a clear rain-shadow area receiving less precipitation can be identified in the northwestern slopes (P. K. E. Pellikka et al., 2013). This affects again the aboveground biomass patterns of the area, that are otherwise concentrated on slightly higher elevations (Adhikari et al., 2017). Taita Hills belongs to the Eastern Arc Mountains chain, which is one of the world's biodiversity hotspots containing various endemic species (Burgess et al., 2007). Some of them are even named after Taita, such as the African violet flower (*Streptocarpus teitensis*) (Martins, 2008) and the critically endangered bird species Taita Thrush (*Turdus helleri)* and Taita White-Eye (*Zosterops [poliogaster] silvan*) (Brooks et al., 1998).

Taita Hills is widely characterized by small-scale farming. Livelihoods of the inhabitants are highly connected to food production. The most important crops for agriculture are maize and bean crops (Castalonge, 2008). Thanks to two rainy seasons and efficient water storing facilities, the Taita inhabitants have a favorable foundation for agriculture compared to those in the nearby lowlands (Soini, 2006). Consequently, a significant portion of the highlands has been converted for agricultural use. However, population growth has reduced available farmland, which means smaller yield for the household to consume (Soini, 2006). As follows, the farmers either discover new means of livelihood, such as beekeeping (Newman, Marchant, Enns, & Capitani, 2021) or expand the agricultural land conversion to larger areas. Thus new challenges are introduced, including growing human-wildlife conflicts in the region (Di Minin, Slotow, Fink, Bauer, & Packer, 2021).

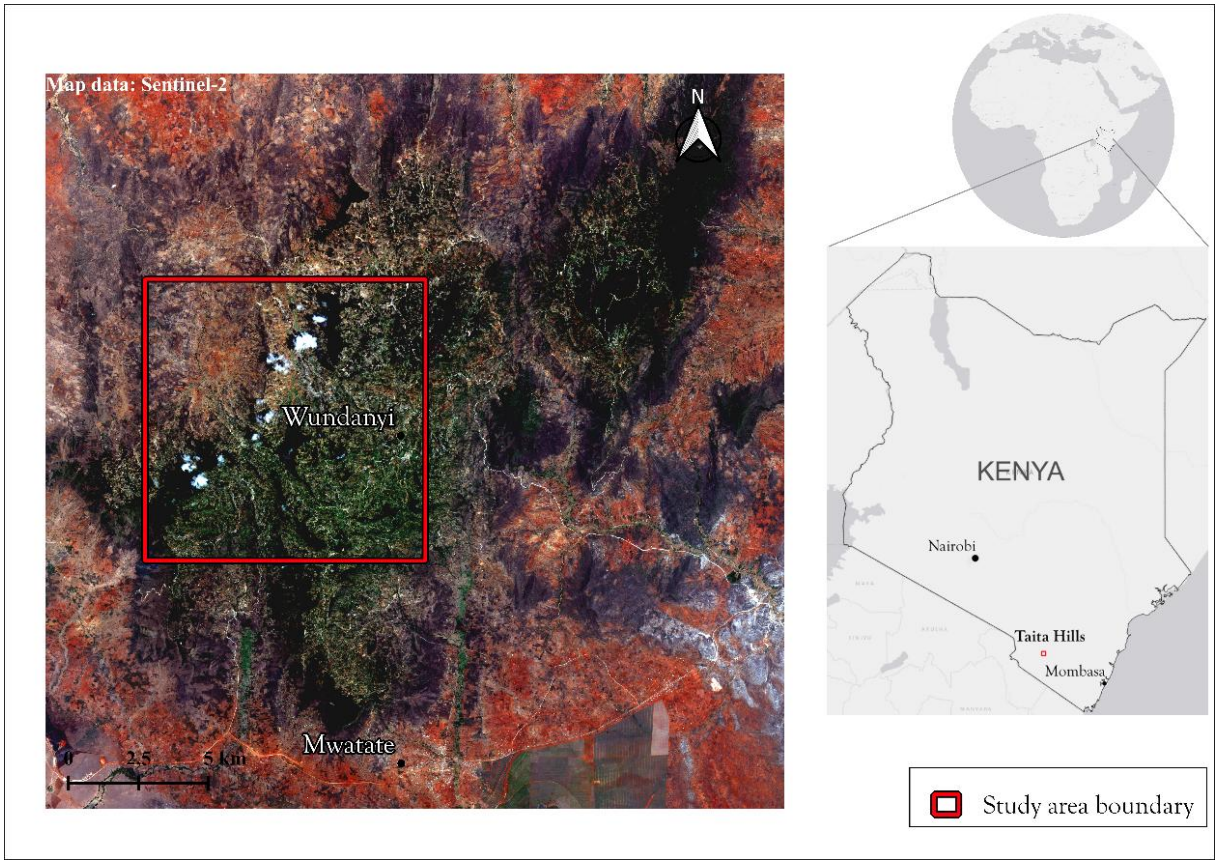

Figure 4. Map of the study area of 10x10 km in Taita Hills.

2.2. Fieldwork

To detect the trees from hyperspectral imagery, geographical coordinates of mango trees as well as trees other than mangoes were measured. The fieldwork was carried out at the end of January 2024 in the span of six days. A total of 220 tree locations were positioned with Trimble GeoExplorer 7 resulting in 110 mango tree locations and 110 other tree locations. Additional tree location data, which was collected earlier during fieldwork in 2013 for another thesis (Piiroinen, 2014) and in 2022 for this analysis, was also used. The non-mango tree datasets consist principally of avocado, but also various other tree species, such as silver oak, *Macadamia tetraphylla*, *Liquidambar styraciflua* and *Croton megalocarpus*. In addition, measurements of tree height, canopy width and diameter of the trunk were measured from all the mango trees for aboveground biomass and carbon assessment. The tree height was measured with a handheld laser range finder, the trunk diameter with a Diameter at Breast height (DBH) measuring tape and the canopy width with a centimeter measuring tape. The trees were found with help from local research assistants and farmers. The field data from 2022

and 2024 were eventually merged, as both were located based on the same hyperspectral data collected in 2022. This merged data is hereinafter referred to as 2022 data.

The fieldwork also included a collection of temperature data to determine the cooling effect of the mango tree canopy compared to its surroundings. Three pairs of Temperature-Moisture-Sensors (Wild et al., 2019) called “Tomst”s were used. One pair of Tomst’s was always set to measure one mango tree, in a way that one thermologger was placed in an open area near the tree, and the other right under the tree in shadow. Tomst’s were placed to measure five different trees in four different sites. One pair stayed in the same position for five days, while for two pairs the location was changed every day. In total there is data of one five-day period from one site, and four sets of 24-hour measurement periods from different sites. Only the ground-level temperatures measured by Tomst’s were used. The temperature measurement sites can be seen from the map in Figure 5.

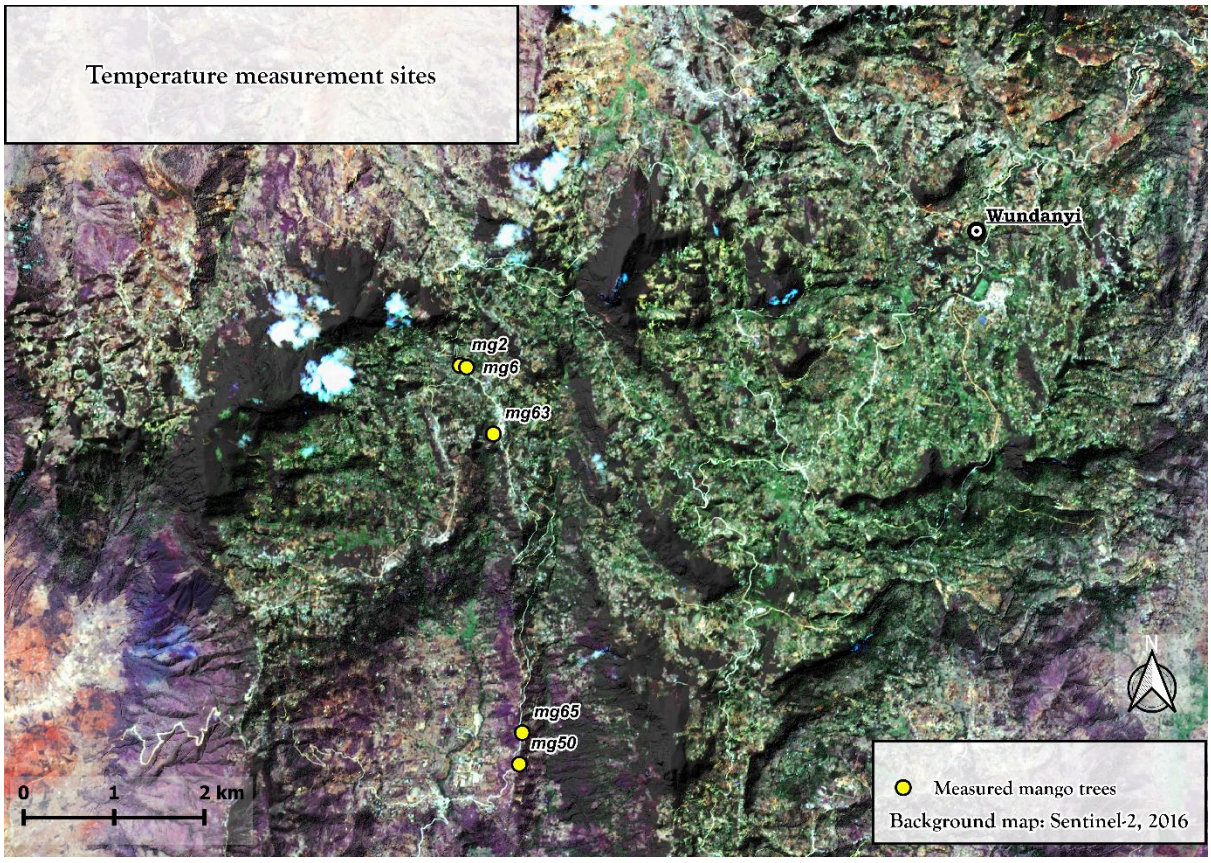

Figure 5. Temperature measured mango tree sites on map. Mango trees 50 and 65 are outside the 10x10 km study area.

Spectral measurements were taken with an SVC HR-1024i field spectroradiometer from two different mango trees and one avocado tree. The measurements with a spectrometer included individual leaves on the sunny side as well on the shadowed side of the mango trees due to bidirectional effects (P. Pellikka, 1998). Because of the equatorial location of the study area, the trees are exposed to sunlight during the day from every angle, but during the image acquisition leaves reflect light differently depending on the direction of incoming radiation. Leaf measurements were taken applying the spectrometer's internal light. The purpose of this was to visualize how the sunlit and shaded mango leaves' spectral reflectance profiles differ on sunny and shadowed side as well as in comparison with avocado leaves, which proposes a high chance of misclassification with mango. The spectral measurements are presented in Figure 6. to show the similarity of mango and avocado leaves in reflectance.

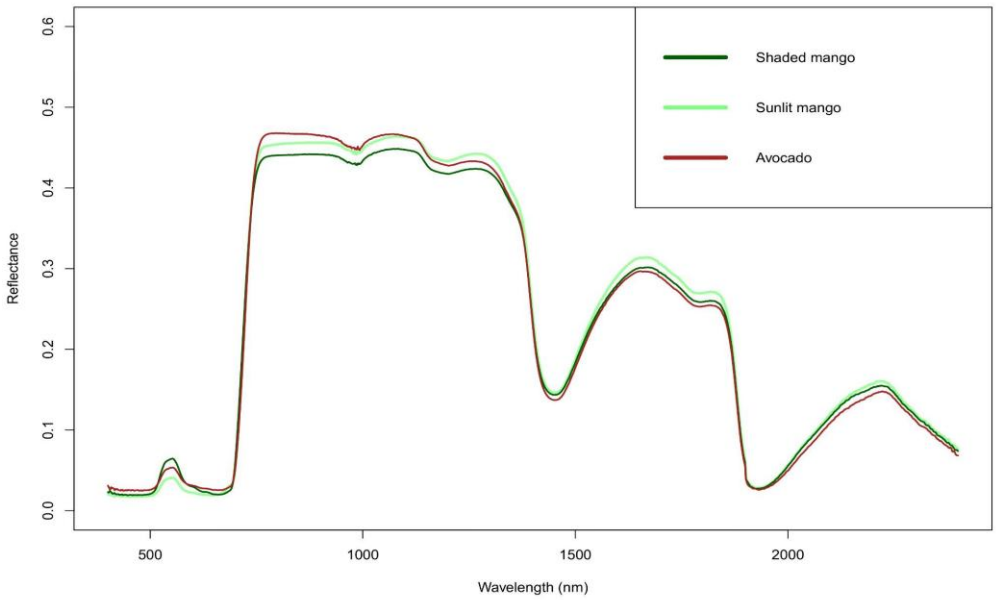

Figure 6. Spectral profiles of young and mature mango leaves. To show the similarity, the graph also includes a spectral profile of an avocado leaf. Green light is reflected at 500–600 nm.

In addition to the data collection, 20 interviews of small-holder farmers cultivating mangoes were carried out during the fieldwork. The questionnaire contained background questions, such as the size of the household and land and the amount of mango trees in the property. A questionnaire was formulated in advance and the responses were collected in a door-to-door method. There were for example questions related to the usage of mango trees: what is it used

for other than the fruits and what benefits and downsides does the tree bring. These interviews were collected from around the research area, each from a different site.

2.3. Data

Airborne hyperspectral imaging data in 2013 and 2022 was used as the material of this research (Table 1). LiDAR data was acquired simultaneously in both years' flight campaigns. A correspondent CHM produced from the LiDAR datasets was used to remove non-tree pixels from the hyperspectral image for both years. A tree mask was further created by combing the CHM, near-infrared bands and NDVI. Hyperspectral data is adequate at detecting tree species, and the hyperspectral datasets have excellent spatial and spectral resolution in both years. Even though the technical specifications of the sensor applied as well as the imaging characteristics of the data differ slightly, the imagery both covers the extent of the study area and complement each other enough to be comparable.

