COLLEGE OF SOCIAL SCIENCES

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL STUDIES

Modeling Water Hyacinth (*Pontederia crassipes*) Distribution Using Geospatial Techniques: The Case of Lake Tana, Ethiopia

By

Matiwos Belayhun

Addis Ababa University

November 2023

COLLEGE OF SOCIAL SCIENCES

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL STUDIES

Modeling Water Hyacinth (*Pontederia crassipes*) Distribution Using Geospatial Techniques: The Case of Lake Tana, Ethiopia

By

Matiwos Belayhun

Advisor

Asnake Mekuriaw (PhD)

A Thesis Submitted to the Department of Geography and Environmental Studies, College of Social Sciences of Addis Ababa University in partial fulfillment of the requirement for the Degree of Master of Arts in Geographic Information System, Remote Sensing and Digital Cartography.

Addis Ababa University

November 2023

ADDIS ABABA UNIVERSITY

COLLEGE OF SOCIAL SCIENCES

DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL STUDIES

APPROVAL SHEET

This is to certify that the thesis prepared by Matiwos Belayhun entitled "Modeling Water Hyacinth (*Pontederia crassipes*) Distribution Using Geospatial Techniques: The Case of Lake Tana, Ethiopia" is Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Arts in Geography and Environmental Studies (specialization in Geographic Information System, Remote Sensing and Digital Cartography) compiles with the regulations of the university and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Chairperson	Signature	Date
______________________	______________	______________
Advisor	Signature	Date
______________________	______________	______________
Internal Examiner	Signature	Date
______________________	______________	______________
External Examiner	Signature	Date
______________________	______________	______________

____

Chair of Department or Graduate Program Coordinator

ADDIS ABABA UNIVERSITY

COLLEGE OF SOCIAL SCIENCES

DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL STUDIES

Declaration

I declare that this thesis prepared for the partial fulfillment of the requirements for the Degree of Master of Arts in Geography and Environmental Studies (specialization in Geographic Information System, Remote Sensing and Digital Cartography) entitled "Modeling Water Hyacinth (*Pontederia crassipes*) Distribution Using Geospatial Techniques: The Case of Lake Tana, Ethiopia" is my original research work prepared independently by my own effort with the close advice and guidance of my adviser. I also declare that this thesis has not been presented in any university and all sources that I have used or quoted have been indicated and acknowledged by means of complete references.

Name: **Matiwos Belayhun**

Signature: ______________________

Data of Submission: **November, 2023**

Table of Content

Contents **Page**

List of Tables

List of Figures

Abbreviations

AUC	Area Under Curve
ANN	Artificial Nural Network
ARLTWBPDA	Amhara Region Lake Tana and other water bodies Protection and Development Agency
BRT	Boosted Regression Tree
CART	Classification and Regression Trees
COR	Coefficient of Rank Correlation
ESA	European Space Agency
EEPO	European and Mediterranean Plant Protection Organization
GAM	General Additive Model
GBM	Generalized Boosting Model
GCP	Ground Control Point
GEE	Google Earth Engine
GLM	Generalized Linear Model
LAUBAR	Land Administration and Use Bureau of Amhara Region
LiDAR	Light Detection and Ranging
LULC	Land Use/Land Cover
Maxent	Maximum Entropy
IDW	Inverse Distance Weight
IUCN	International Union for Conservation of Nature
RF	Random Forest
ROC	Recover Operator Curve
SAR	Synthetic Aperture Radar
SDM	Species Distribution Modeling
SVM	Support Vector Machine
TSS	True Skill Statistics
VH	Vertical Horizontal Polarization
VV	Vertical Vertical Polarization

Abstract

Water hyacinth, an invasive aquatic plant, poses serious environmental and socioeconomic challenges. Remote sensing is essential to understand and predict how species are distributed in different habitats and environmental conditions and helps with monitoring and management activities. This study is intended to model the distribution and detection of the spatiotemporal dynamics of water hyacinth via four machine-learning models in Lake Tana, Ethiopia. The study employs 11 variables for modeling and 16 variables for spatiotemporal dynamics obtained from Sentinel-1 SAR bands, Sentinel-2A bands and indices, and bioclimate data sources. The models used 458 presence and 458 randomly generated pseudoabsence as response variables and tenfold bootstrap sampling. The models were evaluated using the area under the curve (AUC), receiver operator curve (ROC), true skill statistics (TSS), coefficient of rank correlation (COR), sensitivity, specificity, and Kappa coefficient, while the spatiotemporal distribution between 2016 and 2022 was evaluated using the overall accuracy and kappa coefficient. The findings demonstrate that the random forest model outperforms the other models, with AUC values of 0.93 and 0.95, TSS values of 0.77 and 0.82, and kappa values of 0.76 and 0.82 in the wet and dry seasons, respectively. B12 (16.3% and 19.7%), NDWI (14.7% and 12.4%), mean annual temperature (13.4% and 14.2%), and B5 (11.4% and 12.4%) were the most relevant variables during the wet and dry seasons, respectively, while B3, B5, B11, B12, VH, elevation, NDAVI and NDWI were the most relevant features in the spatiotemporal detection. According to the model prediction result, water hyacinths have the highest coverage during the wet season. The spatial coverage was 686.5 and 650.4 ha in 2016 and 1436.5 and 1216.5 ha in 2022 in the wet and dry seasons, respectively. Study results showed manual removal and machine harvesting were used to manage water hyacinth. The research concludes that the integration of Sentinel image indices and bands with bioclimatic variables is essential in the modeling and detection of spatiotemporal dynamics. The research recommends that geospatial technology helps in the regular assessments and timely detection of water hyacinths as a response to new infestations and prompt management actions.

Keywords Phrases: Water hyacinth, remote sensing, machine learning, modeling, spatiotemporal dynamics, Lake Tana.

Acknowledgment

I would like to acknowledge and express my deep gratitude to God for the blessings, guidance, and strength provided. I would like to express my deepest gratitude to my research advisor, Dr. Asnake Mekuriaw, for his valuable guidance and support throughout this research. His expertise and insightful feedback have been instrumental in shaping the direction and outcomes of this study. I am also immensely grateful to the RCMRD GMES Africa Scholarship for providing the necessary funding and resources that made this research possible.

Furthermore, I would like to extend my sincere appreciation to Addis Ababa University for providing me with a conducive academic environment and access to various research facilities. Their commitment to nurturing intellectual growth has played a crucial role in enhancing the quality and depth of this research. I would also like to acknowledge Copernicus Sentinel image providers for their contribution in providing valuable satellite imagery that was essential for the source of the data in this study.

Last, but not least, I would like to express my heartfelt thanks to my family and friends who have supported me throughout this research endeavor. Their unwavering encouragement, understanding, and belief in my abilities have been a constant source of motivation and inspiration. Without the collective efforts of these individuals and institutions, this research would not have been possible. I am truly grateful for their contributions and support.

CHAPTER ONE

1. INTRODUCTION

1.1 Background of the study

An aquatic plant called water hyacinth (*Pontederia crassipes*) is a flowering, free-floating macrophyte and highly harmful plant native to the Amazon, and Brazil in tropical South America (Simpson et al., 2022). The plant's status has evolved over-time from being a beautiful ornamental plant to an invasive alien species that has harmed socioeconomic development and natural water systems. The species is on the EPPO A2 list of weeds that are suggested for management and is one of the most harmful aquatic weeds globally (Laranjeira and Nadais, 2008; Wang and Yan, 2017).

Water hyacinth features glossy, spherical green leaves and large, pale-blue flowers with purple and yellow spots on the petals (Phiri and Navarro, 2000). In favorable conditions, water hyacinths can grow rapidly and reproduce at a high rate (Dersseh et al., 2019b). The plant has sexual and vegetative reproduction (Gaikwad and Gavande, 2017). The process of vegetative reproduction is carried out by stolon, which are stem-like structures that develop along the water's surface and give rise to new plants at their nodes. Sexual reproduction takes place through the production of seeds, which are dispersed by animals and water currents (Otieno et al., 2022).

The plant was first discovered in Africa (South Africa and Egypt) in 1816. Since then, it has spread across the entire continent, invading streams, rivers, natural lakes, and impoundments (Hill and Coetzee, 2008). Globally, in the areas between 40° north and 45° south, it is found in at least 62 different nations (Wimalarathne and Perera, 2019). In the 1950s, it was first brought to Ethiopia as an ornamental plant in the Rift Valley. Lake Koka was the first to become infected, and it later migrated to the Aba-Samuel Dam, Lake Ellen, Koka Dam, and Wonji sites, where it proliferated spectacularly. Currently, it is found throughout the whole Ethiopian Rift Valley system (Dersseh et al., 2020). Lake Tana and Rift Valley region water bodies that are situated at low, mid, and high altitudes are infested (Firehun, 2017). The plant was found in the Abay River's marshy areas early in the 1990s. However, the plant was largely identified in 2011 close to the Megech River. In 2011, it was officially known as a harmful weed that threatens the environment (Tewabe, 2015).