2.3.1. Hyperspectral and LiDAR data for 2013

Aisa EAGLE hyperspectral imagery was used for detecting the positioned trees in the field for the classification model of 2013 (Figure 7). The airborne data was acquired in February of 2013 at 750 m flying height above ground. The hyperspectral data has 129 bands from wavelengths of 400-1000 nm with 4,5–5,0 nm (full width at half maximum) intervals. The spatial resolution is approximately 1 m. The Aisa EAGLE sensor has a field of view of 36,04°. The hyperspectral data was radiometrically corrected and orthorectified with CaliGeoPro 2.2 (Spectral Imaging Ltd. Oulu, Finland), and atmospheric correction was done using ATCOR-4 (ReSe Applications Schläpfer, Wil, Switzerland). The Optech ALTM 3100 LiDAR laser scanner applied simultaneously acquired data with mean pulse density of 9,6 pulses m^{-2} and mean return density of 11,4 returns m^{-2} (P.K.E. Pellikka et al., 2018). The LiDAR data was preprocessed by the data supplier (Topscan Gmbh, Rheine, Germany), and a CHM model was derived from the LiDAR data with TerraScan software (Terrasolid Ltd., Helsinki, Finland).
(Piiroinen et al., 2018)

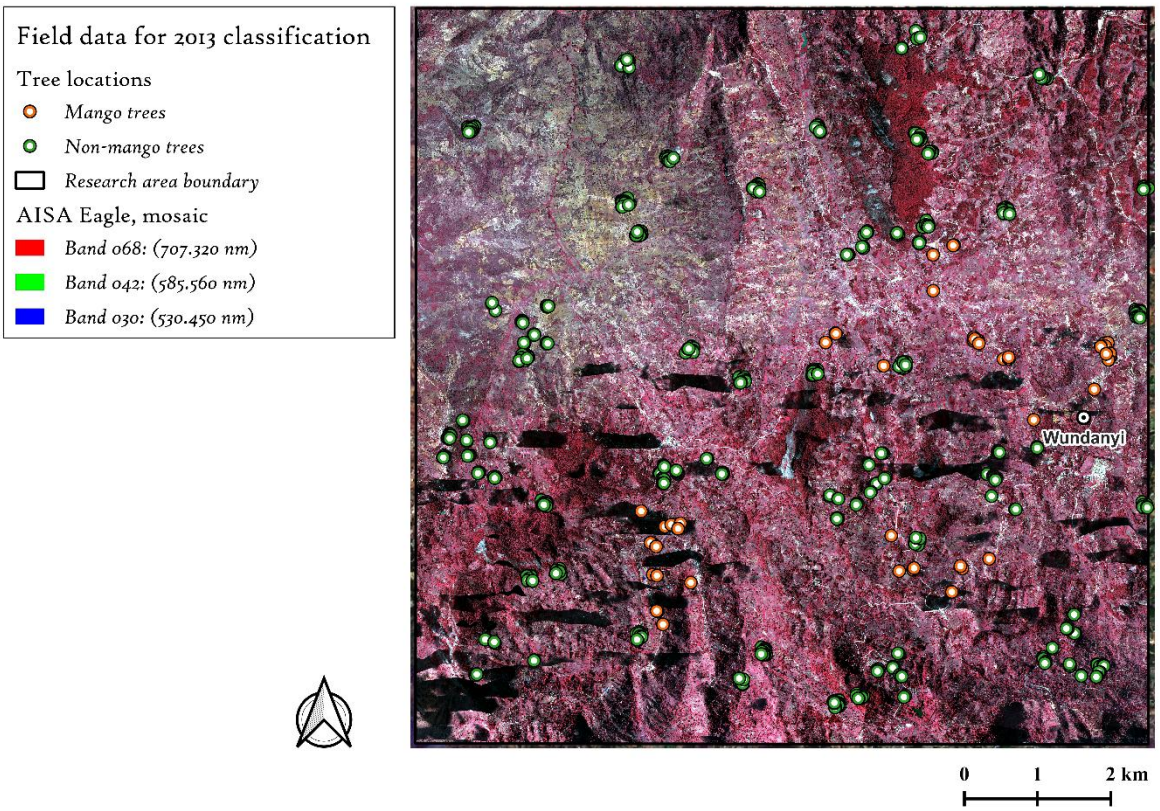

Figure 7. Hyperspectral imagery with measured tree locations for 2013 classification.

2.3.2. Hyperspectral and LiDAR data for 2022

Aisa KESTREL10 hyperspectral imagery was used for the classification model of 2022 (Figure 8.). The hyperspectral data was captured using Aisa KESTREL 10 push-broom imager in March 2022. The data has a spatial resolution of approximately 0,7 m with a flying height of 900 m above ground and a 39° horizontal field of view. 92 spectral bands were used ranging from 381 to 1003 nm, each with a spectral resolution of 6,7–6,9 nm (full width at half maximum). Radiometric calibration and orthorectification for the 2022 data were done by CaliGeoPro 2.2. DROACOR software version 2.0 (ReSe Applications Schläpfer, Wil, Switzerland) was used for atmospheric correction and BRDF correction. Leica ALS60 sensor (Leica Microsystems, Wetzlar, Germany) acquired Airborne Lasor Scanning (ALS) data with point density of 0.5–5 points/m^2 simultaneously for the CHM creation. The preprocessed data received from the vendor (Ramani Geosystems Ltd, Nairobi, Kenya) was generated into a CHM model using LAStools software (rapidlasso, Gilching, Germany).

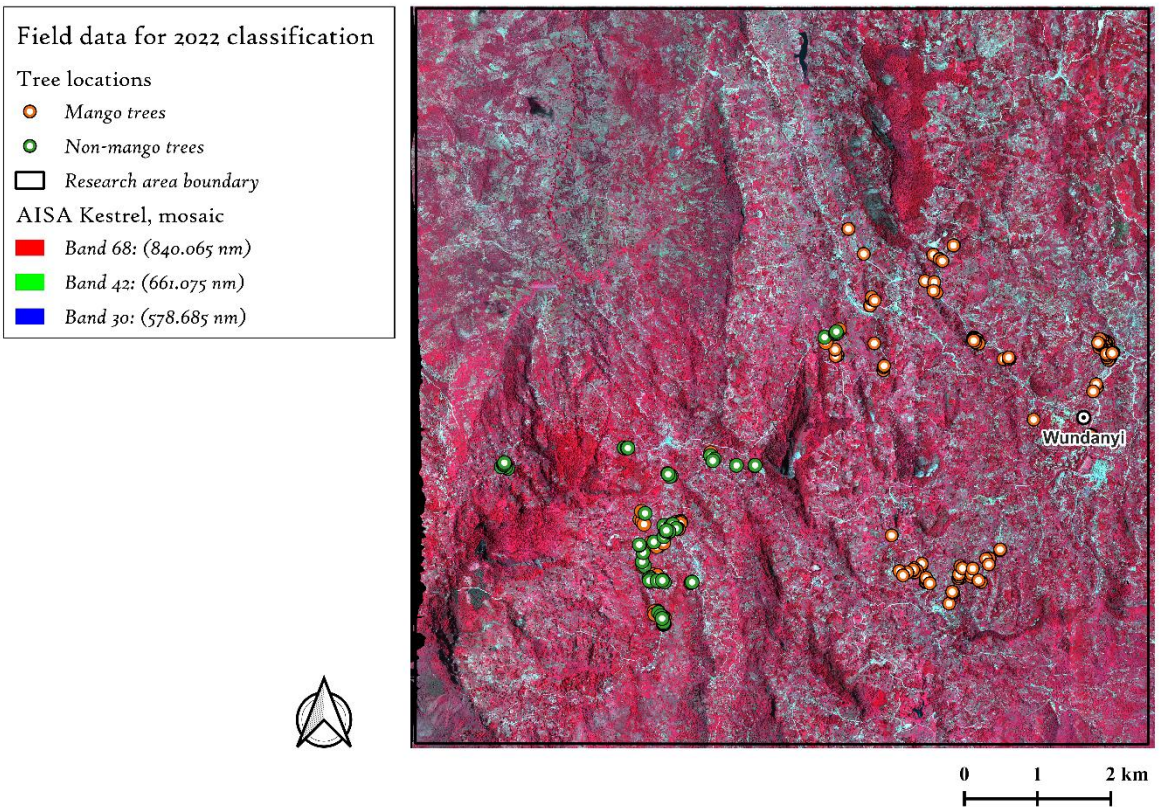

Figure 8. Hyperspectral imagery with measured tree locations for 2022 classification.

Table 1 Aisa EAGLE and KESTREL hyperspectral datasets.

****Might differ in a mountainous terrain***

	Year	Spectral bands	Wavelength range	Spatial resolution*	Spectral resolution	Field of view	Flying height above terrain*
EAGLE	2013	129	400-1000 nm	1 m	4,5-5,0 nm	36,04°	750 m
KESTREL	2022	92	381-1003 nm	0,7 m	6,7-6,9 nm	39°	900 m

3. Methods

This thesis will be adapting a framework developed by (Zhao et al., 2022), where the one-class classification received exceptional results for mapping eucalyptus and black wattle. The framework implements case-control PU learning to study high-resolution hyperspectral image data. Even though the model was initially created for invasive trees, it was also suitable for other species detection. This framework was convenient for several reasons: a) The framework utilized hyperspectral data as input, which is needed for reliable species detection. B) Instead of relying on manually defined features, deep learning was utilized to automatically extract high-level semantic representations, which greatly benefits classification. In addition, c) the model used a one-class classification method. One-class classifiers are well suited to focusing solely on the target species (Kilickaya, Ahishali, Sohrab, Ince, & Gabbouj, 2023). This is beneficial when the interest is in mapping only single class. The classification also implemented d) positive and unlabeled (PU) learning, along with an AbsNegative risk estimator, to integrate additional unlabeled samples and reduce overfitting during model training (Zhao, Zhong, Wang, & Shu, 2023). Overfitting means the too close resemblance of algorithm and training data, which results in inaccurate predictions when the model cannot draw any conclusions outside of the training data. Incorporating more data into the classification or refining the loss function to better reflect real-world conditions can help the model better generalize to new data.

The pipeline of this thesis (Figure 9) begins with the preprocessing of raw hyperspectral and LiDAR datasets as well as the field data for both 2013 and 2022. Positive and unlabeled datasets are created, and they are then assigned into training sets according to Table 2 for the classification model of ITreeDet. Negative dataset is also created, which is used for testing the model together with the positive dataset. Accuracy of the model is assessed by calculating several common indices. Postprocessing includes filtering the prediction results in order to better detect individual tree canopies. The final result is a prediction of the mango tree presence for both years. Calculating the difference produces a change detection map for the whole research area. In addition, aboveground biomass and carbon content are calculated based on the mango tree measurements taken from the field.

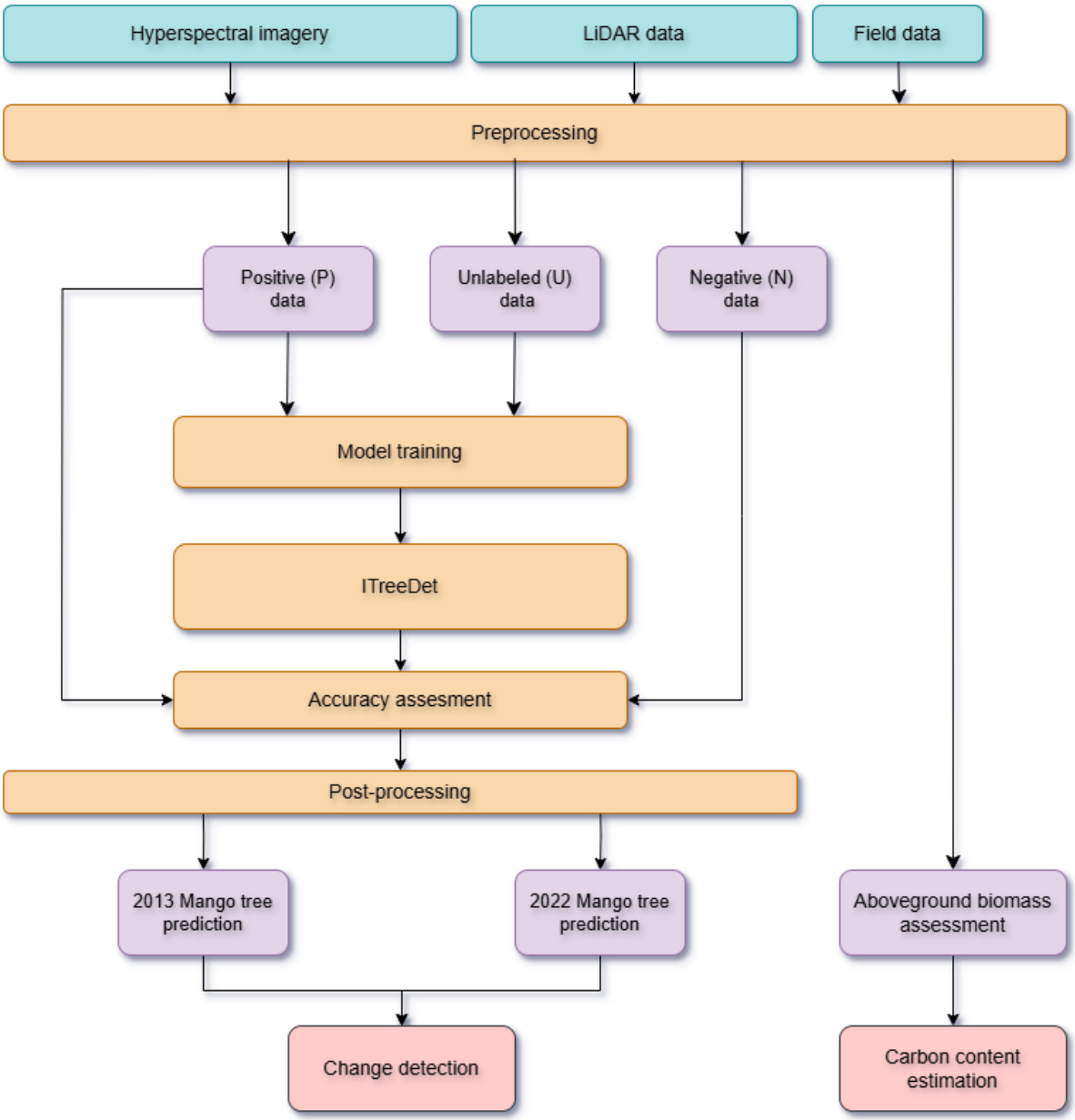

Figure 9. Flowchart of the classification process.

3.1. Preprocessing

3.1.1. Field data preprocessing

During the measurements of the tree locations, a Trimble base station recorded data of its own fixed position located at an open space at the Taita Research Station while the tree locationing was done. This static data was then utilized in Differential Global Navigation Satellite System (DGNSS) correction, a method that enhances positional accuracy by minimizing the errors common to both the base and rover receivers. When combining both the static and rover-

collected data the spatial resolution could be improved considerably. By the end of the fieldwork, all the points gathered were successfully corrected in Trimble Pathfinder Office.

The first step was to create polygons to cover the apparent canopy area of the trees. The trees were identified visually in ArcMap with the help of high-resolution Red, Green and Blue (RGB) images and the tree location points collected in the fieldwork. A polygon was drawn around the apparent canopy area of the identified trees, and each species was then addressed to its own class for differentiation in the classification stage. The above-mentioned progress is called labeling. For the hyperspectral data of 2013, 53 mango trees were labeled. In addition, from the 685 labeled non-mango trees (Piiroinen, 2014), 200 were used. For the 2022 data, a total of 114 mango trees, and 32 negative trees were labeled (Table 2). Negative trees (i.e. trees that are not mangoes) consisted mainly of avocados, liquidambars, crotons and silver oaks, which are typical tree species growing on farms. The overlapping of the tree canopies in the forests caused problems for the identification and therefore the measurements were carried out in open farm areas. Some trees were excluded to ensure that the positive samples are not false positives, i.e. trees incorrectly mapped as mangoes.

3.1.2. Dimension reduction of the HRS data and tree mask generation

The original hyperspectral raster data files were cropped to match the research area extent of 10 x 10 km. Minimum Noise Fraction (MNF) was done as a preprocessing step for both datasets prior to classification. MNF means the reduction of data dimensionality in order to preserve the most relevant information and to enhance data quality (Luo, Chen, Tian, Qin, & Qian, 2016). The MNF for Aisa Eagle (2013) and Kestrel (2022) data was created in ENVI image processing software with the Spectral Hourglass Wizard tool. The resulting MNF-files only consisted of 15 bands needed for the analysis, which helped to reduce the redundant data, as well as to manage the noise in raw hyperspectral data. Tree mask was created in RStudio using the cropped hyperspectral image and a canopy height model created from the LiDAR datasets collected from 2013 and 2022. For both years the mask was calculated with parameters, where reflectance (836 nm) < 20%, NDVI < 0,5 and CHM > 1,5 were combined to remove non-tree pixels. Thus, the model could only focus on the canopy pixels and not confusing them with ground pixels. Patches of 11x11 pixels were created from the tree-masked MNF-file and assigned randomly into training and testing sets 70% and 30% respectively (Table 2). The prior probability for positive class was calculated with the collected ground truth points and the

masked hyperspectral data by gradient thresholding algorithm (Ramaswamy, Scott, & Tewari, 2016), to help the model define how likely a pixel belongs to the target class.

Table 2. Composition for the classification model training and testing sets.

Phase	**Patches**	**2013**	**2022**
Training	Positive pixels	774	3 008
	Positive tree crowns	40	78
	Unlabeled	20 000	20 000
Testing	Positive pixels	393	1 246
	Positive tree crowns	18	35
	Negative pixels	3 056	1 439
	Negative tree crowns	205	32

3.2. One-class classification

The framework used for the classification model in this thesis is called ITreeDet. It was originally created to tackle the challenges of spectral variations between different trees and within the same tree species often occurring with OCC. The ITreeDet framework is composed of several integral components. The first stage is data preprocessing, where data is prepared for hierarchical classification. ITreeDet uses Case-Control PU learning, meaning that there are "cases" of positive samples obtained from tree-level measurements (mangoes) and "controls" of unlabeled samples that are randomly selected from all tree pixels. The ITreeDet framework utilizes a shallow 2D CNN for spectral-spatial feature extraction, where patches are defined as pixels and their surrounding neighborhoods (11x11 pixels) for spatial context (Table 3). The deep spectral-spatial extractor for the tree species is designed as a shallow CNN - meaning it has fewer layers - because of the limited amount of training samples. The main layers are a) convolutional layers with 3x3 kernels and multiple filters, b) batch normalization and rectified

linear unit used for activation, c) average pooling to reduce parameters and d) fully connected layer as the final classifier. Finally, an AbsNegative risk estimator is used along with the ITreeDet-model, which allows the error accumulation to be decreased.
(Zhao et al., 2022)

Table 3. The configuration details of the standard deep spectral-spatial feature extraction CNN in ITreeDet. (Zhao et al., 2022)

Stem	**Block#1**	**Block#2**	**Block#3**
3 x 3 conv, 64	3 x 3 conv, 64	3 x 3 conv, 128	3 x 3 conv, 192
	3 x 3 conv, 128, stride 2	3 x 3 conv, 192, stride 2	

To train the classification model using only positive and unlabeled samples, the PU-learning framework relies on a specific risk estimation strategy, which is described as follows. Let x denote the input tree species patches and $y \in R$ denote the mango class or negative tree species class. The probability marginal distribution of the input features x is $P_P(x) = P(x|y = 1)$ for mango tree species and $P_N(x) = P(x|y = -1)$ for the negative tree species. Let $f : \chi \to \gamma$ be the decision function and $l : \chi \times \gamma \to R$ is a loss function. Thus, the CNN can be optimized using the traditional risk estimator:

$$\hat{R}_{PN}(f) = \pi_P \hat{R}_P^+(f) + (1 - \pi_P)\hat{R}_N^-(f) \quad (1)$$

where $\hat{R}_P^+(f) = E_P[l(f(x), +1)]$ and $\hat{R}_N^-(f) = E_N[l(f(x), -1)]$. $\hat{R}_P^+(f)$ and $\hat{R}_N^-(f)$ are the expectation of the loss of the positive tree species samples and negative tree species samples, respectively, and $\pi_P = P(y = +1)$ represents the class prior of the positive species. Given that the negative classes are not available, the marginal distribution of the negative species can be estimated by (Du, Niu, & Sugiyama, 2015):

$$P(x) = \pi_P P_P(x) + (1 - \pi_P)P_N(x) \quad (2)$$

It results in $\hat{R}_N^-(f) = \frac{\hat{R}_U^-(f) - \pi_P \hat{R}_P^-(f)}{1-\pi_P}$, where $\hat{R}_U^-(f) = EU[l(f(x), -1)]$, which is the expectation of loss when the unlabeled tree species samples are considered to be negative tree species, and $\hat{R}_P^-(f) = EP[l(f(x), -1)]$, which is the expectation of loss when the positive samples are considered to be negative (Zhao et al., 2022). Then the general estimator $\hat{R}_{PU}(f)$ can be estimated by:

$$\hat{R}_{PU}(f) = \pi_P \hat{R}_P^+(f) + \hat{R}_U^-(f) - \pi_P \hat{R}_P^-(f) \quad (3)$$

However, there is a possibility for the uPU risk estimator to overfit, as $\hat{R}_N^-(f)$ is not always greater than 0. To solve the overfitting problem, the absNegative risk estimator was created. The formula for the negative risk estimator can be calculated by:

$$\tilde{R}_{PU}(f) = \pi_P \hat{R}_P^+(f) + \left|\hat{R}_U^-(f) - \pi_P \hat{R}_P^-(f)\right| \quad (4)$$

3.3. Post-processing and change detection

The classification results for both 2013 and 2022 data were post-processed in order to reduce misclassification errors in the change detection analysis. All the areas representing mango trees with less than seven pixels or more than 200 pixels were filtered out. Most of the misclassifications or false positives in the data were dispersed in pixel groups of under ten pixels. Value seven was chosen after visual inspection of the results, because no seven-pixel area had covered the size of a whole tree canopy. Some of the mango tree predictions covering only a part of a tree canopy were unfortunately also under that limit. The limit for the large pixel groups was set, because the mangoes in Taita Hills mainly grow as individual cultivated trees in the field, and not tightly next to each other. Thus, if a group of mango trees were detected they were most likely other tree species. For the purpose of cleaning the results, the large pixel groups were also filtered.

After the filtration process, the results were further processed into a more compact form, and thus a density map was created. To enhance visualization, the research area was divided into 100x100-meter grid cells, and the number of individual tree canopies was counted within each cell. This was done by identifying and extracting connected regions (tree crowns) from the

classification raster file and projecting these trees to a specified 100m x 100m grid. In case of intersection, the tree was counted into the grid cell where it occupied more area (pixels). Although the method makes the results more robust, the higher spatial resolution was chosen to improve clarity in presenting the results, considering the size of the research area. Finally, the results from both years were compared to create a change detection map that calculates the difference between the 2013 and 2022 data. This was done by subtracting the 2013 values from the 2022 values. The resulting change detection map shows positive values as increase in mango trees and negative values as decrease.

3.4. Aboveground biomass retrieval

As a part of the climate change mitigation aspect of the mango trees, an aboveground biomass (AGB) and aboveground carbon storage (AGCS) estimations were calculated. The research article by (Kuyah et al., 2024) suggests a number of different formulas for calculating aboveground biomass for fruit trees with reliable and statistically significant results. The formula with the best accuracy considering the data available was chosen. In this case, the best allometric equation estimating AGB in mango was found combining mean DBH as primary predictor and tree height as the secondary predictor. The equation was number 19 from the list in the article, and it performed well with RMSE of 51,88, Adjuster R^2 of 0,96 t-value of 18,47 and p-value of 0,001. (Kuyah et al., 2024)

The formula for AGB estimation is as follows:

$$AGB = 0{,}0845 \times \mu DBH^{1{,}6995} \times Height^{1{,}0853} \quad (5)$$

where Height is the height of the highest point of the tree, and μDBH is the mean diameter of the trunk at breast height.

Prior calculations had to be made for the AGB estimations. The DBH of multi-stemmed mango trees was calculated as the sum of the total diameter of the largest trunk and one-half the diameter of each additional trunk (Magarik, Roman, & Henning, 2020). Aboveground carbon storage (AGCS) was calculated roughly as 45% of the aboveground biomass content, as it is also an accurate estimation for a mango tree (Naik, Sarkar, Das, Singh, & Bhatt, 2019). The field measurements were then calculated with the earlier-mentioned formula in

RStudio into AGB and AGCS estimations. The mean DBH of multi-stemmed trees can be calculated as:

$$\mu DBH = \sqrt{\sum_{i=1}^{n} DBH_i^2} \quad (6)$$

where DBH is the diameter of each individual stem at breast height, in a case where there are six stems in the measured tree, n meaning the number of branches. (Magarik et al., 2020)

3.5. Classification assessment

The accuracy of the classification model was assessed using five metrics. These metrics are calculated based on "ground truth" i.e., true labels and predicted labels. Precision, recall, and F1 score are metrics, that are computed at a specific decision threshold. By varying the threshold, the classification can be wired towards higher precision or higher recall trade-off. These curves help identify the threshold that best suits the application requirements

The precision curve shows how precision changes as the threshold varies and tells the classifier's ability to avoid labeling a negative sample as positive. Precision is calculated as a ratio of $\frac{tp}{(tp+fp)}$, where *tp* is the number of true positives and *fp* the number of false positives. Precision function receives values ranging from 0 to 1, highest meaning the best performance.

The recall curve shows how recall varies with threshold and tells the classifier's ability to find all the positive samples. It is calculated as a ratio of $\frac{tp}{(tp+fn)}$, where *fn* is the number of false negatives. Recall function receives values ranging from 0 to 1, the highest meaning the best performance.

F1 score is the harmonic mean of precision and recall. It combines both precision and recall showing the balance across thresholds, which is especially useful when the class distribution is imbalanced. F1 score is calculated as:

$$\mathrm{F1} = \frac{2 * \mathrm{tp}}{2 * \mathrm{tp} + \mathrm{fp} + \mathrm{fn}} \quad (7)$$

F1 score receives values ranging from 0 to 1, highest meaning the best performance.

The Precision–Recall (PR) curve shows precision against recall at various thresholds. It is particularly useful for imbalanced classification problems, when the positive class is rare. Average Precision (AP) is used to summarize the PR curve. The increase in recall is used to calculate the weighted mean of precision achieved. AP is calculated by:

$$\text{AP} = \sum_{n} (R_n - R_{n-1}) P_n \tag{8}$$

where P_n and R_n are the precision and recall at n^{th} threshold.

Receiver Operating Characteristic (ROC) curve is created by plotting the true positive rate against the false positive rate. Area Under ROC-curve (AUC) summarizes the ROC curve in one number. AUC tells the probability that the model will correctly rank a randomly chosen positive example higher than a randomly chosen negative example. AUC also receives values ranging from 0 to 1, a value of 1.0 representing perfect classification.

3.6. Implementation details

The processing steps were done with R in RStudio, Python in Google Colab -Jupyter notebook service and in the CSC (Center for Scientific Computing, Finland) Puhti interface. Puhti is one of the seven supercomputers run by CSC. ArcGIS was used for the initial tree labeling, and QGIS for testing purposes. Part of the codes were provided from previous research projects for the preprocessing and classification step of the analysis.

The one-class classification model was run in Puhti-supercomputer web interface. The model was trained with the previously created patches. The number of epochs to be trained was set to 110 epochs with 0,001 learning rate. The requested resources for the batch job were NVIDIA V100 GPU, 1 node per job and 1 task for job step, 10 CPUs per task and 50 G memory. The model showed best performance when the class prior values of 0,06 for 2013 and 0,15 for 2022 were used. The threshold to convert continuous predicted values into class labels was adjusted after several reruns. The values of 0,85 for 2013 and 0,95 for 2022 were used for the threshold.

4. Results

4.1. Mango tree classification

4.1.1. Model performance

The model performed well in detecting trees within their canopy areas. The positively classified pixels effectively corresponded to the visible canopy pixels, including both previously labeled trees and newly identified ones. This can be studied from Figure 10 and 11, where turquoise areas represent predictions for 2013, and pink areas correspond to predictions for 2022. The model seemed to effectively identify tree canopies, capturing at least a portion of each canopy, with many canopies being fully covered. Notably, the model largely avoided predicting pixels beyond the canopy edges (also visible in Figure 10), suggesting a high degree of spatial precision. Tree growth can also be detected from the results: canopies of the 2022 predictions are larger than those of the same trees of 2013 predictions. The emergence of new trees is also detected. Figure 11 shows a collage of different trees located in the fieldwork that are known to be mango trees. The majority of the located and labeled trees were successfully detected by the model.

Figure 10. The model detected the canopy areas well. The green points mark the located trees; the mango tree prediction for 2013 is marked in turquoise and 2022 in pink.

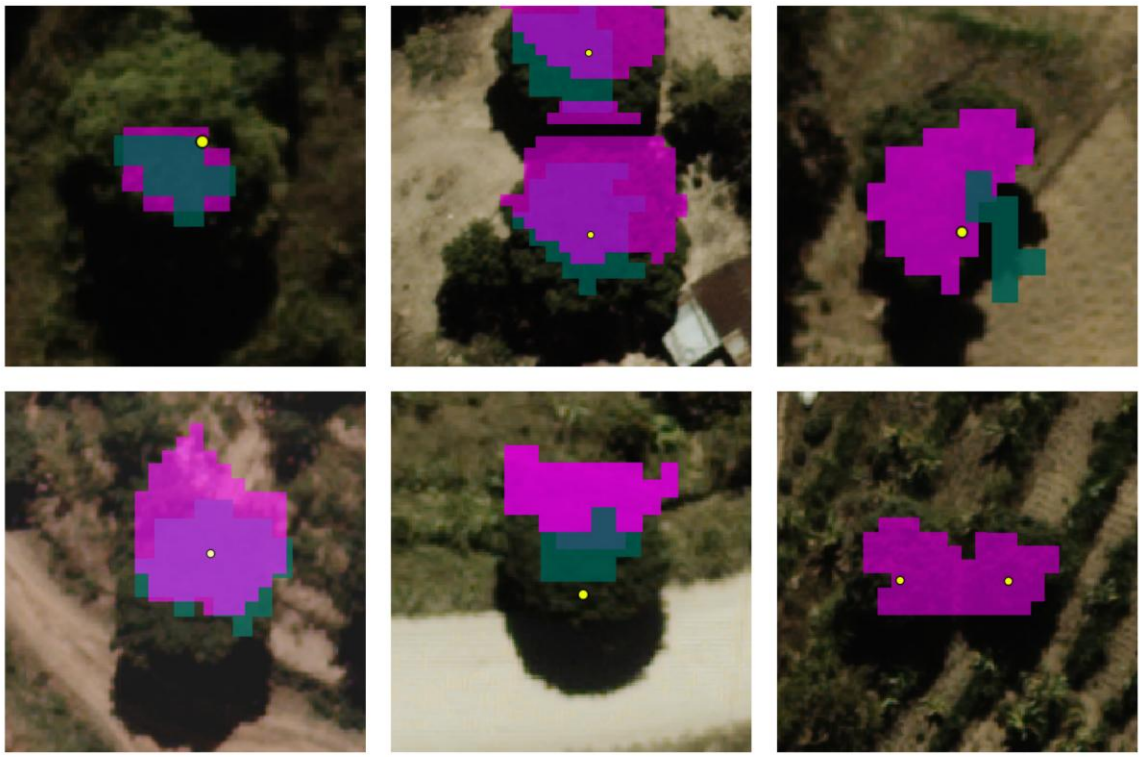

Figure 11. A collage of different field located trees, that are also classified positive by the model. The green points mark the located trees; the mango tree prediction for 2013 is marked in turquoise and 2022 in pink. The spatial resolution of the RGB-images is slightly different than in the hyperspectral images, hence the subtle displacement of the predictions from the trees.