Water hyacinth has become a danger to many aquatic habitats worldwide. It conquers vast areas, invades neighboring plant species, and reproduces quickly, creating dense, free-floating mats that fully cover freshwater surfaces. Infestation has several detrimental ecological and social effects that endanger aquatic systems, prevent people from using surface waters, and impair hydraulic structures such as canals and pumping places (Gerardo and de Lima, 2022a). It obstructs navigation and fishing (Hill and Coetzee, 2008), harms infrastructure, raises service costs, weakens riparian populations, and changes ecology (Laranjeira and Nadais, 2008; Julien, 2008).

To control and manage weeds, water hyacinth distribution modeling is crucial. Geospatial methods are effective in modeling the distribution of water hyacinth and provide important management information. Geographic Information System (GIS) and remote sensing offer useful tools for mapping and predicting possible habitat and geographic dispersion. To identify and assess the abundance and scope of infestation, remote sensing can acquire multispectral images of water bodies at various spatial and temporal resolutions (Wanyonyi, 2019). GIS can integrate remote sensing data with additional biophysical and environmental factors, such as climate, topography, hydrology, and land use, to develop spatial models that forecast the suitability and risk of water hyacinth invasion. Modeling with geospatial techniques provides large-scale, synoptic, and objective information on water hyacinth occurrence and abundance (Keebine, 2019). The model allows us to consider varieties of elements, such as temperature, precipitation, elevation, and human activities that affect growth and diffusion. Additionally, they can make it possible to analyze various management strategies and evaluate climate change effects on the dispersal of water hyacinth (Thakuri et al., 2019; Mukarugwiro et al., 2019).

1.2 Statement of the problem

Water hyacinth, which displaces native aquatic plant and animal groups, is regarded as the most severe weed in freshwater ecosystems throughout the world's warm and tropical regions (Rezene, 2005). Ethiopia currently hosts 35 invasive alien plant species (Taswe and Wendimagegnehu, 2022). The main danger to biodiversity loss among those species is water hyacinth. Roadsides, urban green spaces, waterbodies, wetlands, different vegetation types, and range lands are all in danger (Shiferaw et al., 2018). It adversely affects the ecosystem and human health. It affects the production of fish, crops, and cattle as well as electric power, irrigation, transportation on waterways, tourism, and human health (Dechassa and Abate, 2020). According to Enyew et al.

(2020), water hyacinth has impacted the Lake Tana vital resources, which are essential to locals' livelihoods. The weed mats hindered farming operations and reduced crop production by deteriorating crop fields. It significantly impacts fishing activities by blocking fishing entry points and fishing communities by deteriorating fish populations. The weed affects cattle through its negative impacts on many species of tasty grass by outcompeting and invading the area.

In Lake Tana, the coverage of water hyacinth has grown rapidly (Cai et al., 2023). It covered between 80 and 100 hectares (ha) during the time of its inception in 2011. After a year, it covers approximately 20,000 ha (Tewabe, 2015). Even though significant efforts such as time and money have been invested by stakeholders each year, its coverage has increased from time to time (Enyew et al., 2020). The weed grows beyond its original region of coverage toward the northeastern direction of the lake region. In particular, weeds are expanding and moving toward the lake at a far faster rate than other coverage types, which is hazardous for the lake's sustainability and the riparian community, whose livelihood depends on lake resources (Asmare et al., 2020). It has impacted fishing activities, including the livelihood of more than 5400 fishers (Asmare, 2017), water quality and biotic communities (Abera et al., 2017; Gezie et al., 2018), pasture and agricultural land use (Dechassa and Abate, 2020; Damtie and Mengistu, 2022), crop production, human health and livestock (Enyew et al., 2020), and transport and tourism (Tewabe et al., 2017). Despite the time and resources invested over many years to reduce the rapid expansion and its adverse impact, it remains the most harmful aquatic weed in Lake Tana and its environs.

For a variety of reasons relating to water resource management, early identification and detection of the spatiotemporal spread of weeds in domestic hydrologic systems are essential (Thamaga and Dube, 2018b). Effective and efficient monitoring largely rests on the accurate detection of the spatial distribution of weeds. Since water hyacinth has the potential to impact biodiversity, ecosystem function, and services, it is imperative that it be dealt with in aquatic habitats (Robles et al., 2015; Pádua et al., 2022). Therefore, it is essential to have up-to-date spatially explicit data and information about its dispersion and early identification to comprehend its spatial arrangement and propagation rate (Thamaga and Dube, 2019).

The spatiotemporal distribution of water hyacinth in Lake Tana has been determined by previous studies using a variety of classification techniques, including the maximum likelihood classifier (Asmare et al., 2020; Worqlul et al., 2020; Damtie et al., 2021), decision tree (Asmare et al., 2020), random forest (Dersseh et al., 2020; Bayable et al., 2023) and SVM (Bayable et al., 2023). However, most of the studies have only used spectral bands, indices, and satellite imagery that are optical as a data source and classification-based machine learning models that are highly affected by atmospheric conditions such as clouds (Mukarugwiro et al., 2019; Bayable et al., 2023). In addition, optical data have limitations in capturing backscattering differences between water hyacinth and other vegetation. Contrary to optical data, radar data have the advantage of capturing images of water hyacinth that are cloud-free, can operate in any time of the day or night regardless of the solar illumination, have high penetration capacity to submerged aquatic plants and can detect the backscattering differences between water hyacinth and other plants of the aquatic environment (Datta et al., 2021; Simpson et al., 2022). Moreover, previous studies rely entirely on remote sensing data sources. However, the detection of water hyacinth is more precise and reliable when several environmental variables, including temperature and rainfall, are combined with remote sensing variables (Thamaga and Dube, 2018a). Hence, to the knowledge of the researcher, no study has been conducted on the detection of water hyacinth using regression-based machine learning models that integrate multiple data sources. Therefore, this study aimed to model the spatial dispersal of water hyacinth in Lake Tana by combining Sentinel-1 SAR and Sentinel-2A optical data and climatic variables (precipitation and temperature).

1.3 Objective of the study

1.3.1. General objective

The main objective of the study is to model the spatial distribution of water hyacinth in Lake Tana.

1.3.2. Specific objectives

The study has the following specific objectives:

1. To model the spatial distribution of water hyacinth using machine learning models with the integration of remote sensing indices, bands and bioclimate variables in the study area
2. To examine the spatiotemporal dynamics of water hyacinth from 2016-2022 using a random forest classification algorithm in the study area
3. To assess the effectiveness of the water hyacinth management system in Lake Tana

1.4 Research questions

The leading research questions for this research include the following:

1. How can machine learning models be utilized to model the spatial distribution of water hyacinth by integrating remote sensing indices, bands, and bioclimate variables?
2. What are the spatiotemporal dynamics of water hyacinth in Lake Tana from 2016-2022?
3. What are the management practices of water hyacinth in Lake Tana and its vicinity, and what challenges have been faced?

1.5 Scope of the study

The geographic extent of the study area is Lake Tana and its surrounding kebeles. Thematically, the study is delimited to applying regression-based machine learning models (RF, SVM, BRT and Ensemble) and examining the predictive capacity of variables derived from Sentinel-2A, Sentinel-1 SAR, and climate data. Additionally, the spatial and temporal dynamics from 2016 to 2022 and the challenges of water hyacinth management practices are covered. The year 2016-2022 is selected due to the availability of Sentinel imagery. The study was conducted from September 2022 to August 2023.

1.6 Significance of the study

Decisions about species conservation and management heavily rely on knowledge of where they are found; however, this information is frequently vague or unreliable. Water hyacinth distribution models can be used to map habitats and can generate reliable, repeatable information that can be used to guide decisions. Before removal and management procedures can be used, it is crucial to locate and map the water hyacinth's spatial distribution to understand its pattern and extent. As a result, the study will have an indispensable role in the identification of the spatial extent and thereby help actors that have priority areas for environmental conservation. In addition, the study will help development actors and NGOs to make appropriate interventions with regard to their area of focus. The local community will obtain awareness and information about the extent of water hyacinth expansion and impediments related to management practices to plan appropriate measures. The study will help novice researchers in the future who will conduct studies on the same and related topics as a source of reference.

1.7 Limitations of the study

The study was limited by the dense mat of the water hyacinth, which hinders navigation to the whole extent of the weed during the data collection period. In addition, since data collection was carried out in April and June 2023, the GCP of the 2022 wet season was not primarily collected from the field by the researcher. To fill this gap, GCP points from unreachable areas and the 2022 wet season were obtained from Amhara Region Lake Tana and other Water Bodies Prevention and Development Agency (ARLTWBPDA). In addition, the study was limited by lack of secondary sources such as report and published materials. Security challenges, in accessibility of some of the study areas were also a limiting factor of the study. Furthermore, the study was limited to employ metrological station data due to the coarse resolution of CHIRPS data.

1.8 Organization of thesis

The document is arranged in 5 chapters. Chapter one focuses on the introductory aspect, which comprises the background of the study, problem statement, objectives, research questions, scope, significance, and limitations of the study. In the 2nd chapter, literature that has thematical, empirical and theoretical connections with the topic of the study was critically reviewed. The third chapter covers the methodology section. It incorporates the study area description and specific methods and materials pertinent to the study topic. Chapter four is about the interpretation of data and discussion based on similar work that has been undertaken from different literature sources. The last chapter is allotted for conclusion and recommendation.