Table 4. Model accuracy table at selected thresholds, as well as metrics of the highest F1 score for both years

	Threshold	Precision	Recall	F1 score	AUC	AP
2013	0,85	0,96	0,45	0,61	0,97	0,86
Optimal 2013	0,41	0,83	0,77	0,80	0,97	0,86
2022	0,95	0,99	0,65	0,78	0,95	0,95
Optimal 2022	0,86	0,88	0,88	0,88	0,95	0,95

The evaluation plots for the classifications of 2013 and 2022 are shown in Figures 12 and 13, respectively. The model on the data of 2013 shows a slightly unstable performance. When the recall is low, the precision is high. AUC instead reaches the best score of 0,97, and seems to perform well, together with the ROC curve (Table 4). The precision curve reaches values over 0,8 already by the threshold of 0,4. The F1 score is highest at the threshold of 0,41, which

would mean the best model predictive performance, yet at that threshold the precision curve does not reach its highest value. The threshold was finally set to 0,85 for the best performance of the model. In this threshold the precision is strong with a score of 0,96 while the F1 score is 0,61 and recall 0,45 (Table 4), meaning that at this point almost half of the trees have already been mapped, but the remaining trees are mapped with high accuracy and preferable precision-recall.

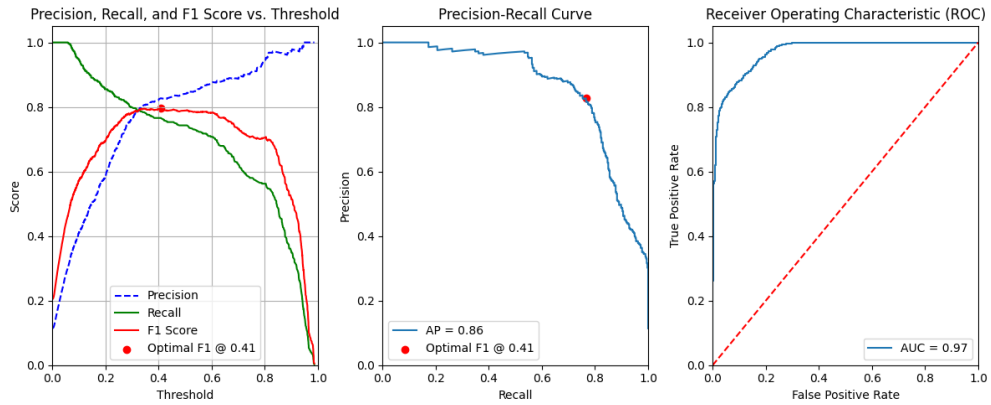

Figure 12. The evaluation plots for the classification model of 2013.

For the classification of 2022, the inference evaluation plots perform better. The precision-recall curve is more stable, and the precision stays high and only decreases after recall rises over 0,8. AUC reaches a value of 0,95, but the ROC curve stabilizes dilatory, which affects false positive rate of the model. The precision curve increases exponentially as the threshold increases. The recall curve is clean. The F1 score receives high values of over 0,6 and reaches a score of near 0,9 at around 0,85 threshold. The threshold for the classification was set to 0,95, where while the F1 score is around 0,6 so slightly lower, the classification model precision reaches remarkable values of 0,99. This threshold was decided in order to reduce the amount of misclassifications in the results.

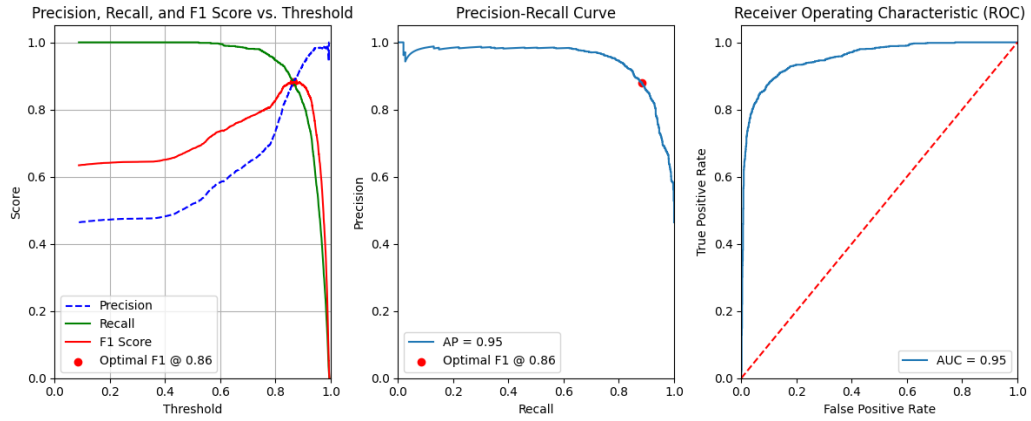

Figure 13. The evaluation plots for the classification model of 2022.

4.1.2. Mango tree mapping

The final results produced a raster classification map, where each pixel is classified as either presence or absence of mango tree canopy. The visual classification of the 100x100m grid cells is based on the number of individual tree canopies present in them (Figure 14). The prediction maps in Figures 15 and 16 depict the number of trees inside each 100x100m grid cell. The prediction results of each year were visualized classifying the values according to a green color scheme. Grid cells with only one predicted tree are displayed with 30% transparency for better readability. Cells with no predicted trees (value of 0) are not displayed. Contour lines were added for elevational context.

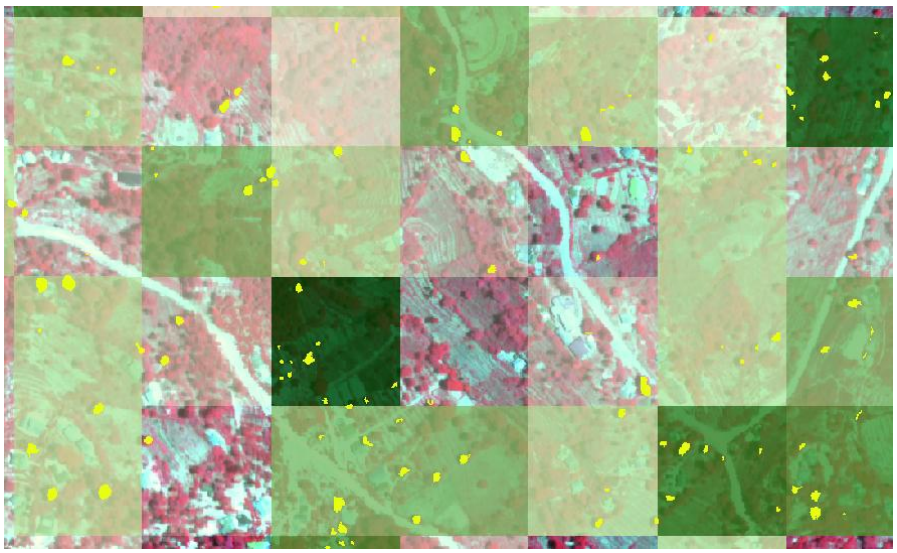

Figure 14. A zoomed-in screenshot of the 2022 prediction results on top of the grid cells. The positively classified trees are marked yellow. The grid cells are classified according to the number of trees present in them.

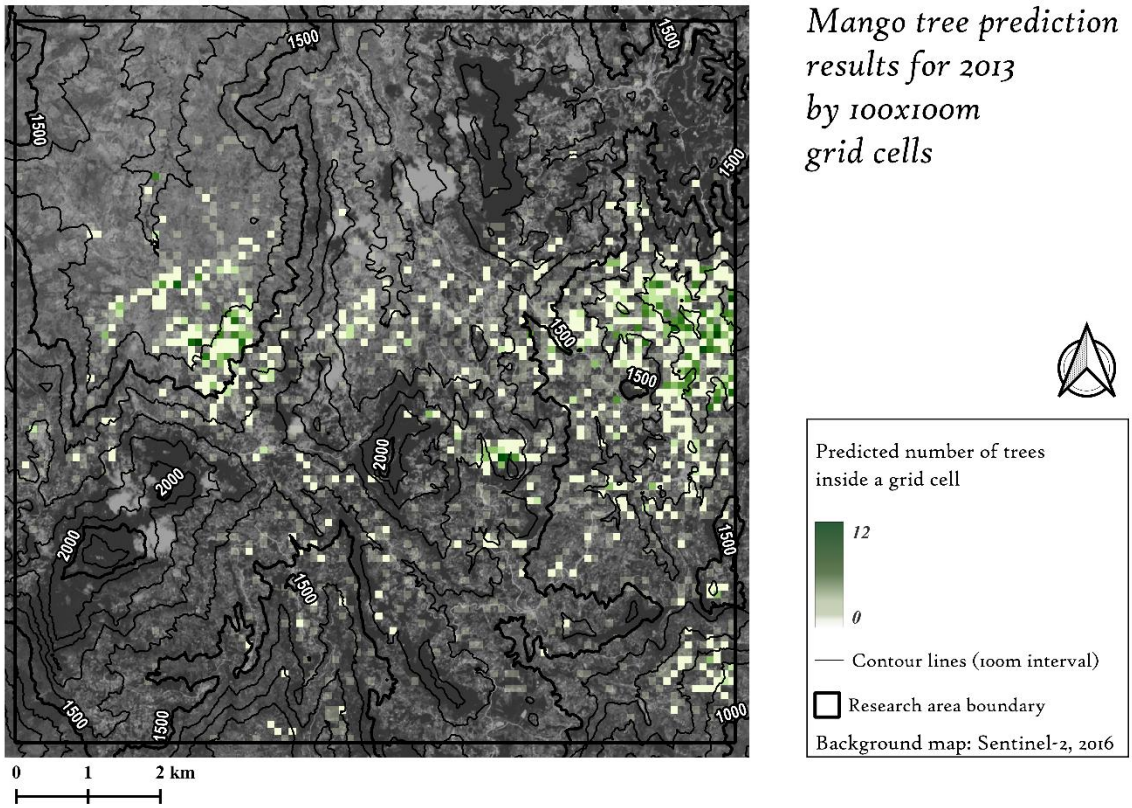

Figure 15. The predicted locations for mango trees in 2013. The predictions are shown in a green color scheme, displayed values ranging from 1 to 12. The map has contour lines with 100-meter intervals.

The classification results for 2013 are presented in Figure 15. The predicted positive trees are sparsely distributed across the study area, although the trees seem to be merely focused on the eastern side, as well as on elevations around 1300 to 1500 meters. The map reveals individual mango trees along with a few distinct clusters. Individual trees are more commonly found in the southern part, while the two largest clusters in the study area are located towards west from the central point, as well as towards east. Both of these clusters have elevations of around 1400 m a.s.l. Additionally, smaller but denser clusters are visible in the central region with one very distinctive cluster northeast of the center. The number of tree pixels per grid cell is higher in these clustered areas, suggesting a higher tree density within single farms. The effect of the rain-shadow area can also be seen from the map on the northwestern side, where there are little to no mango trees present. Very few predictions can be found in elevations above 1500 meters.

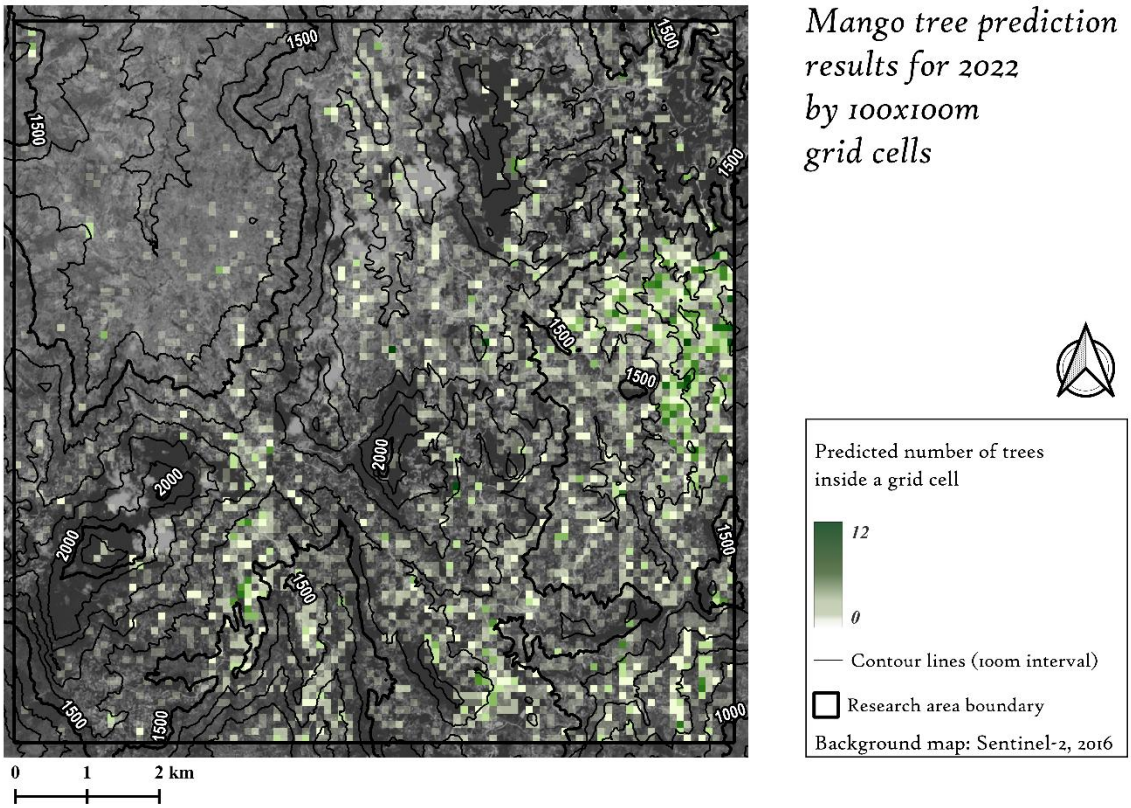

Figure 16. The predicted locations for mango trees in 2022. The predictions are shown in a green color scheme, displayed values ranging from 1 to 12. The map has contour lines with 100-meter intervals.

Figure 16 illustrates the predicted mango tree locations for 2022. The classification results show a more widespread presence of mango trees compared to 2013. The trees are more distributed around the research area, whereas 2013 predictions are more concentrated. The largest clusters appear recurrently around the elevation of 1400, but some predictions are present in areas as high as 2100 m a.s.l. Similar to the 2013 classification, trees remain clustered toward the eastern side of the study area. The clusters, however, cover a larger area. Individual trees are present throughout the area, and greater values are found outside the main clusters as well. Only the west side remains less represented, again likely because of the rain-shadow area, though some smaller predictions are still present.

4.1.3. Change detection

The mango tree prediction results from both years were compared to each other to create a change detection map, where the total number of mango trees in 2013 was subtracted from the number in 2022. The result of the change detection analysis is shown in Figure 17. Instead of counting the total number of pixels, the 100x100m grid cells were classified based on the amount of individual tree crowns within each cell. This approach highlights the presence of new trees, which is independent of natural canopy growth over the study period. It allows the focus to remain on significant changes in tree presence.

Similarly to the tree prediction maps, a green color scheme represents positive change, meaning an increase in mango trees within a grid cell. In contrast, an orange color scheme indicates a negative change, or a decrease of mango trees within each cell. For a clearer presentation, values ranging from -1 to 1 were left transparent.