CHAPTER TWO

2. REVIEW OF RELATED LITERATURE

2.1. Concepts, history and spread of water hyacinth

Seven different species of water hyacinth have been identified, and they all belong to the *Pontederia* genus, of which water hyacinth (*Pontederia crassipes*) is widespread and common around the globe (Ghoussein et al., 2023). It is a member of the *Pontederiaceae*, sometimes known as the *Pickerelweed* family, and it breeds a large number of persistent seeds. The plant is a macrophyte and floats freely in water. It produces thick floating mats (Nang'alelwa, 2008). It is perennial, erect, and stoloniferous. It has buoyant leaves that may change in size according to the growth environment and reach a height of 1 to 1.5 m (Verma et al., 2003). It reproduces both vegetatively and sexually, with the former being more crucial for the plant's quick growth and stolon development (Téllez et al., 2008). The vegetative process of reproduction and growth is quick. In 120 days, two plants could multiply to 120,000, whereas in just 23 days, two parent plants could give birth to 30 children (Ouma et al., 2007).

Water hyacinth can spread through various methods, including the fragmentation of established plants, the resprouting of rhizomes and the germination of seeds. The dispersal of this invasive plant can occur through water-borne seeds, as well as through seeds that adhere to the feathers of birds. Notably, birds have a paramount role in the long-distance dispersion of water hyacinth. However, the most challenging and uncontrollable means of dispersal is through the active transport of the plant by people who, often unaware of its detrimental effects, intentionally introduce it to water bodies. Furthermore, humans inadvertently contribute to the spread of water hyacinth in certain areas by using it as packaging material and boat cushions (Twongo et al., 2009). The way that water hyacinth spreads from one place to another differs depending on whether it is accidentally or actively disseminated by people. It has moved from one continent to another, from one nation to another, and from one body of water to another, primarily by human activity (Cilliers et al., 2003; Téllez et al., 2008).

Water hyacinth, one of the most effective plant invaders, has spread from its native Brazil to at least fifty different countries in the last century (Verma et al., 2003). This species first appeared in records in Brazil in 1816. C. Von Martius, who was a German naturalist, researched the Brazilian flora and found the species in 1823. *Pontederia crassipes* is the name he gave it. Solms added it to the 1829 Kuntz description of the *Eichhornia* genus sixty years later (Téllez et al., 2008). The late 1800s saw the introduction of this plant as an ornamental species to North America, followed by its introduction to Africa at the beginning of the 1900s and to Europe in the 1930s. Currently, it has become a native of more than 50 nations in Africa, Central America, Asia, Australia, and New Zealand (Yan et al., 2017).

The entry of species into Africa and Asia occurs through humans. Humans are the primary force behind the species' spread throughout Africa. Because of the beauty of its leaves and blossoms, it is a species with great ornamental value that is used in gardens and attracts people. However, it is one of the 100 most invasive species that are harmful according to the IUCN. Water hyacinth is responsible for the majority of problems due to its rapid growth, competitiveness, and simple proliferation (Téllez et al., 2008).

Water hyacinth has periodically entered different nations throughout Africa. For instance, the majority of invasions in West Africa have taken place since the early 1980s. The first reports of the weed appeared in Benin in 1980–1988, followed by Ghana and Nigeria in 1984, Mali in the 1990s, Burkina Faso in 1991, and Niger in 1990–1994. It was first discovered in Egypt in the late 1880s, but it was not until 1932 that the potential threat to water supplies was recognized. It was first noticed in South Africa in 1910, but it was not labeled as toxic until 1983 (Phiri and Navarro, 2000; Hill and Coetzee, 2008).

Since it was first discovered in Zimbabwe in 1937, it has been a common plant in southern and eastern Africa. It continued to invade significant bodies of water, including the Incomati River in Mozambique in 1946, the Zambezi River and a few significant rivers in Ethiopia in 1956, rivers in Rwanda and Burundi in the late 1950s, the Pangani and Kafue rivers in Tanzania, the Shire River in Malawi, Lake Naivasha in Kenya, and rivers in Rwanda and Burundi in the 1960s. The lakes of Kyoga in Uganda, Victoria in 1989, Malawi-Nyasa in 1996, and Tanganyika in 1997 are where infestations were most recently documented (Phiri and Navarro, 2000).

2.2.Biology of water hyacinth

The biology of weeds includes their morphology, growth, the chemical makeup of their various parts, phenology, vegetative and sexual reproduction, and how they interact with their environment in terms of factors such as temperature, light, pH, dissolved oxygen (DO), and salinity (Yan et al., 2017).

2.2.1. Morphology and environment

Water hyacinth morphology is described as an environmental adaptation, highlighting a species-specific strategy for competitive success in a particular habitat and a propensity to take over ecosystems. To make the most of the available light and nutrients, the plant grows horizontally on the water's surface as space permits, giving rise to multiple new generations. When there is not enough space for them to spread out, water hyacinths often grow vertically with long floats and broad blades to store nutrients and energy for future growth and sexual reproduction. The ball-type floats develop into elongated floats in high-density mats (Zhang and Guo, 2017).

The glossy, broad, thick, and egg-shaped leaves are 10-15 cm wide. It can grow up to 1 meter, but its typical height is between 20 and 30 cm. The roots are purple-black, freely dangling, lengthy treads that are 2.54 cm under the water's surface. It is always found in colonies since it cannot survive on its own. The weed's petals contain dots that are both purple and yellow. Under ideal circumstances, the coverage can be up to 25 kilograms per square meter (400 tons per ha), 500,000 plants per ha, 400 seeds per flower, and 5000 seeds per plant. The seeds can survive for up to 15 years in dirt, water, and silt (Masifwa et al., 2001; Gaikwad and Gavande, 2017).

Both sexual (through seeds) and asexual (through vegetative) reproduction are possible in water hyacinths. Both reproduction methods offer a greater potential for production in a short time frame. For example, through vegetative reproduction, 3,418,800 plants and 43 daughter clusters of leaves can be created in approximately 200 days (Gaikwad and Gavande, 2017). The weed reproduces sexually by producing seeds from its flowers with the help of insects. A single plant can produce flowers after 26 days. Many other methods, including human and bird legs, can be used to distribute seeds (Dersseh et al., 2019a).

2.2.2. Water content and nutrients

The biomass of freshwater hyacinths contain between 90% and 95% water, depending on the growth stage and environmental factors. Compared to new and younger shoots, mature and older plants have a reduced water content. Due to delayed growth and a significant accumulation of carbohydrates in tissue, a poor nutritional situation results in a decrease in water content. The ability of water hyacinth to hyperaccumulate nutrients in various plant portions, which results in a significant variety in its chemical makeup, is another intriguing biological trait (Zhang and Guo, 2017).

2.2.3. Habitat

Water Hyacinth colonizes slow-moving or still water, forming thick, broad mats. It can be found in wetlands, lakes, waterways, and estuary ecosystems. It can withstand extreme fluctuations in water level, periodic flow changes, fluctuated nutrient supply, pH, temperature, and toxin levels, but it cannot withstand brackish or saltwater (Laranjeira and Nadais, 2008). As a free-floating macrophyte, water hyacinths generate two distinct canopies at the air-water interface: leaf canopies are made up of above-water structures, while root canopies are made up of underwater structures. Mature weeds comprise roots, rhizomes, leaves, inflorescences, and fruit clusters (Abera et al., 2017).

2.2.4. Temperature and vegetative reproduction

The growth and development of water hyacinths can be significantly influenced by water and air temperature. Perennial in nature, water hyacinths thrive in tropical and subtropical climates where they can grow and flourish all year long. Depending on the temperature and the quantity of nutrients in a certain area, it can grow well in the temperate climate zone during spring, summer, and autumn. However, June, July, and August are the most conducive to growth. Water hyacinth infestation is considerable from September to December, while it is least prevalent from January to May (Yitbarek et al., 2019).

A single plant generated 19 new ramets over 15 days at an average air temperature of 24.3°C, according to the vegetative reproduction rate, with no space restrictions. However, the vegetative reproduction pattern of water hyacinth would be entirely different under restricted growth conditions or with an increase in population density, as the reproduction rate slowed down gradually with an increase in population density (space limitation intensified) (Zhang and Guo, 2017).

The most conducive environments for water hyacinth growth are nutrient-rich water, temperatures between 28°C and 30°C, a pH value between 6.5 and 8.5, and a salinity of less than 2%. Nitrogen, potassium, phosphorous, and sulfates are favorite nutrients. Within 5 to 15 days, the invasive plant doubles in size (Yan et al., 2017; Dersseh et al., 2019a). The weed has a mean leaf length of 9.3 cm, petiole length of 28.1 cm, leaf blade area of 55.6 cm^2 and leaf area index of 11343.9 cm^2 (Damtie, 2022).

2.3. Drivers of expansion

Water hyacinth may thrive in a variety of habitats. Water hyacinths reproduce quickly due to two main factors: the climate and aquatic environments (Gaikwad and Gavande, 2017).

2.3.1. Bioclimatic drivers

Seasonal climatic variables help to predict the spread and site of water hyacinth. Due to rapid growth rates, the mobility of plant mats, and weather, the area and density of aquatic plant infestations can alter drastically both within a peak period and from year to year (Ouma et al., 2005).

According to Chen et al., (2021), water hyacinth is significantly impacted by bioclimatic elements such as rainfall, wind speed, temperature, sunshine hours, evapotranspiration, and relative humidity. Water hyacinth may grow and spread easily in warm, humid areas with frequent or heavy rain. The speed and area of the infestation are determined by the amount of precipitation and the speed of the wind, respectively.

On the other hand, water hyacinth photosynthesis, seed germination, and autochory require energy and an environment, which are determined by temperature, sunshine hours, and relative humidity. These factors also affect how quickly and widely water hyacinth propagates (Pan et al., 2006). Due to sunlight's role in supplying the energy source for transpiration and photosynthesis and in regulating the activity of numerous enzymes involved in the carbon absorption cycle, sunshine hours show a continuous positive link with water hyacinth growth and distribution. Water hyacinths thrive and spread in large part as a result of temperature. To absorb water and nutrients and carry out synthesis, transformation, and other biological processes, a yearly average temperature of 20 to 25°C is extremely suitable (Chen et al., 2021).

2.3.2. Waterbody drivers

pH is an element that affects water hyacinth growth. It must fall between 6 and 8. pH 7 promotes the most development (in terms of plants and dry weight), with pH values of 3.2 to 4.2 being extremely poisonous to plants, 4.2 to 4.3 being inhibitive, and 4.3 to 4.5 probably inhibitive. The primary driving nutrient is nitrate. In the areas with the most infestation, it ranged from 19.63 to 23.52 mg/L. The range of phosphate values was 0.02 to 3.31 mg/L. (Téllez et al., 2008). A study conducted by Gaikwad and Gavande (2017) shows that chemical fertilizers and pesticides and runoff from cultivated land water hyacinth contain phosphorus (P) and nitrogen (N0). A water hyacinth must have at least 5.5 mg/L N and 1.66 mg/L P to survive. These nutrients may have been introduced to the water by agricultural fields, industrial regions, and populated areas.

On the other hand, the growth of water hyacinth species depends on both variations in water level and water depth. When floating in deep water as opposed to shallow water, the plants develop more roots and grow more frequently. The most suitable circumstances are calm water and shallow water depth (<6 m) (Makhanu, 1997).

2.4.Water hyacinth control and management

Water hyacinth is incredibly hard to eliminate. The majority of handling initiatives aim to reduce financial outlays and ecological disruption. Many nations have made an effort to manage or control water hyacinth (Abera et al., 2017). These methods range from biological control to chemical use to mechanical and material hyacinth removal, but none of them have controlled the threat. The selection of a control mechanism should consider factors such as the size of a region and spatial arrangement, seasonal weather patterns, and the designated functions of the waterbody (Villamagna and Murphy, 2010). Integration and institutional partnership can secure successful management and control efforts at the catchment level. Successful management therefore depends on choosing proper and sound ecological methods suitable for the local context (Van Wyk and Van Wilgen, 2010). The most common and widely applied control and management techniques are discussed below.

2.4.1. Biological control

Reducing weeds by introducing biological agents such as weevils is one of the biological methods used to control water hyacinth (Greenfield et al., 2006). Arthropods and diseases are applied to the affected areas as part of biological approaches. Arthropods feed on water hyacinth leaves, but infections cause the plant to become ill and eventually succumb to bacterial decomposition. The weevil, pathogen, bacteria, fungus, and virus groups are the most prevalent and efficient arthropods (Van Wyk and Van Wilgen, 2010; Dersseh et al., 2019a). The time it takes for the plant to die once the weevils are released varies depending on temperature, the plant's nutrient condition, the climate, the hydrology, and the quantity and insect population health. These methods are more effective in large, deep-water bodies. High wind and wave movements in these aquatic bodies diminish plant biomass by sinking and washing out weed mats and leaves (Julien, 2000). Herbivores (such as grass carp or insects), bacterial and fungal diseases, and insects are a few examples of biological control. Water hyacinth can be eliminated by either directly eating it, infecting undesirable plant species, or utilizing resource competition (Barreto et al., 2000; Greenfield et al., 2006). Because it reduces the biomass of water hyacinth and is environmentally beneficial, it is a recommended method. The most frequent biological water hyacinth control strategies are the weevil species *Neochetina eichhorniae* and *Neochetina bruchi* (Center and Dray, 2010).

2.4.2. Physical and mechanical control

The physical approach involves manually removing the plant using hands and hand tools or physically collecting, cutting, and removing it using harvesting machinery (Dersseh et al., 2019a). These techniques, which include rot ovation, weed raking, hand pulling, dredging, channel cleaning, and excavation are used to remove water hyacinth. Different machines are used to carry out each of these actions. Physical approaches entail modifying the nearby plant's habitat's water body to decrease its survival and growth (Greenfield et al., 2006). This is done to prevent water hyacinth from growing in recreational areas, water supplies, pumps, and landing beaches (Mallya et al., 2001). There are many other physical methods, such as the act of being taken away and contained. Plants were removed manually or mechanically. Pulling boats and conveyor belts, grapple buckets hung from shore or boats, and self-propelled collecting equipment that picks up weeds and then discharges their burden onshore via conveyors have all been used in mechanical removal. They are especially helpful when used in conjunction with the installation of booms to keep young vegetation from encroaching into defined regions that are relatively small and constrained (Julien, 2008). Many nations, such as South Africa and China, use manual removal techniques such as hand pulling or pitch forking. The main purpose of this method is to create jobs because it is extremely labor intensive, only works on tiny infestations, and requires much labor (Hill and Coetzee, 2008).

2.4.3. Chemical control

Herbicides and pesticides, both conventional and unconventional, can be applied directly or indirectly to regions that are infected with weeds to remove or slow their growth (Dersseh et al., 2019). According to Tewabe (2015), glyphosate as a foliar spray and copper complexes alone as a foliar spray can be used to control water hyacinth. However, compared to terrestrial systems, aquatic systems have stricter regulations on the use of herbicides. In certain nations with more developed economies, chemical management through the use of specific herbicides such as 2,4-D or glyphosate appears to be a financially viable method. Additionally, the use of substances in drinking water is widely opposed by the general population in many nations. Hence, the application of chemical controls in the management of water hyacinth requires critical consideration.

2.4.4. Integrated control approach

Since every management method has its drawbacks, is ineffective in specific contexts or habitats, or cannot deliver the desired results, the majority of weed problems need an integrated approach. The foundation of integrated strategies is the coordinated application of a variety of control alternatives. The main methods used to do this are biological control, manual removal, and nutrient enrichment management (Mallya et al., 2001). An adaptive management approach should be used to develop how and when they are deployed in the field because every situation is unique. To prevent one action from negatively affecting another, up-to-date decisions are necessary based on information on the weed, biological control agents, herbicide action, and repercussions of eradicating the weed (Julien, 2008). Accordingly, the coordination of various management methods and organizations will bring a better outcome in the endeavor to manage water hyacinth.

2.4.5. Preventive method

This method emphasizes eliminating water hyacinth by addressing its primary causes, such as eutrophication. This method necessitates a cooperative approach to trash management from residential and commercial spaces (Ingwani et al., 2010). According to Tewabe (2015), preventing water hyacinth from entering a freshwater system is the best way to manage it. This can be accomplished by educating the general public about the issues that arise when unwanted aquarium or water garden plants are disposed of in freshwater systems, as well as transferring equipment to another freshwater or inside the lake's region without first thoroughly cleaning it to remove any plant debris from boats, trailers, other water sports equipment, bait buckets, and fishing materials.

2.5. Water hyacinth in Ethiopia

Lake Koka and the Awash River were the first to be infested approximately 60 years ago. The initial introduction was primarily for its ornamental purposes in 1950 close to Aba-Samuel Dam. Since then, Ethiopia has seen widespread water hyacinth invasion in many of its water bodies. The Lake Koka, Lake Tana, Sobate, Baro, Gillo, and Pibor Rivers, Lake Ellen, Lake Abaya, and Lake

Elltoke are the water bodies where water hyacinth is invading (Firehun, 2017). Currently, it is found in low-, mid-, and high-altitude Rift Valley regions. Low-altitude water bodies such as Lake Abaya, Lake Koka, Koka Dam, and irrigation and drainage systems along the Awash River are affected. Lake Ellen and Lake Elltoke and the Aba-Samuel Dam are among the water bodies that are contaminated at mid- and high altitudes, respectively (Firehun, 2017).