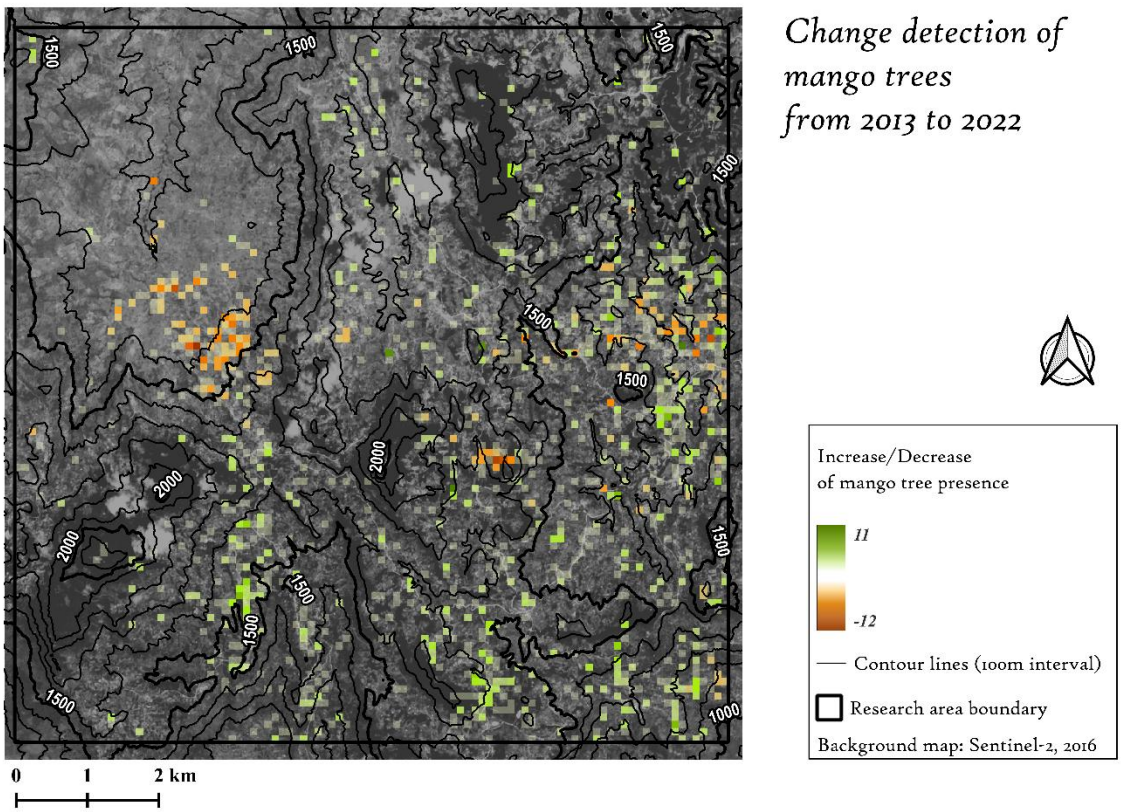

Figure 17. The change in mango tree presence in the study area based on the predictions. Increase is displayed in green and decrease in orange. Values ranging from -1 to 1 are transparent.

The change detection map shows an overall increase of mango tree presence in the study area of Taita Hills. The increase is apparent in the cluster area on the east side, though some cells within that cluster also show a decline. The mango trees have expanded in area mostly at the elevation of 1400 m. The cluster area present in the central area towards west in 2013 appears to have shifted away from the rain-shadow to new areas in the south that are slightly higher in elevation. Additionally, multiple new individual trees have appeared all throughout the study area. Although largest clusters are still found at the same elevations, the trees are occupying new territories, some of which are as high as over 2000 m a.s.l. This concludes that the trees might have spread upwards in elevation.

4.2. Aboveground biomass analysis

Figure 18 shows aboveground biomass estimations of the mango trees that were measured during the fieldwork in 2024. The values were classified into six different classes, which visualized the bigger the greater the AGB. This visualization can be seen to highlight the size of the trees.

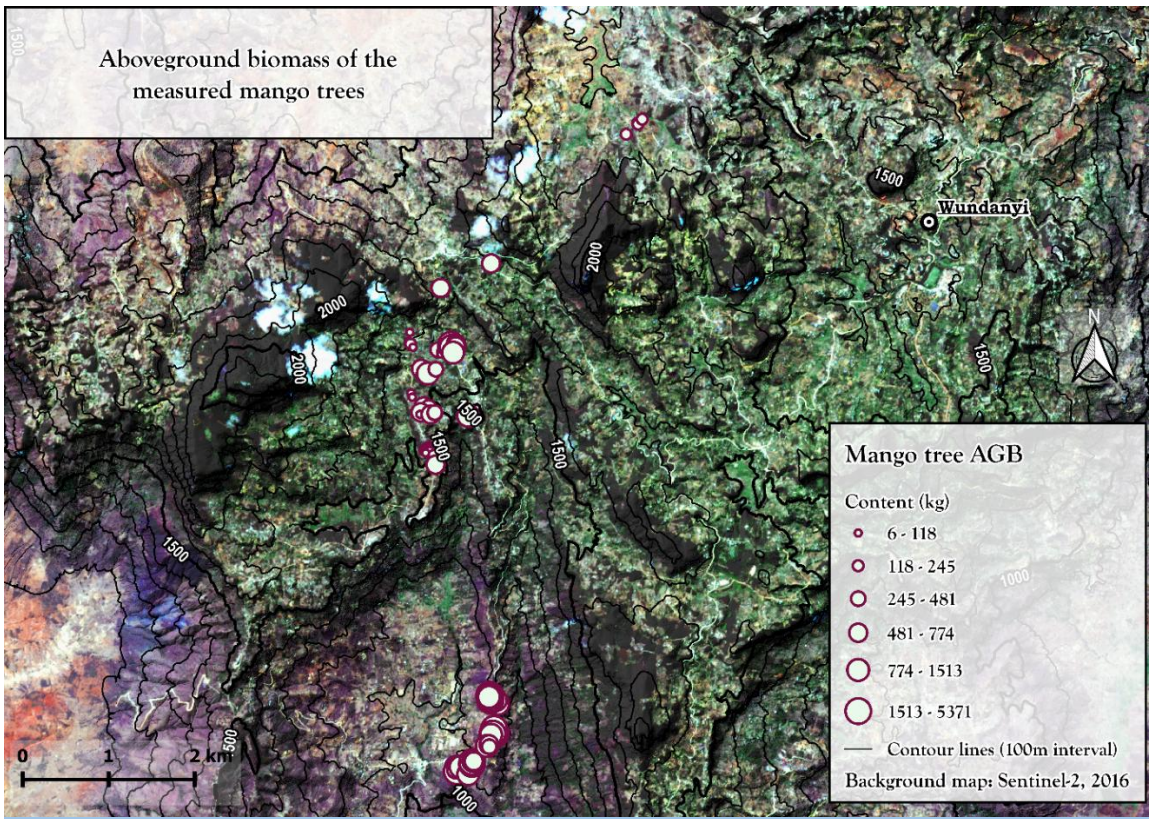

Figure 18. Map of the aboveground biomass estimations. The AGB estimations are calculated and classified larger the bigger the estimation.

The trees measured differed in ages as well as sizes, which explains the vast variation. The trees tend to grow smaller in farms where there are multiple trees, because the space is limited for roots and canopy to spread out. Individual trees have the means to grow bigger. The limited space in soil is a problem especially in higher elevations, where the conditions are otherwise harsher for trees. This can be seen in the northern part of the area in Figure 18, where clusters of the smallest class are growing tightly together, whereas the individual trees are showing greater estimations. In the lower elevations the trees are bigger overall. The tree sites in the southern part of the map are from lower elevations than the ones in the north. Despite growing near other trees, the mangoes still reach greater biomass estimations in the lower areas. However, this map naturally does not take other trees than mangoes into account.

Summary of the AGB and AGCS estimations are shown in Table 5. The AGB estimations range from 5 to 5000 kilograms. The greatest part of the measured mango trees are of medium size and fall between the values of around 193 and 1047 kilograms of AGB or 87 and 471 kilograms of AGCS. A few young trees bring down the reached minimum value of under 6 kg of AGB, which is unexpectedly low. Surprisingly, the minimum value isn't an outlier, and there are a lot of low values both in higher and lower elevations. There were seemingly old and strong ones in the dataset too. The maximum value is one of them, reaching the incredible value of 5371 kg of AGB or 2417 kg in carbon. From Figure 19. it is clear that the maximum value is an outlier. The figure also proves the theory of the trees in lower elevations reaching higher carbon contents than those in the highlands. Even though there is a gap in the elevations visited and R-squared doesn't reach higher than 0,216, a pattern is still visible.

Table 5. Summary of AGB and AGCS contents of the measured mango trees.

value	Aboveground biomass (kg)	Carbon (kg)
Min	5,85	2,63
1st Qu.	193,31	86,99
Median	481,47	216,66
Mean	767,32	345,29
3rd Qu.	1046,85	471,08
Max	5371,27	2417,07
n	110	110

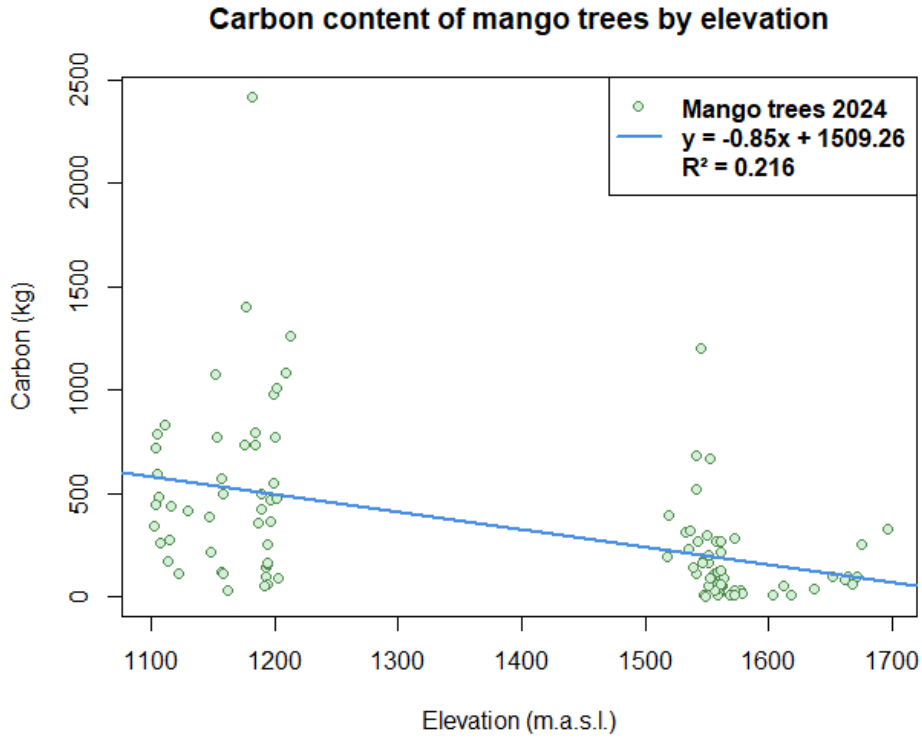

Figure 19. The distribution and regression line of the carbon contents of the measured mango trees by their growing elevation.

4.3. Regression analysis of elevation and tree count

Linear regression of the amount of mango trees and elevation was also studied for a better understanding of the tree distribution in Taita Hills. The linear regression models are presented in Figures 20 and 21. The data of mango trees is counted from the 100x100m prediction grid file, and the elevation is calculated from a Digital Terrain Model (DTM) of Taita Hills. All values under 0,9 were filtered from the analysis.

The scatterplots on the linear regression between mango trees and elevation show a vague but clear negative correlation indicating that there are more mango trees in lower elevations than higher. It is worth noting that the 2022 classification predicts mango trees from higher elevations than 2013, and that this study area only includes elevations above 1000 m a.s.l.. The vast majority of the observations are plots of under 10 trees between the elevations of 1200 to 1700 meters. Observations outside of that cluster reveal the descending trend of the regression line. The correlation coefficients in Table 6 show very similar results for both years, with around -0,2 Pearson and -0,3 Spearman correlations, thus not very strong. R^2 and adjusted R^2 are likewise very low. For comparison, Table 6 also includes these values for the regression model of carbon storage and elevation, which are a bit stronger. Despite the clear tendency for mango trees to grow more in lower elevations, no strong correlation or explanatory power can be drawn from these analyses. However, the scatterplots reveal the distribution of mango trees: the 2022 plot receives higher overall values and for even higher elevations.

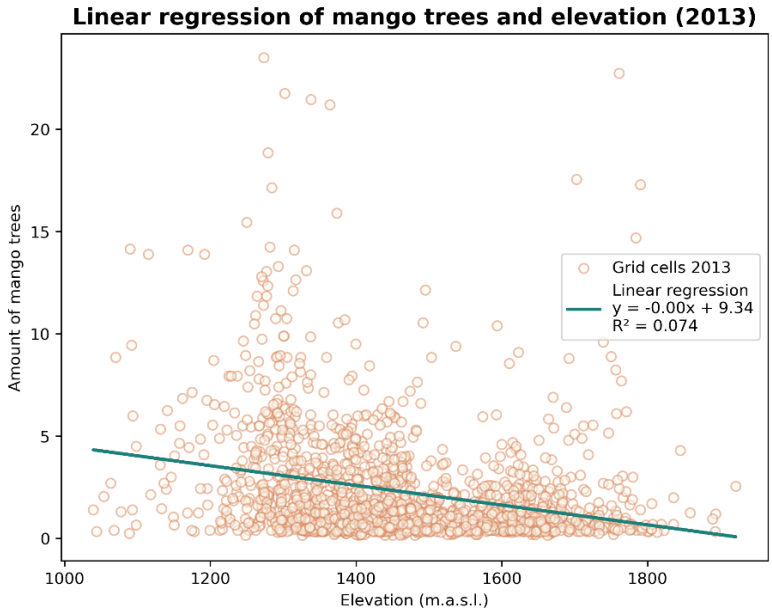

Figure 20. Linear regression line between the amount of mango trees in 2013 and elevation.

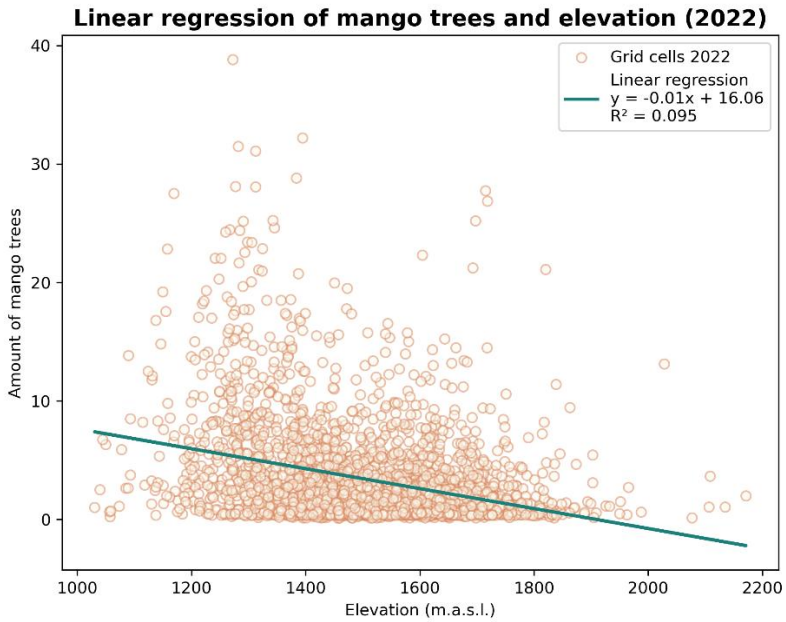

Figure 21. Linear regression line between the amount of mango trees in 2022 and elevation.