2.5.1. Water hyacinth in Tana

On the northern lakeshore in the Achera kebele of Dembya Woreda, close to the Megech River, the first infestation was discovered. The total coverage was approximately 80–100 ha during the first period of infestation (Tewabe, 2015). Approximately 10-15 ha of stationary water hyacinth mats blanketed Lake Tana to their highest extent, according to a 2012 survey, and they were spread out along 60–80% of the shoreline length where they were found. When it was originally discovered, the weed was found only at the northernmost points of the lake. The area is suitable due to turbulence and the absence of spacious, sheltered, and shallow banks. Reports show that extensive mats were floating and drifting into the lake, particularly close to where cattle grazed (Stave et al., 2017). Lemba Arbaytu, Tana Woyna, Jarjar Abanor, and Adisgie Dingie Kebeles, which are next to Lake Tana's coastal areas, have high infection rates. It was estimated that these heavily contaminated areas covered 3000 ha. However, the majority of infested areas were classified as coastlines with a medium, low, or very low infestation. Particularly sparsely infested were the three Kebeles in Libo Kemkem Woreda (Teza Amba, Kab Abo, and Agid Kirehna) (Anteneh et al., 2015).

In 2011, Bahir Dar University declared it to be a harmful and a threat to the lake. The term "*Emboch,"* which comes from the Amharic proverb "*Wuha Biwoktut Emboch*," which translates to "whatever you pound water, remains water," is used by the locals to refer to water hyacinth. It alludes to the difficulty of human eradication and the quick reinfestation of the plant (Dersseh et al., 2019a).

Infestations of water hyacinth are currently widespread and have occasionally expanded quickly in this lake (Belayneh et al., 2022). The lake is highly infested with water hyacinths, unlike other water bodies in Ethiopia. Only the northeast corner of Lake Tana is where weeds are highly concentrated (Damtie et al., 2021). In 2012, it spread widely and covered 15% of the lake's northern side. The estimated area covered by water hyacinth infestation in 2015 was 34,500 ha (3,000 ha dense, 2,500 ha moderate and 29,000 ha scattered). The local community's mass physical eradication campaign and protracted dry season dormancy are the main reasons why thick mat-covered areas decreased from year to year (Asmare, 2017).

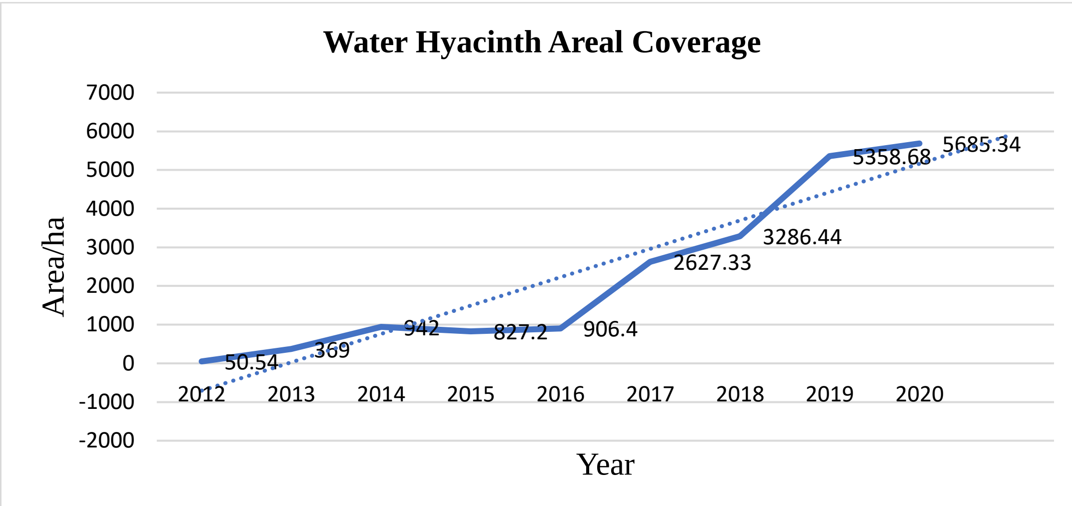


Figure 2.1 Lake Tana water hyacinth area coverage. Source: Bahir Dar University, (2020)

In terms of administrative district, Dembia woreda has a higher concentration of water hyacinth than the other woredas. Agricultural practices are prevalent along this woreda, and crop yields are significantly impacted. Water hyacinth growth is frequently influenced by the local topography. The growth and greenness of this invasive plant are higher when large trees provide shade as well as along lakeshores with shallow depths as opposed to those with deep depths. Around the lake, there is just a minimal amount of water hyacinth present in rocky coastal sections (Ayalew, 2014). The weed is spreading in both time and space, infesting approximately thirty kebeles and 9 woredas (Takusa, West Dembia, East Dembia, Gonder Zuria, Libokemkem, Fogera, Dera, and Bahir Dar Zuria) (Bahir Dar University, 2020).

2.6. GIS and remote sensing for water hyacinth distribution modeling

Invasive aquatic species have long been linked to environmental change brought about by humans, with detrimental implications for ecosystems and humans (Pejchar and Mooney, 2009). Identification of the geographical locations of species is a prerequisite for management and action. Traditional invasive plant detection typically entails extensive field surveys that can be time- and money-consuming. However, remote sensing data, as opposed to traditional field surveys, are currently the main data source for recording and regulating the operation and rate of invasion of water bodies, as well as for spotting potentially exposed locations (Thamaga and Dube, 2018a). In inland hydrological systems, geospatial techniques aid in the early identification and detection of the spatiotemporal distribution, which is crucial for managing a variety of water resources. Remote sensing helps to identify the spectral and/or morphological differences among plants. Maps of invasive aquatic plant species are increasingly being mapped using remote sensing image processing. The resulting distribution maps can be used to model potential invasion risk and target early infestation treatment. The most widespread technique, which often makes use of hyperspectral data, is the remote detection of harmful aquatic plants based on variations in spectral profiles (Bradley, 2014; Thamaga and Dube, 2018a).

Studies on managing water resources have increasingly included remote sensing data from satellite platforms. Indicators are frequently calculated according to the ability provided by this technology to record spectral profiles at relevant spatial and temporal resolutions for a variety of applications. The spectral indices of vegetation and water have been frequently used. To compute indices that are useful in evaluating various characteristics of aquatic plants at the image pixel scale, which is affected by the attribute of reflectance, data from various reflectance bands are obtained from remote sensing sensors, particularly those that depend on the red, green, and NIR wavelength bands (Gerardo and de Lima, 2022a).

Large, statistically sound species distribution data can be produced using remote sensing. By supplying synoptic, spatial, and ecologically relevant predictor factors over a wide geographic area, these technologies have already accelerated improvements in predictive distribution modeling. Remote sensing data such as land use/land cover, habitat classifications, canopy cover, tree height, vegetation morphology, biomass, leaf/wood density, water cover, spectral and textural indices, topography, and climate have been applied widely in species distribution modeling and mapping using remote sensing (Andrew and Ustin, 2009).

In species distribution modeling using geospatial data, two datasets are very important: occurrence and predictors. The response variable (in the form of GCP) data can be collected from atlases or directly by using ground survey methods. Response variables are commonly obtained from geographically interpolated data (e.g., climate, topography, and other remote sensing indices) from different sources, such as WorldClim (Andrew and Ustin, 2009; Bradley, 2014).

With regard to predictor variables, the normalized difference vegetation index (NDVI) is a useful proxy for vegetation analysis and is an example of an index used as an indicator of the habitat suitability of aquatic weeds. Given the recent advancements in remote sensing technology and products, continuous remote sensing metrics are now a crucial component of species distribution investigations and will provide a sizable amount of data for modeling and multiple taxonomic distributions (Buermann et al., 2008).

Satellite systems that are launched into space or airborne either in the form of active or passive sensors contribute to the data that are used to build a species distribution model. Active sensors, such as light detection and ranging (LiDAR) and microwave, as well as passive sensors, such as panchromatic (single-band, high-resolution aerial photography in the grayscale range), multispectral (moderate resolution sensors such as Landsat), and hyperspectral (data from AVIRIS with >100 narrow spectral bands), have been providing data for species distribution mapping and modeling (He et al., 2015).

Broad spatial scale prospective species distributions have frequently been predicted using climate data (New et al., 2002). The datasets obtained from worldwide meteorology stations contain data on temperature, precipitation, solar radiation, and soil moisture (as well as numerous derived bioclimatic integrations of temperature and precipitation). As a result, continuously observed remote sensing climate data are free of interpolation and geographic bias. As a result, climatic predictor factors in species distribution and mapping could be improved by satellite-based measurements of temperature, precipitation, and radiation. For instance, the monitoring of plant species across various locations has made extensive use of MODIS land surface temperature data for terrestrial and aquatic surfaces (He et al., 2015).

Characteristics of vegetation can be useful indicators of suitable habitat, serving as a stand-in for supplies of food or shelter (Buermann et al., 2008). The assessment of the potential effects of climate change on spatial dispersals and extinction threats, which is a significant motivator for the study of species distribution models, has been done using NDVI. Studies on aquatic vegetation, particularly floating and emergent aquatics such as water hyacinth, have utilized the NDVI to show the spatial distribution (Cho et al., 2008). The NDVI, whose signal relates to reflectance in the near-infrared spectral band, can be used to monitor infestations of water hyacinth (Gerardo and de

Lima, 2022a). NDVI helps to assess the vegetation density, morphological structure, leaf area, and chlorophyll amount of water hyacinth (Mucheye et al., 2022).

Through remote sensing, the temporal and spatial dynamics of water hyacinth were identified and quantified. Sentinel-2 imagery with high resolution has demonstrated the seasonal dynamics of water hyacinths in lakes. High-resolution multisensor remote sensing imaging can properly depict the distribution of vegetation over space and time (Mucheye et al., 2022). Sensors that have high spectral (hyperspectral) and spatial (< 10 m) have significantly improved detection capabilities compared to broadband multispectral data. The detection, observation, and mapping of these invasive plants using vegetation indices and water indices, such as NDVI and NDWI, calculated from Sentinel-2 are better than other coarse-resolution imagery (Gerardo and de Lima, 2022a).

Although remote sensing undoubtedly has the potential to identify and map invasive species such as water hyacinth, distinguishing aquatic weeds from other neighboring plants still requires resolutions that range from moderate to high in the visible, NIR and shortwave infrared areas due to the near similar reflectance of plant species found in aquatic habitats (Hestir et al., 2015). Hence, the advancement of hyperspectral sensor technology and the accompanying enhancement of spatial, temporal, spectral and radiometric resolutions enable remote sensing technology to identify fine spectral differences between aquatic vegetation types. Similarly, although alien plants can be spotted and mapped using multispectral remote sensing, weeds are frequently hidden by surrounding green vegetation, which makes them challenging to detect or even map at a fine scale. For effective ecological monitoring in this scenario, sensors with high spatial, multispectral, multitemporal, and radiometric resolutions are required on a larger scale. This will help to better characterize water hyacinths and enhance management techniques in both open and diversified environments (Cheruiyot et al., 2014).

2.6.1. Indices in remote sensing

Vegetation indices are numerical measurements that show how vigorous the vegetation is. They exhibit greater sensitivity for detecting green vegetation than particular spectral bands. Their value is in supporting the interpretation of images, the monitoring of changes in land use, the assessment of biomass density, forestry, vegetation identification, and crop vegetation prediction. Vegetation indices are fairly simple and reliable methods for examining vegetation cover, robustness, and growth dynamics, among other methods (Xue and Su, 2017). Currently, it is a widespread practice to extract aquatic vegetation using remote sensing data in conjunction with several categorization algorithms. However, classification-based methods are ineffective since they must rely on a large number of assessed samples to produce accurate conclusions. As a result, researchers have tried to create a variety of vegetation indicators that are sensitive to aquatic green vegetation. Numerous indices have been promoted for use in identifying and enhancing aquatic vegetation. These include the normalized difference vegetation index (NDVI), modified normalized difference water index (MNDWI), floating leaf vegetation sensitive spectral index (FVSI), planktonic algal vegetation index (floating algal index, FAI), and macro aquatic vegetation index (Singh et al., 2020; Mucheye et al., 2022; Wang et al., 2022). In addition, remote sensing water-based indices such as NDWI are best for separating water bodies from terrestrial and aquatic weeds. It has the potential to monitor aquatic vegetation and is useful for water detection. It also can distinguish between features of aquatic and terrestrial plants, especially in conditions of low vegetation.

2.6.2. Species distribution models (SDMs)

SDMs are used to both forecast a species' geographic distribution and infer its ecological needs. A variety of applications, such as biodiversity monitoring and assessment, conservation activities, and wildlife management and conservation planning, have grown to rely on these models (Naimi et al., 2011). As more people become aware of environmental dynamics and their effects, the application of SDMs to identify and track fauna and flora ranges has grown in significance. The statistical techniques and digitized biological and environmental data used to construct SDMs in a GIS have changed since they were first developed as tools for resource assessment and conservation mapping (Miller, 2010).

To forecast species distributions based on the connection between occurrences and environmental elements, numerous SDMs are available. Each SDM has its strengths and shortcomings. According to Li and Wang (2013), various types of regression models, such as the GLM, GAM, and hierarchical modeling as well as classification models, such as mixture discriminant analysis, GBM, and CART have been extensively utilized in species distribution modeling (SDM). Additionally, complex models such as ANN, random forest, and Maxent approaches have also gained significant popularity in the field of SDM.

Based on a location's environmental circumstances, SDMs are used to determine whether a species is likely to occur. Data on species occurrence and environmental conditions are combined to create

SDMs. In the former, there are geographic data (for instance, coordinates in the form of pairs of longitude and latitude) concerning the location where the species was seen or collected. The latter kind consists of a collection of variables and georeferenced rasters that are known to affect the distribution of species. The set of rasters is often chosen using pre-analysis techniques or expert knowledge (Zhang and Li, 2017).

2.6.2.1. Data in species distribution modeling

A. Biological data

Biological data in SDMs can be nominal (occurrence versus absence or type), ordinal (ordered), and ratio levels (abundance and richness). This affects the modeling techniques that should be used and, ultimately, the statistical level of the SDM result (e.g., probability or suitability of occurrence, type, and expected mean) (Parra et al., 2004; Miller, 2010).

B. Environmental data

To characterize an optimal integration of direct, resource, and indirect gradients, SDM obtains GIS layers of environmental variables. The most common predictors in SDM are climatic and topographical, which characterize biological tolerances linked to water and temperature as well as finer-scale spatial variation in site energy and moisture availability, respectively (Parra et al., 2004).

C. Climate

One of the most significant elements affecting which plants thrive is thought to be the geographic variance in temperature and the amount of moisture and light that are available. In the context of SDM, those variables can encompass both direct factors such as temperature, and resource-related factors such as moisture. These variables can be represented as summary measures, such as annual average precipitation, or more intricate and ecologically significant variables, such as seasonal values and humidity (Parra et al., 2004; Miller, 2010).

Although climate is unquestionably crucial in affecting the distributions of species, GIS bioclimate layers are often created by interpolating ground station data and are thus scarce and of poor quality in many regions (Parra et al., 2004).

D. Topography

Topographic data such as digital elevation models provide the basis for computing more multifaceted topographic variables in SDMs. Topographic variation is typically linked to climate relating to water accessibility, temperature, and solar radiation as predictor variables. In numerous studies, topographic factors such as elevation, slope, and aspect have been extensively exploited. To define a particular mix of soil attributes related to depth, texture, and water retaining capacity, more complicated topographic variables, such as potential solar radiation, topographic moisture index, and landscape location, can be determined (Guisan and Zimmermann, 2000; Rollins et al., 2004).

E. Other

A species' access to moisture and nutrients can also be represented by factors such as geology and soil type. A GIS can calculate more specific variables, such as distance to roads, water, and edges, to characterize closeness to disturbances or significant resources (Osborne et al., 2001). The structure of the habitat, its biophysical characteristics, its pattern, and its heterogeneity can all be described using remotely sensed data and metrics from landscape ecology. The most popular remote sensing outputs, land-cover maps, have enhanced model performance when utilized hierarchically with regional climate. Additionally, remote sensing-based indices such as vegetation, soil and water can also be used as potential predictor variables for various SDMs (Osborne et al., 2001; Gottschalk et al., 2005).

2.6.1.2. Machine learning modes in species distribution models

A. Random Forest

Random forest is a machine-learning approach that is used to predict species distribution. It is a nonparametric, data-driven technique that creates a binary tree using explanatory predictors to characterize the circumstances at sites that are appropriate for a species to occur and those that are not. Using subsets of the environmental predictor factors and species occurrence data that have been randomly chosen (with replacement), random forests create numerous trees. Based on the average of every tree, a final tree is created. Random forest is based on presence and absence data (Hernandez et al., 2008). A group of classification or regression trees are used in random forest to prevent overfitting. The following process is commonly used to implement the model: To create multiple sets of training data, it is necessary to (i) bootstrap the samples from the original dataset;

(ii) create unpruned regression trees using the samples; (iii) select the best split for each node of the tree by arbitrarily selecting a subset of the variables to define the split; and (iv) make predictions by averaging the predictions of the n tree regression trees (Liaw and Wiener, 2002; Luan et al., 2020).

B. Support Vector Machine (SVM)

This system employs the hypothesis space of a linear function in a high-dimensional feature space and is taught with a learning algorithm derived from optimization theory that incorporates a learning bias obtained from statistical learning theory. A mathematical concept, SVM, is a method for maximizing a specific mathematical function with regard to a given set of data. Regression or classification can both be performed using the SVM. In an infinite-dimensional space, SVM constructs a hyperplane or a set of hyperplanes to find the best separation between classes. In higher-dimensional spaces, SVM uses input vectors to build a maximal separation hyperplane. This hyperplane is flanked by two parallel hyperplanes that divide the data. The goal is to maximize the distance between these parallel hyperplanes, as it is believed that a larger margin or separation would result in lower generalization error for the classifier (Noble, 2006; Durgesh and Lekha, 2010; Rodrigues and De la Riva, 2014).