Table 6. The correlation and statistical significance of the linear regression models of both the number of trees and carbon content against elevation.

Coefficient	2013	2022	Carbon content (2024)
Pearson	-0,247	-0,298	-0,464
Spearman	-0,33	-0,309	-0,589
R^2	0,074	0,095	0,216
Adjusted R^2	0,086	0,105	0,208

4.4. Temperature data analysis

The mango trees' cooling effect results are based on temperature measurements taken from the trees that were located during the fieldwork. The results shown in Table 7. indicate a clear decrease in the maximum temperatures measured under the tree compared to an open area nearby the tree. This proves the point of mango trees having a cooling effect on their

surroundings. Most of the maximum temperatures outside and under the tree had nearly 10 °C difference, while greatest maximum and minimum temperatures largely exceeded 15 °C. This can be explained by the rapid heating of bare ground during the day. Vegetation slows down these rapid temperature shifts, and mango trees also seem to store the heat for a while longer during the night. The bare soil also cools down quickly at night, and for four out of five trees the minimum temperatures are lower in the open area. The mean and median temperatures of the ground surface are also mainly lower under the tree than outside, with small differences. These results point towards a conclusion that the trees do affect the ground temperatures of their surroundings by protecting them from the heat and from extreme temperature shifts.

Table 7. Ground surface temperatures of the field measures mango trees.

	value	**Mango 2**	**Mango 6**	**Mango 63**	**Mango 65**	**Mango 50***
Under tree	min	15,38	17,38	18,38	18,88	20,62
	median	19,97	19,5	20,19	22	21,88
	mean	20,62	19,75	20,75	22,38	22,03
	max	28,88	27,87	26,44	29,63	25,75
Open area	min	**15**	15,81	17,75	19,25	20,56
	median	18,88	18,13	20,78	21,88	23,06
	mean	21,7	19,56	22,46	23,03	25,05
	max	**38,25**	31	35,19	33,63	35,75

**Monitored for five days, instead of for 24 hours.*

4.5. Interview results

20 interviews were collected during the fieldwork period from different areas in Taita Hills (Figure 22). The respondents of the door-to-door interviews were all mango tree cultivators. The farmers came from household sizes ranging from 1-8 people and land areas of 0,3 acres to 7 hectares. The farmers usually had one or two mango trees, but the highest number of trees was 12. All the farmers practiced agroforestry likewise with other trees, such as guava (*Psidium guajava*), macadamia and grevillea. All of the farmers also grew avocados. Frequently named advantages with mango trees were fruits, shade and firewood, whereas challenges were posed mostly by insects and low or inconsistent production of fruits. Many used the tree leaves for different purposes, such as composting, land nourishing or animal food. Some respondents mentioned that the trees are strong and do not fall easily. A few named that the mango tree casts too much shade, and therefore nothing can grow under it. Otherwise, the trees were often seen attracting rain and fresh air. Only four of the respondents did not have a problem with pests, and all of the other respondents named fruit flies as a pest problem. The mango trees were planted in hopes of fruit yields for domestic use or selling. Only one of the trees in this study was not intentionally planted. The trees' ages varied from one to over 50, the majority still being over 10 years old. Many farmers had recently planted new trees.

Worth mentioning is that none of the respondents were impolite. The interview process was very pleasant to conduct and one of the most insightful things done for this thesis. The interviewees were all welcoming and warm, many were quick to offer chairs and even beverages upon our arrival.

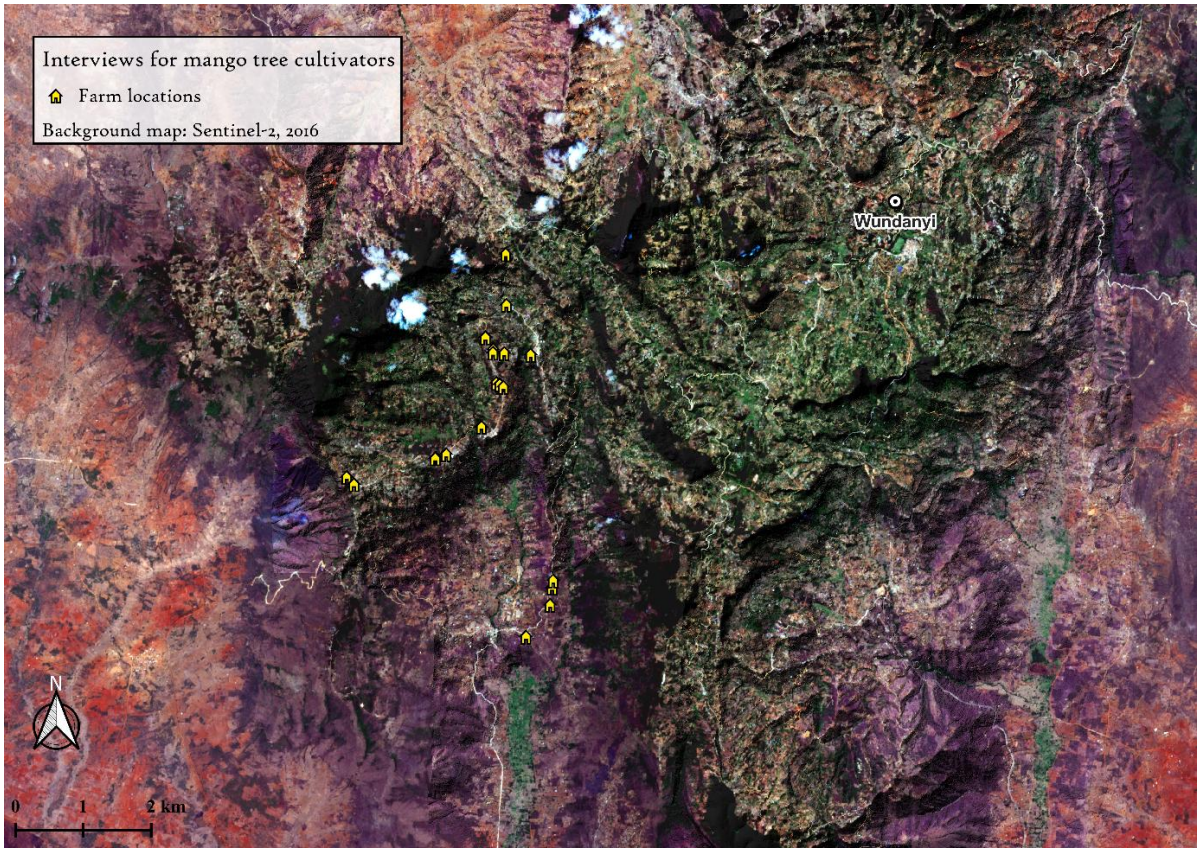

Figure 22. Map of the questionnaires done in the research area. The interview sites are marked yellow.

5. Discussion

Hyperspectral remote sensing and LiDAR data have also been used to map trees in the European setting. A classification of major tree species in a Finnish boreal forest was done using field-detected tree species to train a classification model (Mäyrä et al., 2021), similarly to this thesis' workflow. European aspen (*Populus tremula L*) has been mapped with Overall Accuracy (OA) of 84% combining HRS techniques together with other indices such as Cellulose Absorption Index and Plant Stress Index 3 to discriminate coniferous and deciduous trees (Viinikka et al., 2020). A study conducted in Bavarian Forest national park, Germany, showed that the use of an OCC approach effectively distinguished silver fir (*Abies alba)* from Norway spruce (*Picea abies*) despite their spectral similarities, achieving a puAUC (AUC for positive and unlabeled samples) of 0.95 (Shi et al., 2021). One of the aims was to distinguish the spectral as well as morphological feature differences between silver fir and Norway spruce, both metrics attained using HRS and LiDAR data. The Brazilian Atlantic Forest study demonstrated that combining Unmanned Aerial Vehicle (UAV) hyperspectral data and LiDAR metrics can significantly improve tree species classification in complex environments (Pereira Martins-Neto et al., 2023). Their approach provided accurate tree crown detection by using semi-automatic superpixel segmentation followed by manual correction. Although this thesis focused on mango trees using a deep learning approach, future research in the Taita Hills could benefit from testing a similar segmentation method to improve accuracy. Additionally, exploring the integration of hyperspectral and LiDAR data with principal component analysis, as applied in the Brazilian study, could further refine mango tree mapping techniques in challenging environments. Combining that with more information about the vegetation indices as well as structural tree metrics of the research area and adding that into the model could help to obtain better results.

Mangoes have been widely studied with hyperspectral imaging before. It has been proven a great non-destructive grading method for mango fruit inspection regarding for example maturity (Wendel, Underwood, & Walsh, 2018), ripeness (Gutiérrez, Wendel, & Underwood, 2019), fungal diseases (Siripatrawan & Makino, 2024; Velásquez et al., 2024) as well as changes in the fruit's physiochemical properties (Rungpichayapichet et al., 2017). However, none of these studies have utilized remote sensing, and the scope and areal extent is therefore small. On orchard level, canopy height estimates of mango trees has been found reliable with Airborne Laser Scanning (ALS) derived data (Wu, Johansen, Phinn, Robson, & Tu, 2020), and

yield estimations have been studied by using tree structural parameters mapped with UAV-based LiDAR data (Sarron, Malézieux, Sané, & Faye, 2018). To the best author's knowledge, no previous wide-scale studies have been done on mango tree distribution before. That being said, this study has been innovative and outstandingly the first of its kind. It provides an understanding of the patterns within agroforestry in a tropical Afromontane environment and contributes to the research on the effects of agroforestry on climate change mitigation.

5.1. Improvements of remote sensing techniques

The results of this research seem plausible. The model detected trees well within their canopy areas. The model received high accuracy values, which indicates reliability of the results. Despite receiving conceivable results, there are a lot of matters in the process that could have been improved.

5.1.1. Imbalance in training and testing data

The most distinctive improvement for the one-class classification model would be to even the number of training samples. Piiroinen et al., 2018 conducted an invasive tree species classification analysis in the same research area. For the classification model the training data was derived from around 40 positive labeled tree crowns. For testing they used 22 positive and 215 negative labeled tree crowns, which means around 10/90 ratio. As for the testing samples in this thesis, the ratio of tree crowns in 2022 classification was around 50/50. This imbalance is a result of the lack of negative tree crowns for 2022. For 2013 classification the negative tree amount of over 600 was cut down to 200 in order to avoid this imbalance, but adding more trees is harder than filtering the oversupply of them. The datasets were used consequently to identify more trees for the other year. In other words, 2013 data was used as a reference to identify more negative trees for 2022, and 2022 data to identify mango trees for 2013. Despite these efforts, the 2022 negative trees were really few.

One of the reasons for the lack of adequate negative samples was the subtle inaccuracy of the field data. This can be better obtained in a few ways, one option being to only collect locations from individual trees in the field, so that they can be recognized and labeled with better accuracy. In the data collection phase, it is important to avoid collecting trees from forests or from areas where the canopies overlap even a little. This might mean more time for the

fieldwork, if the research area is this big and the target trees are few and sparse. Another improvement to make is to have the research area on a GNSS locator, or other methods to be certain about the research area boundaries. It is also important to only label the trees with high confidence. Any false positives in the sets disturb the training of the model and should be reduced to a minimum.

5.1.2. Effect of spectral variance in training data

Unsuitably, the classification model classified many trees as falsely positive. False positive means the misclassification of a tree that is incorrectly classified as belonging to the target group, in this case as a mango. Some of these misclassified areas were visited during the fieldwork, which means that they would have been located and collected if they were mango trees. Individual and isolated false positive classifications can be seen across the area, primarily in regions corresponding to the shadowed portions of tree canopies. Given that mango leaves are relatively dark, this characteristic may contribute to these false positive classifications. Moreover, some trees recorded during fieldwork were not classified as positive.

The most difficult thing for the classification model seemed to be the separation between mango and avocado trees. The resulting prediction map showed large forest stands where the mango couldn't possibly grow in such large quantities, hence a harsh filtration process had to be done. Many of the trees known to be avocados were predicted as mangoes in the classification, which couldn't be post-processed. To put it another way, the model produced a lot of false positives. One of the reasons might have been the imbalance in the testing and training sets described in the earlier chapter. Another reason has to do with the very similar spectral reflectance of the trees, also earlier depicted in the methods section. According to Zhao et al., 2022, a challenging problem in tree species detection is the large spectral variation within the same species, where there is subtle variation between different species. As negative class is not clearly defined, similar reflectance of the unlabeled dataset can confuse the learning of the classifier. In this case, the use of avocado as the main non-mango tree in the unlabeled dataset might have affected the results negatively. For 2013, the received dataset of negative tree species varied greatly within itself, while the negative trees for 2022 were mainly avocados. The spectral variance within the same species can also be influenced by bidirectional effects, varying slopes and slope aspects (P. Pellikka, 1998). However, the accuracy of the predictions is not straight-forward to review. Most of the clustered areas were not visited during the

fieldwork, and therefore it is hard to tell empirically if the results are accurate. The most reliable, yet costly way of accuracy assessment would be to visit the research sites afterwards for on-site evaluation, and check if the predicted results are true.

5.1.3. Alternative improvements

There are several other factors that might have altered the results towards better accuracy. One potential is to use pseudo label method, which is a semi-supervised learning technique developed for deep neural networks. The accuracy of the method relies on the class that has maximum predicted probability, which is treated the same way as true labels (Lee, 2013). The advantage of pseudo-labeling is that it can be done with a minimal amount of labeled data. In tree mapping, pseudo-label method has shown to receive improved detection accuracies compared to unsupervised methods (Xiao, Qin, & Huang, 2020). Additionally, pseudo-labeling has also found useful for improving low-resolution markup (Mirpulatov, Illarionova, Shadrin, & Burnaev, 2023), which could have been beneficial in this study as well.

The choice of input data sources is another factor worth contemplating. The possibilities are many, and each data acquisition method has its limitations. For a research area of the size of 10 x 10 km where the topic of interest is to map individual trees, airborne data appears to be the most appropriate option. Drone-based data often lacks the coverage for such large areas, while satellite hyperspectral data is too robust. However, with ever-evolving innovative technology, these arguments might soon be outdated. Tree counting studies have already been conducted successfully using satellite imagery and deep neural networks (Yao, Liu, Qin, Lu, & Zhou, 2021), and efforts to maximize drone coverage have also received promising results in for example planting drone base stations (Azari et al., 2017). This is particularly relevant, given that the flight campaigns are often time-consuming and costly. However, there is yet no functional hyperspectral imaging satellite with a spatial resolution applicable for tree canopy level detection.