C. Boosted Regression Tree (BRT)

The boosted regression tree (BRT) is a machine learning technique that builds models in a stagewise fashion similar to other boosted models. It starts with simple trees and gradually adds more trees to improve the model fit. Each sequential tree aims to correct the residuals of the previous tree. This boosting process continues until no significant improvements in model fit can be achieved by adding more trees. The resulting additive model has excellent predictive ability due to its ability to represent complex relationships. BRT is useful for species distribution modeling because it can incorporate nonlinear relationships between predictor variables and responses and handle interactions among predictors. This allows BRT to closely fit complex response curves relating environmental factors to species occurrences. BRT has proven useful for modeling the current and future distributions of many plant and animal species. It has been widely used in ecological research and conservation planning (Elith and Leathwick, 2009; Yu et al., 2020; Hallman and Robinson, 2020).

CHAPTER THREE

3. METHODS AND MATERIALS

3.1. Study area description

3.1.1. Location

Lake Tana basin, the source of the Abay River, is situated in the northwestern highlands of Ethiopia at an elevation ranging from 1784-4109 m (Bogale, 2020). In terms of geographical coordinates, it spans from 11° 00' to 12° 40' 00" N and from 36° 40' 00" to 38° 20' 00" E. It is located in Amhara Regional State. The total catchment area of the basin is approximately 15,000 km^2 of which Lake Tana covers approximately 3,060 km^2 (Kumlachew et al., 2023). Lake Tana has an average depth of 8 m and a maximum depth of 14 m (Tibebe et al., 2019). More than 40 rivers feed into the lake, the largest of which are the Gilgel Abay from the south, the Ribb and Gumara from the east, and the Megech River from the north (Enyew et al., 2020). There are 347 Kebeles and 21 Woredas that make up the Lake Tana watershed.

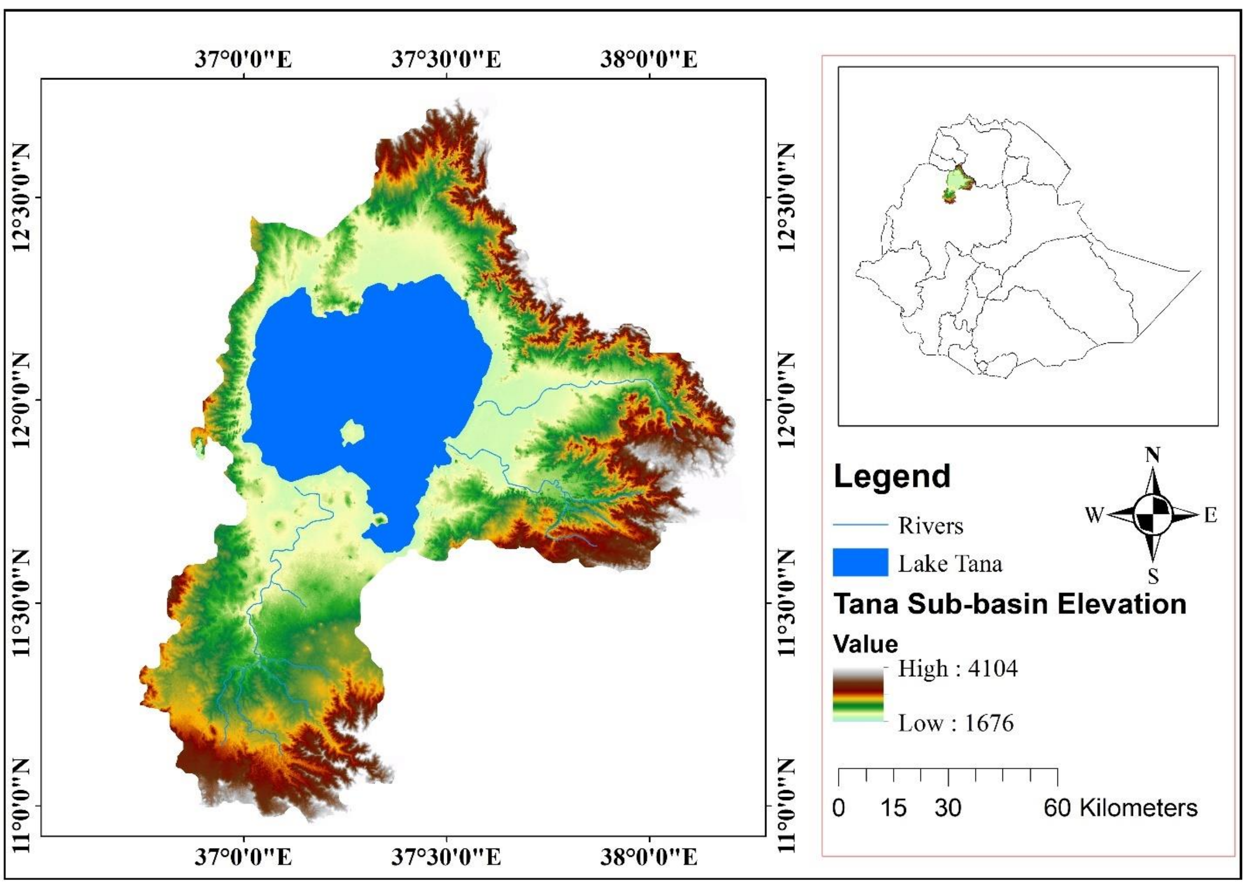


Figure 3.1 Location map of the study area.

3.1.2. Climate and topography

The region experiences a tropical highland monsoon climate. The main rainy season occurs from June to September, accounting for approximately 70-90% of the total annual rainfall. The basin experiences a mean annual rainfall of approximately 816 mm-1200 mm (Mequanent et al., 2021). The average air temperature in the basin throughout the year is approximately 19°C- 23°C (Worqlul et al., 2020). In addition, Mt. Choke and Mt. Guna are the highest points found in the region. Conversely, the western side of the basin exhibits a steep drop toward the neighboring Beles and Dinder basins (Abebe et al., 2017).

3.1.3. Demography and socio-economic setting

The study area has a total population of 3103231, with 1563276 males and 1539955 females (Demissie and Abebe, 2017). This represents approximately 15.8% of the overall population of the region. The majority of the population, 75.8%, of the region lives in rural areas, while 24.2% is densely settled in urban areas. The average population density is 292 people per km^2 (Berihun, 2017).

Lake Tana serves as a multipurpose water resource that supports the livelihoods of millions of individuals in the region. The area surrounding the lake is known for its fertile soil, which has facilitated productive agricultural practices and vibrant fishing activities. Approximately 80% of the population relies on agriculture and engages in subsistence smallholder farming that encompasses a diverse range of crops, such as sorghum, millet, rice, and maize (Abebe and Minale, 2017). Aside from agriculture, the Lake Tana region also benefits from other sectors, such as tourism, fishing, livestock breeding, and small-scale manufacturing and marketing enterprises (Birara et al., 2018). In addition to its agricultural significance, the lake serves as a transportation hub. It is worth mentioning that Lake Tana is home to 37 islands, each with its unique charm, housing intriguing churches and monasteries. Some of these religious sites have a rich history dating back to the 13th and 14th centuries. It is the most prominent tourist destination in Ethiopia. In addition to its appeal to tourists, the region is also renowned for its significant fish resources, boasting a remarkable variety of over 67 fish species. Notably, 70% of these fish species are exclusive to the region, adding to their ecological significance (Anteneh et al., 2015; Stave et al., 2017).

3.2. Research design and approach

The research design provides guidelines for data gathering, measurement, and analysis. This study demands an accurate description of the distribution, spatiotemporal dynamics, and management of water hyacinth. Hence, the study employs a descriptive research design. According to Kothari (2012), the who, what, when, where, and how questions related to a particular research subject can be addressed with the help of descriptive research designs.

On the other hand, a mixed research approach (qualitative and quantitative) was applied. The mixed research approach helps the researcher to integrate various data about water hyacinth distributions and management practices. Additionally, based on this approach, the researcher collected and analyzed both types of data at the same time.

3.3. Data sources, types and acquisition

3.3.1. Data source

Accurate measurement of water hyacinth distribution and assessment of management practices requires diverse, triangulated, up-to-date and reliable data sources. To this end, the researcher carefully explores a variety of data that are pertinent to the study topic. Therefore, both primary and secondary sources were employed. The primary data were gathered through key informant interviews from local farmers, development agents, and ARLTWBPDA that function in relation to Lake Tana water hyacinth. In addition, validation and training ground control points (GCPs) of water hyacinth and other land use data were obtained as primary data to train machine learning models and classify land use/land cover.

Secondary data were collected via intense desk review from both published and unpublished institutional reports, books, peer-reviewed journals, online data, and digitally created platforms. Satellite imagery data from different freely available sources were obtained. Historical GCPs were collected for 2016, 2018, 2020 and 2022 from the Amara Region Environmental Protection Bureau.