Spatial resolution has a critical role in remote sensing-based studies. As previously discussed, even a few meter inaccuracies in the source data can significantly affect the labeling of an object. In tree mapping, a small flaw in both hyperspectral and tree location datasets can quickly become crucial. In the change detection analysis, two different datasets that might

differ from each other in terms of spatial accuracy and precision are compared, and thus the results might be biased to some degree. Even though the change detection is calculated based on the grid cells to minimize this spatial inaccuracy, some trees near the borders of a grid cell might fall into another grid cell's side in the other year's classification result. Another downside in the change detection map is that the effect of tree growth remains unexamined. Instead of calculating the results based on individual tree canopies, it could also be done on the basis of the total amount of pixels. However, in that case it would be harder to distinguish newly emerging trees. Because tree growth happens naturally, the results would indicate an increase regardless.

5.2. Climate change mitigation

The classification results show that the number of trees has increased in the study area. The visual inspection and comparison of the predictions from both years indicate tree growth. The trees have also found new growing habitats and thus migrated upwards in elevation, which is also reported more or less throughout the whole research area. This is a markable result, coupled with the reported cooling effect the tree and its canopy have. Moreover, the increase in tree presence means again an increase in carbon storage.

5.2.1. Tree amount change and the cooling effect

Free-air CO_2 enrichment (FACE) experiments are conducted in effort to study the impact of rising CO_2 levels on plants and ecosystems in natural conditions. These studies have shown that photosynthesis is stimulated by raised CO_2 concentrations, and thus report an approximate 23% biomass increase in tree species in a scenario where atmospheric CO_2 levels are doubled (Kallarackal & Roby, 2012). In addition, LAI in trees has increased by an average 21% in all FACE study sites (Ainsworth & Long, 2005). Since these findings predate the time of our study period, elevated CO_2 concentrations could help explain why the mango trees have increased in Taita Hills. Trees play a significant role in climate change mitigation, and the change detection map can help assess their impact in Taita Hills. Moreover, as trees grow better in areas with more carbon dioxide, studying their distribution can also be seen as an indicator of climate change.

However, FACE studies do not come without limitations. Long-term exposure to elevated CO_2 can drive species to acclimate and down-regulate in the changed conditions, and as a result potentially increasing the water use efficiency of trees (Kallarackal & Roby, 2012). This might cause serious challenges to agricultural areas that already suffer from repeated or prolonged droughts. Another major drawback is the absence of FACE studies in the tropics, limiting the applicability of these models to new ecosystems and climatic conditions.

Trees have been used as a cooling effect in urban areas for a long time. In addition to the trees' transpirative cooling, they also provide shade that increases thermal comfort (Manickathan, Defraeye, Allegrini, Derome, & Carmeliet, 2018). Trees are especially effective in urban environments, for example in preventing Urban Heat Islands (UHI) from overheating (Manickathan et al., 2018). One of the findings was that trees help regulate temperature shifts within their surroundings. The same finding is also studied by (Gosme, Dufour, Inurreta Aguirre & Dupraz, 2016), who argue that the effect that trees have on microclimate means lower temperatures during the day but higher temperatures during the night in agroforestry compared to full sun. A study done in Taita reveals that also the closure of the canopy affects its cooling effect of the canopy cover, which however decreases with increasing altitude (Aalto, 2020). The cooling effect of agroforestry is also mentioned by nearly all the mango tree cultivators interviewed.

5.2.2. AGB prediction and AGCS calculation

Even though destructive sampling to this day is the most efficient way to accurately estimate tree biomass, the study by (Kuyah et al., 2024) strives to develop an allometric equation for a non-destructive estimation. Equation 19. was used in the AGB estimations for mango trees. The equation in question was chosen for this study based on the best accuracy and statistical significance results using the available tree metrics that were measured during the fieldwork. According to the article, the most accurate allometric equations for mango tree AGB were calculated using the measurement of mean diameter of primary branches (DPB) as the primary predictor variable. This article was published only after the fieldwork was already done, and thus the measurements of the primary branches were no longer conveniently attainable.

Regardless, the estimation of aboveground biomass and carbon content of mango trees produced strong results. The prediction based on on-site tree measurements without destructive

sampling has been proven reliable (Kuyah et al., 2024). Similar accuracy was likewise achieved in this study by using the above-mentioned equations. While the overall calculations propose reliable results, the estimations might have slight variation due to rounding factors. Although the estimations are considerably qualitative, the quantitative predictions of AGB and carbon contents are somewhat limited. Allometric equations rely on tree measurements such as height and DBH, which are essential but often challenging to obtain remotely. These measurements require direct in-situ contact with the tree. In turn, the measuring of individual trees is time-consuming, and very inefficient for covering large areas. For carbon content calculations a connection to elevation was also identified, supporting the previous findings. Earlier research shows that models on AGCS distribution, 89%, are related to the variation of elevation, slope and population density (Odeny, 2020). Another study conducted in the same 10 x 10 km research area argues, that 69,8% of the AGB variation is explained by mean annual precipitation, the distribution of croplands and slope (Adhikari et al., 2017).

The main focus of this thesis is on farmlands of Taita Hills. While tropical montane forests contribute unprecedented stocks of AGCS (Cuni-Sanchez et al., 2021) and most significant woody AGB are found in riverine forests (Amara et al., 2020), mango trees provide a large contribution to the AGB in Taita Hills croplands and homesteads. Mango trees contributed to 11,6% of the total AGB as well as received the second highest tree species Importance Value Index (IVI) in Taita Hills (Amara et al., 2023), which is calculated by comparing the number, dominance and frequency of the target tree species and other tree species in the area (Tolangara, Ahmad, & Liline, 2019). After thickets, croplands also account for the most carbon stock in Taita Hills with 4 Mg per hectare (P. K. E. Pellikka et al., 2018). Even though this is mostly explained by the larger area of the croplands, there is no denying that exotic trees can have a large contribution to AGB.

Notably, remote sensing methods can be used for AGB estimations. LiDAR data can be obtained from moderately large areas. LiDAR data can be used to calculate the CHM of the area, which can then further be used for AGB regression. This AGB estimation would not be as accurate as the estimation done with allometric equations, but it would be suitable for a large research area. Additionally, as there is LiDAR data available from both 2013 and 2022, a change detection map could be obtained. This would deepen the analysis of tree growth and carbon sequestration development in the whole research area. Furthermore, a comprehensive

AGB change detection map of the whole area would be very helpful for analyzing the tree growth in Taita Hills and should be taken into consideration in future research.

5.3. Future of mango trees

However, mango trees do seem to have a future. Because of their relatively high and thick canopy, mango trees have a cooling effect on the surroundings by providing shade. The trees store carbon, which is the main component of the greenhouse effect. The mango shows no morphological or physiological adaptations for wind pollination (Kumar, 2016), and it therefore relies on attracting pollinators. Besides, mango trees also produce fruit, which provides food as well as economical safety. This being said, mango trees are certainly important in reducing the negative effects of climate change. In addition, the mango tree is frankly a very well-liked tree species with drawbacks of only few and far between.

It is noted that mango trees do not only migrate upwards all by themselves, but rather this is a result of a change in cultivation practices in Taita Hills. No mangoes were found growing in the wild during the fieldwork period, and all of the 110 mango trees except for one were intentionally planted. There is a possibility that the farmers have been planting mango trees to the highlands before in hope for it to grow, but the seedling has not survived. As mentioned before, the distribution of mango trees has not been studied earlier on a wider scale, which means that the exact timeline of the migration is hard to determine. The interview results show that the trees growing in higher elevations are quite young for the most part. However, the quantity of interviews is low, so the results cannot accurately be implemented for the whole research area. It is also possible that the popularity of mango has been spreading as soon as the first highland farmers have managed to get yield.

6. Conclusions

This thesis performed a previously created deep OCC model framework. The presence of mango trees was mapped inside a 10x10 km study area in Taita Hills, Kenya. Taita Hills is a mountainous biodiversity hotspot with elevations ranging from 1100 to 2200 m a.s.l. towering over the surrounding lowlands. Field work was conducted in January of 2024 across the Taita Hills region. The fieldwork included the collection of tree locations of mangoes and other trees, as well as measuring mango tree dimensions for the AGB estimations and conducting interviews. Methods included training the classification framework called ITreeDet. Training data included MNF patches derived from the hyperspectral imagery that contained spectral reflectance data of the tree canopies.

The research found that mango trees have increased in the study area, as well as mitigated upwards in elevation. The most increase has, however, happened in their previous growing habitats around elevations of 1400-1600 m a.s.l. A PU-learning based one-class classification approach using hyperspectral imaging was proven suitable for mango tree mapping in the Taita Hills area. However, closer attention should be paid to the balance of training and testing sets in the model training phase, as well as to the adequacy of the non-mango trees in the selection of unlabeled samples.

This study provides valuable insights into the distribution patterns of mango trees, as well as potential in climate change mitigation in a tropical montane environment. Expanding the dataset with more adequate training and testing sets would greatly enhance the reliability of the results. A comprehensive map on the mango trees' contribution to the local AGB could be produced with the existing LiDAR datasets along with the knowledge of the measurements needed for the AGB estimations. These findings highlight the importance of combining remote sensing techniques with field-based measurements to improve our understanding of tree dynamics in montane ecosystems, paving the way for more effective research within the remote sensing field.

7. Acknowledgements

This research was conducted as a part of the ESSA project (Earth Observation and Environmental Sensing for Climate-Smart Sustainable Agropastoral Ecosystem Transformation in East Africa, 2020-2024) funded by EU DeSIRA Lift (Contract number: Ref. Ares(2020)7862067 - 22/12/2020).

I would like to thank everyone who has supported me during the composition of this thesis to make it possible. I want to express my gratitude to Mwadime Mjomba for the initial idea for this study, Peter Mwasi and Darius Mwambala for the assistance and good company during the fieldwork period, Ashfak Mahmud and Julius Omondi for sharing your work, not to mention all the lovely staff of Taita Research Station. Thank you Ilja Vuorinne for sharing your knowledge and Janne Heiskanen for the background support throughout the thesis process.

My sincere thanks go to my supervisors: Professor Petri Pellikka for this incredible opportunity and encouragement that has allowed me to have so many new experiences I would never have even dreamed of, and Zhaozhi Luo for the assistance and continuous support you have provided me with everything ranging from a problem with the code to collected guidance for the process to move forward.

Lastly, I extend my appreciation to my family and friends in and outside of university. Without your kind words, encouragement and peer support throughout my life as a student, this thesis would have been long overdue.

Declaration of generative AI

Artificial Intelligence, namely DeepL and ChatGPT, have been used in the elaboration of this thesis for text formatting ideas and source summarizing purposes, as well as for code generation assistance. Nothing has been directly copied, and the text is written completely by the author herself.

8. References

Aalto, I. (2020). *Assessing the cooling impact of tree canopies in an intensively modified tropical landscape* (Master's thesis). University of Helsinki, Master's Programme in Geography.

Abera, T. A., Vuorinne, I., Munyao, M., Pellikka, P. K. E., & Heiskanen, J. (2022). Land Cover Map for Multifunctional Landscapes of Taita Taveta County, Kenya, Based on Sentinel-1 Radar, Sentinel-2 Optical, and Topoclimatic Data. *Data*, *7*(3), 36. doi: 10.3390/data7030036

Adhikari, H., Heiskanen, J., Siljander, M., Maeda, E., Heikinheimo, V., & K. E. Pellikka, P. (2017). Determinants of Aboveground Biomass across an Afromontane Landscape Mosaic in Kenya. *Remote Sensing*, *9*(8), 827. doi: 10.3390/rs9080827

Ainsworth, E. A., & Long, S. P. (2005). What have we learned from 15 years of free-air CO_2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO_2. *New Phytologist*, *165*(2), 351–372. doi: 10.1111/j.1469-8137.2004.01224.x

Albrecht, A., & Kandji, S. T. (2003). Carbon sequestration in tropical agroforestry systems. *Agriculture, Ecosystems & Environment*, *99*(1–3), 15–27. doi: 10.1016/S0167-8809(03)00138-5

Alonzo, M., McFadden, J. P., Nowak, D. J., & Roberts, D. A. (2016). Mapping urban forest structure and function using hyperspectral imagery and lidar data. *Urban Forestry & Urban Greening*, *17*, 135–147. doi: 10.1016/j.ufug.2016.04.003

Amara, E., Adhikari, H., Heiskanen, J., Siljander, M., Munyao, M., Omondi, P., & Pellikka, P. (2020). Aboveground Biomass Distribution in a Multi-Use Savannah Landscape in Southeastern Kenya: Impact of Land Use and Fences. *Land*, *9*(10), 381. doi: 10.3390/land9100381

Amara, E., Adhikari, H., Mwamodenyi, J. M., Pellikka, P. K. E., & Heiskanen, J. (2023). Contribution of Tree Size and Species on Aboveground Biomass across Land Cover Types in the Taita Hills, Southern Kenya. *Forests*, *14*(3), 642. doi: 10.3390/f14030642

Bourel, M., & Segura, A. M. (2018). Multiclass classification methods in ecology. *Ecological Indicators*, *85*, 1012–1021. doi: 10.1016/j.ecolind.2017.11.031

Brantley, S. T., Zinnert, J. C., & Young, D. R. (2011). Application of hyperspectral vegetation indices to detect variations in high leaf area index temperate shrub thicket canopies. *Remote Sensing of Environment*, *115*(2), 514–523. doi: 10.1016/j.rse.2010.09.020

Brooks, T., Lens, L., Barnes, J., Barnes, R., Kihuria, J. K., & Wilder, C. (1998). The conservation status of the forest birds of the Taita Hills, Kenya. *Bird Conservation International*, *8*(2), 119–139. doi: 10.1017/S0959270900003221

Burgess, N. D., Butynski, T. M., Cordeiro, N. J., Doggart, N. H., Fjeldså, J., Howell, K. M., … Stuart, S. N. (2007). The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. *Biological Conservation*, *134*(2), 209–231. doi: 10.1016/j.biocon.2006.08.015

Castalonge, O. W. (Ed.). (2008). *Agricultural systems: Economics, technology and diversity*. New York, NY: Nova Science Publ.

Cho, L., Yoon, J., & An, G. (2017). The control of flowering time by environmental factors. *The Plant Journal*, *90*(4), 708–719. doi: 10.1111/tpj.13461

Collier, P., Conway, G., & Venables, T. (2008). Climate change and Africa. *Oxford Review of Economic Policy*, *24*(2), 337–353. doi: 10.1093/oxrep/grn019

Cuni-Sanchez, A., Sullivan, M. J. P., Platts, P. J., Lewis, S. L., Marchant, R., Imani, G., … Zibera, E. (2021). High aboveground carbon stock of African tropical montane forests. *Nature*, *596*(7873), 536–542. doi: 10.1038/s41586-021-03728-4

Di Minin, E., Slotow, R., Fink, C., Bauer, H., & Packer, C. (2021). A pan-African spatial assessment of human conflicts with lions and elephants. *Nature Communications*, *12*(1), 2978. doi: 10.1038/s41467-021-23283-w

Dian, Y., Pang, Y., Dong, Y., & Li, Z. (2016). Urban Tree Species Mapping Using Airborne LiDAR and Hyperspectral Data. *Journal of the Indian Society of Remote Sensing*, *44*(4), 595–603. doi: 10.1007/s12524-015-0543-4

Dianda, O. Z., Wonni, I., Ouédraogo, L., Sankara, P., Tollenaere, C., Del Ponte, E. M., &

Fernandez, D. (2023). Lasiodiplodia species associated with mango (Mangifera indica L.) decline in Burkina Faso and influence of climatic factors on the disease distribution. *Physiological and Molecular Plant Pathology*, *126*, 102041. doi: 10.1016/j.pmpp.2023.102041

Ferreira, M. P., Zortea, M., Zanotta, D. C., Shimabukuro, Y. E., & De Souza Filho, C. R. (2016). Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. *Remote Sensing of Environment*, *179*, 66–78. doi: 10.1016/j.rse.2016.03.021

Gosme, M., Dufour, L., Inurreta Aguirre, H. D., & Dupraz, C. (2016). *Microclimatic effect of agroforestry on diurnal temperature cycle*. *1230*, 466. Montpellier, France.

Gutiérrez, S., Wendel, A., & Underwood, J. (2019). Ground based hyperspectral imaging for extensive mango yield estimation. *Computers and Electronics in Agriculture*, *157*, 126–135. doi: 10.1016/j.compag.2018.12.041

Haile, G. G., Tang, Q., Hosseini-Moghari, S., Liu, X., Gebremicael, T. G., Leng, G., … Yun, X. (2020). Projected Impacts of Climate Change on Drought Patterns Over East Africa. *Earth's Future*, *8*(7), e2020EF001502. doi: 10.1029/2020EF001502

Jahurul, M. H. A., Zaidul, I. S. M., Ghafoor, K., Al-Juhaimi, F. Y., Nyam, K.-L., Norulaini, N. A. N., … Mohd Omar, A. K. (2015). Mango (Mangifera indica L.) by-products and their valuable components: A review. *Food Chemistry*, *183*, 173–180. doi: 10.1016/j.foodchem.2015.03.046

Kallarackal, J., & Roby, T. J. (2012). Responses of trees to elevated carbon dioxide and climate change. *Biodiversity and Conservation*, *21*(5), 1327–1342. doi: 10.1007/s10531-012-0254-x

Khan, S. S., & Madden, M. G. (2014). One-class classification: Taxonomy of study and review of techniques. *The Knowledge Engineering Review*, *29*(3), 345–374. doi: 10.1017/S026988891300043X

Kilickaya, S., Ahishali, M., Sohrab, F., Ince, T., & Gabbouj, M. (2023). Hyperspectral Image Analysis with Subspace Learning-based One-Class Classification. *2023 Photonics &*

Electromagnetics Research Symposium (PIERS), 953–959. Prague, Czech Republic: IEEE. doi: 10.1109/PIERS59004.2023.10221460

Kumar, S. (2016). Role of Insects in Pollination of Mango Trees. *International Research Journal of Biological Sciences*, *5*(1), 64–67.

Kuyah, S., Muthuri, C., Wakaba, D., Cyamweshi, A. R., Kiprotich, P., & Mukuralinda, A. (2024). Allometric equations and carbon sequestration potential of mango (Mangifera indica) and avocado (Persea americana) in Kenya. *Trees, Forests and People*, *15*, 100467. doi: 10.1016/j.tfp.2023.100467

Litz, R. E. (2009). *The mango: Botany, production and uses* (2nd ed). Wallingford: CABI.

Liu, L., Coops, N. C., Aven, N. W., & Pang, Y. (2017). Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data. *Remote Sensing of Environment*, *200*, 170–182. doi: 10.1016/j.rse.2017.08.010

Luo, G., Chen, G., Tian, L., Qin, K., & Qian, S.-E. (2016). Minimum Noise Fraction versus Principal Component Analysis as a Preprocessing Step for Hyperspectral Imagery Denoising. *Canadian Journal of Remote Sensing*, *42*(2), 106–116. doi: 10.1080/07038992.2016.1160772

Magarik, Y. A. S., Roman, L. A., & Henning, J. G. (2020). How should we measure the DBH of multi-stemmed urban trees? *Urban Forestry & Urban Greening*, *47*, 126481. doi: 10.1016/j.ufug.2019.126481

Makhmale, S., Bhutada, P., Yadav, L., & Yadav, B. K. (2016). Impact of climate change on phenology of Mango – The case study. *Ecology, Environment and Conservation*, *22*(9), S127–S132.

Manickathan, L., Defraeye, T., Allegrini, J., Derome, D., & Carmeliet, J. (2018). Parametric study of the influence of environmental factors and tree properties on the transpirative cooling effect of trees. *Agricultural and Forest Meteorology*, *248*, 259–274. doi: 10.1016/j.agrformet.2017.10.014

Martins, D. J. (2008). Pollination Observations Of the African Violet In the Taita Hills, Kenya. *Journal of East African Natural History*, *97*(1), 33–42. doi: 10.2982/0012-

8317(2008)97[33:POOTAV]2.0.CO;2

McMullin, S., Njogu, K., Wekesa, B., Gachuiri, A., Ngethe, E., Stadlmayr, B., … Kehlenbeck, K. (2019). Developing fruit tree portfolios that link agriculture more effectively with nutrition and health: A new approach for providing year-round micronutrients to smallholder farmers. *Food Security*, *11*(6), 1355–1372. doi: 10.1007/s12571-019-00970-7

Mitchell, J. F. B. (1989). The “Greenhouse” effect and climate change. *Reviews of Geophysics*, *27*(1), 115–139. doi: 10.1029/RG027i001p00115

Mjomba, M. (2023). *Oral information*.

Musvoto, C., & Campbell, B. M. (1995). Mango trees as components of agroforestry sytsems in Mangwende, Zimbabwe. *Agroforestry Systems*, *32*(3), 247–260. doi: 10.1007/BF00711713

Naik, S. K., Sarkar, P. K., Das, B., Singh, A. K., & Bhatt, B. P. (2019). Biomass production and carbon stocks estimate in mango orchards of hot and sub-humid climate in eastern region, India. *Carbon Management*, *10*(5), 477–487. doi: 10.1080/17583004.2019.1642043

Newman, R. J. S., Marchant, R., Enns, C., & Capitani, C. (2021). Assessing the impacts of land use and climate interactions on beekeeping livelihoods in the Taita Hills, Kenya. *Development in Practice*, *31*(4), 446–461. doi: 10.1080/09614524.2020.1854689

Normand, F., Lauri, P.-E., & Legave, J.-M. (2015). CLIMATE CHANGE AND ITS PROBABLE EFFECTS ON MANGO PRODUCTION AND CULTIVATION. *Acta Horticulturae*, (1075), 21–31. doi: 10.17660/ActaHortic.2015.1075.1

Odeny, D. (2020). *Spatial modeling of Biodiversity and Carbon storage along the inhabited slopes of Mount Kilimanjaro (Tanzania) and Taita Hills (Kenya)*. University of Nairobi, School of Engineering, Department of Geospatial and Space Technology.

Pantera, A., Mosquera-Losada, M. R., Herzog, F., & Den Herder, M. (2021). Agroforestry and the environment. *Agroforestry Systems*, *95*(5), 767–774. doi: 10.1007/s10457-021-00640-8

Pellikka, P. (1998). Development of correction chain for multispectral airborne video data. *IGARSS '98. Sensing and Managing the Environment. 1998 IEEE International Geoscience and Remote Sensing. Symposium Proceedings. (Cat. No.98CH36174)*, 2722–2725 vol.5. Seattle, WA, USA: IEEE. doi: 10.1109/IGARSS.1998.702330

Pellikka, P. K. E., Clark, B. J. F., Gosa, A. G., Himberg, N., Hurskainen, P., Maeda, E., … Siljander, M. (2013). Agricultural Expansion and Its Consequences in the Taita Hills, Kenya. In *Developments in Earth Surface Processes* (Vol. 16, pp. 165–179). Elsevier. doi: 10.1016/B978-0-444-59559-1.00013-X

Pellikka, Petri, Ylhäisi, J., & Clark, B. (2004). *Taita Hills and Kenya, 2004: Seminars, reports and journal of a field expedition to Kenya*. Helsinki: University of Helsinki, Department of Geography.

Piiroinen, R. (2018). *Airborne imaging spectroscopy in mapping of heterogeneous tropical land cover in Eastern Africa* (ACADEMIC DISSERTATION, University of Helsinki). University of Helsinki, DEPARTMENT OF GEOSCIENCES AND GEOGRAPHY A69. Retrieved from http://urn.fi/URN:ISBN:978-951-51-3992-4

Piiroinen, R., Fassnacht, F. E., Heiskanen, J., Maeda, E., Mack, B., & Pellikka, P. (2018). Invasive tree species detection in the Eastern Arc Mountains biodiversity hotspot using one class classification. *Remote Sensing of Environment*, *218*, 119–131. doi: 10.1016/j.rse.2018.09.018

Piiroinen, R., Heiskanen, J., Maeda, E., Viinikka, A., & Pellikka, P. (2017). Classification of Tree Species in a Diverse African Agroforestry Landscape Using Imaging Spectroscopy and Laser Scanning. *Remote Sensing*, *9*(9), 875. doi: 10.3390/rs9090875

Piiroinen, R., Heiskanen, J., Mõttus, M., & Pellikka, P. (2015). Classification of crops across heterogeneous agricultural landscape in Kenya using AisaEAGLE imaging spectroscopy data. *International Journal of Applied Earth Observation and Geoinformation*, *39*, 1–8. doi: 10.1016/j.jag.2015.02.005

Prates, A. R., Züge, P. G. U., Leonel, S., Souza, J. M. A., & Ávila, J. D. (2021). Flowering

induction in mango tree: Updates, perspectives and options for organic agriculture. *Pesquisa Agropecuária Tropical*, *51*, e68175. doi: 10.1590/1983-40632021v5168175

Pu, R. (2017). *Hyperspectral Remote Sensing: Fundamentals and Practices* (1st ed.). Boca Raton : Taylor & Francis, [2017]: CRC Press. doi: 10.1201/9781315120607

Ramaswamy, H., Scott, C., & Tewari, A. (2016). Mixture Proportion Estimation via Kernel Embeddings of Distributions. *Proceedings of The 33rd International Conference on Machine Learning*, *48*, 2052–2060.

Roy, P. S. (1989). Spectral reflectance characteristics of vegetation and their use in estimating productive potential. *Proceedings / Indian Academy of Sciences*, *99*(1), 59–81. doi: 10.1007/BF03053419

Sarron, J., Malézieux, É., Sané, C. A. B., & Faye, É. (2018). Mango Yield Mapping at the Orchard Scale Based on Tree Structure and Land Cover Assessed by UAV. *Remote Sensing*, *10*(12), 1900. doi: 10.3390/rs10121900

Schupp, E. W. (1993). Quantity, quality and the effectiveness of seed dispersal by animals. *Vegetatio*, *107–108*(1), 15–29. doi: 10.1007/BF00052209

Singh, L. B. (1960). *The Mango: Botany, Cultivation, and Utilization*. L. Hill. Retrieved from https://books.google.fi/books?id=iNJJAAAAYAAJ

Soini, E. (2006). *Livelihood, land use and environment interactions in the highlands of East Africa* (ACADEMIC DISSERTATION, University of Helsinki). University of Helsinki, Department of Geography. Retrieved from https://helda.helsinki.fi/server/api/core/bitstreams/d2d71601-4988-44f1-9730-b5987756356c/content

Tax, D. M. J., & Duin, R. P. W. (2001). Combining One-Class Classifiers. In J. Kittler & F. Roli (Eds.), *Multiple Classifier Systems* (pp. 299–308). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/3-540-48219-9_30

Tharanathan, R. N., Yashoda, H. M., & Prabha, T. N. (2006). Mango *(Mangifera indica L.)* , "The King of Fruits"—An Overview. *Food Reviews International*, *22*(2), 95–123. doi: 10.1080/87559120600574493

Thijs, K. W., Aerts, R., Van De Moortele, P., Aben, J., Musila, W., Pellikka, P., … Muys, B. (2015). Trees in a human-modified tropical landscape: Species and trait composition and potential ecosystem services. *Landscape and Urban Planning*, *144*, 49–58. doi: 10.1016/j.landurbplan.2015.07.015

Tolangara, A., Ahmad, H., & Liline, S. (2019). The Composition and Important Value Index of Trees for Wildlife Feed in Bacan Island, South Halmahera. *IOP Conference Series: Earth and Environmental Science*, *276*(1), 012037. doi: 10.1088/1755-1315/276/1/012037

USDA. (2011). *USDA Agroforestry Strategic Framework*. Retrieved from https://www.usda.gov/sites/default/files/documents/AFStratFrame_FINAL-lr_6-3-11.pdf

Wallace, L., Sun, Q. (Chayn), Hally, B., Hillman, S., Both, A., Hurley, J., & Martin Saldias, D. S. (2021). Linking urban tree inventories to remote sensing data for individual tree mapping. *Urban Forestry & Urban Greening*, *61*, 127106. doi: 10.1016/j.ufug.2021.127106

Wendel, A., Underwood, J., & Walsh, K. (2018). Maturity estimation of mangoes using hyperspectral imaging from a ground based mobile platform. *Computers and Electronics in Agriculture*, *155*, 298–313. doi: 10.1016/j.compag.2018.10.021

Wild, J., Kopecký, M., Macek, M., Šanda, M., Jankovec, J., & Haase, T. (2019). Climate at ecologically relevant scales: A new temperature and soil moisture logger for long-term microclimate measurement. *Agricultural and Forest Meteorology*, *268*, 40–47. doi: 10.1016/j.agrformet.2018.12.018

Xiao, Q., & McPherson, E. G. (2005). Tree health mapping with multispectral remote sensing data at UC Davis, California. *Urban Ecosystems*, *8*(3–4), 349–361. doi: 10.1007/s11252-005-4867-7

Zhang, C., & Qiu, F. (2012). Mapping Individual Tree Species in an Urban Forest Using Airborne Lidar Data and Hyperspectral Imagery. *Photogrammetric Engineering & Remote Sensing*, *78*(10), 1079–1087. doi: 10.14358/PERS.78.10.1079

Zhang, X., Treitz, P. M., Chen, D., Quan, C., Shi, L., & Li, X. (2017). Mapping mangrove forests using multi-tidal remotely-sensed data and a decision-tree-based procedure. *International Journal of Applied Earth Observation and Geoinformation*, *62*, 201–214. doi: 10.1016/j.jag.2017.06.010

Zhao, H., Zhong, Y., Wang, X., Hu, X., Luo, C., Boitt, M., … Pellikka, P. (2022). Mapping the distribution of invasive tree species using deep one-class classification in the tropical montane landscape of Kenya. *ISPRS Journal of Photogrammetry and Remote Sensing*, *187*, 328–344. doi: 10.1016/j.isprsjprs.2022.03.005

Zhao, H., Zhong, Y., Wang, X., & Shu, H. (2023). One-Class Risk Estimation for One-Class Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–17. doi: 10.1109/TGRS.2023.3292929

Appendices

Appendix 1. Interview questions

Size of the household?

Size of the land area?

How many mango trees do you currently have?

Were the mango trees in the fields intentionally planted or was it as a result of seed disposal after eating the fruit? If so, what was the reason for that decision considering the altitude?

What are you using the mango tree for?

Do you grow crops under the tree or use it as a shadow to something?

Do you use the branches for timber? (yes/no)

Do you sell mango fruits? (yes/no)

How old is/are the mango tree(s)?

How long did it take to start producing fruit? Or has it ever produced fruit?

At what time of the year are mango trees in Taita Hills starting to flower?

Which varieties of mangoes do you have?

Are there more pests attacking mangoes in upper Taita compared to lower Taita? Which pests?

What other trees do you have than mangoes?

What are the challenges and advantages of mango production in Taita Hills in relation to other trees?