3.3.2. Data type and acquisition

3.3.2.1. Remote sensing data

Species distributions have been frequently mapped using remote sensing data and derived variables such as vegetation and water indices (Feilhauer et al., 2012). To this end, to drive variables related to remote sensing vegetation and water indices, Sentinel-2A and Sentinel-1 SAR imagery were utilized. The Sentinel-2A image collection has been used to accurately identify water hyacinth (Singh et al., 2020; Gerardo and de Lima, 2022b). The blue, red, red-edge, NIR, SWIR1, and SWIR2 spectral areas of the Sentinel-2A satellite dataset are crucial and excellent for identifying and mapping water hyacinths (Zhang et al., 2017; Thamaga and Dube, 2018a). In addition, it has a large number of bands, better temporal resolution, and higher spatial resolution. Therefore, Sentinel-2A was employed. Sentinel-2A satellite images were freely downloaded from Sentinel Hub (https://scihub.copernicus.eu/) with cloud cover less than 10%. The image was downloaded for the 2022 wet season (June, July, August and September) and 2023 dry season (January, February, March, April and May). This is due to the variation in geographic infestation coverage across seasons. In the wet season, the geographic coverage is large, and during the dry season, it decreases (Thamaga and Dube, 2019; Bayable et al., 2023). On the other hand, to address the second objective (spatiotemporal dynamics), Sentinel-2A images for 2016, 2018, 2020 and 2022 were employed with the same acquisition date specified above. The choice of those years is related to the availability of sentinel imagery.

Table 3.1 Sentinel-2A Band Designation

Band	Central Wavelength (nm)	Resolution (m)	Description
B1	443	60	Coastal aerosol
B2	490	10	Blue
B3	560	10	Green
B4	665	10	Red
B5	705	20	Red edge 1
B6	740	20	Red edge 2
B7	783	20	Red edge 3
B8	842	10	Near-infrared
B8A	865	20	Red edge 4
B9	940	60	Water vapor
B10	1375	60	Cirrus
B11	1610	20	Shortwave infrared 1
B12	2190	20	Shortwave infrared 2

Source: Zhang et al., (2017)

From the Sentinel-2A satellite, 10 vegetation and water indices were derived as covariates for modeling. According to Mund et al., (2015), vegetation cover and floating biomass can be detected using a variety of remote sensing indices, such as NDVI, NDWI, SAVI, and MSAVI. Accordingly, the following vegetation indices were selected based on literature sources. However, for the second objective of the study (spatiotemporal distribution), an additional normalized difference built-up index (NDBI) and bare soil index (BSI) are added.

Table 3.2 Remote sensing predictor variables (vegetation and water indices)

Index	Formula	Description	Source
Normalized Difference Vegetation Index (NDVI)	$\frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}}$	It quantifies the level of vegetation vitality by analyzing its spectral reflection in the red and near-infrared range, which indicates the extent of greenness.	Rouse et al., (1974)
Normalized Difference Water Index (NDWI)	$\frac{\mathrm{GREEN} - \mathrm{NIR}}{\mathrm{GREEN} + \mathrm{NIR}}$	The water content of vegetation is quantified by analyzing its spectral reflectance in the green and near-infrared range.	Gao, (1995)
Enhanced Vegetation Index (EVI)	$2.5 \times \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + 6 \times \mathrm{RED} - 7.5 \times \mathrm{BLUE} + 1}$	Similar to NDVI, this method minimizes the impact of atmospheric factors and background interference, resulting in improved sensitivity in regions with high biomass.	Huete et al., (2002)
Soil Adjusted Vegetation Index (SAVI)	$(1 + \mathrm{L}) \times \frac{\mathrm{NIR}-\mathrm{RED}}{\mathrm{NIR}+\mathrm{RED}+\mathrm{L}}$ where L= 0.5	Similar to NDVI but includes a soil adjustment factor L to reduce the effects of soil background. L is a constant that varies depending on soil brightness, usually between 0 and 1.	Huete, (1988)
Modified Soil Adjusted Vegetation Index (MSAVI)	$\frac{(2 \times \mathrm{NIR} + 1) - \sqrt{(2 \times \mathrm{NIR} + 1)^2 - 8 \times (\mathrm{NIR} - \mathrm{RED})}}{2}$	Similar to SAVI but uses an empirical formula to estimate L dynamically from NIR and RED bands. It reduces soil noise and increases the dynamic range of vegetation index values.	Qi et al., (1994)
Modified Normalized Difference Water Index (MNDWI)	$\frac{\mathrm{Green} - \mathrm{SWIR1}}{\mathrm{Green} + \mathrm{SWIR1}}$	Similar to NDWI but uses shortwave infrared instead of near infrared to enhance the water signal and suppress the built-up land signal.	Xu, (2006)

Green Normalized Difference Vegetation Index (GNDVI)	$\frac{\mathrm{NIR} - \mathrm{Green}}{\mathrm{NIR} + \mathrm{Green}}$	Helpful for determining the variability of leaf chlorophyll when the leaf area index is somewhat high and used to identify aquatic plants	Gerardo and de Lima, (2022b)
Water Adjusted Vegetation Index (WAVI)	(1 + L) (NIR-Blue)/(NIR+ Blue + L) where L=0.5	Can improve the ability to distinguish between features of aquatic and terrestrial plants, especially in conditions of low vegetation.	Villa et al., (2013)
Red-edge Normalized Difference Vegetation Index (RENDVI)	$\frac{\mathrm{NIR} - \mathrm{RE}}{\mathrm{NIR} + \mathrm{RE}}$	This method, akin to NDVI, utilizes the red-edge band instead of the red band to capture the significant rise in reflectance near the red edge of chlorophyll absorption. It exhibits greater sensitivity to chlorophyll content and canopy structure compared to NDVI.	Gitelson et al., (2003)
Normalized Difference Aquatic Vegetation Index (NDAVI)	$\frac{\mathrm{NIR} - \mathrm{BLUE}}{\mathrm{NIR} + \mathrm{BLUE}}$	Utilizing the visible shortwave range and the near infrared range, which are sensitive to the presence of water in vegetation and can distinguish between submerged and floating aquatic plants.	Villa et al., (2013)

In addition to optical sensors, according to Simpson et al., (2022), the distribution of aquatic vegetation can be identified using Sentinel-1 SAR image data. Sentinel-1 SAR is a satellite-mounted sensor that can capture C-band images of the Earth's surface with various polarizations (VV and VH). SAR data are excellent for monitoring water hyacinth in areas with regular cloud cover since they can pass through clouds and function in all-weather situations. As a result, a total of 2 bands (VV and VH) were used as predictive variables (Table 3.4).

Table 3.3 Sentinel-1 SAR satellite characteristics

Platform	Sensor	Spatial Resolution	Wavelength range	Temporal Resolution	Band
Sentinel-1A	Imaging Radars	10 m	C band =5.4 cm	12 days	Two band (VV and VH)

Source: (Abdikan et al., 2016)

Table 3.4 Sentine-1A SAR bands and indices

Index	Formula	Description	Source
VV	$VV = \text{average}(VV_{\text{annual}})$	VV polarization can be used to differentiate between water and vegetation since it is sensitive to surface roughness and moisture.	Jiang et al., 2021
VH	$VH = \text{average}(VH_{\text{annual}})$	VH polarization can be used to identify various vegetation types since it is sensitive to volume scattering and double-bounce scattering.	Jiang et al., 2021

With regard to the detection of spatiotemporal dynamics of water hyacinth, a total of 16 variables were selected. The variables that are selected to map the distribution include indices, bands and topography. Except for topographic variables (elevation and slope), the others were derived from Sentinel-2A and Sentinel-1 SAR, which were adopted for species distribution modeling. The slope and elevation data were downloaded from SRTM (https://www.earthdata.nasa.gov) with a spatial resolution of 30 m.

Table 3.5 List of variables for spatiotemporal change detection

Data Source	Variable
Sentinel-2A bands	B4, B5, B8, B11 and B12
Sentinel-2A indices	SAVI, EVI, NDWI, NDAVI, NDBI, NDVI and BSI
Sentinel-1 SAR Bands	VV and VH
SRTM	Elevation and Slope

Land Use/Land Cover

The main target of this study is detecting the spatiotemporal dynamics of water hyacinth in Lake Tana. Therefore, the study classifies the land use/land cover of the Lake Tana region into three groups, namely, water body, water hyacinth and mixed. The mixed class is employed to train the model and accurately distinguish water hyacinth from riparian green vegetations that have similar spectral reflectance.

Table 3.6 Land use/land cover category and description

LULC type	Description
Water Body	Areas covered by perennial rivers, lakes, ponds, reservoirs
Mixed	Areas covered by bushes and small trees, forest, grasses, shrubs, croplands and built-up
Water hyacinth	Area covered by water hyacinth invasive species

3.3.2.2. Bioclimate data

Bioclimate elements such as precipitation and temperature have a substantial impact on water hyacinth growth (Chen et al., 2021). Temperature and rainfall readings are frequently employed in methodologies for species distribution and associated ecological modeling. The bioclimatic variables encompass various aspects of the environment that impact living organisms. These variables capture the patterns of seasonality, such as fluctuations in temperature and precipitation throughout the year and the presence of extreme or constraining environmental factors. Examples of such factors include the temperature of the coldest and warmest months and the amount of precipitation during the wettest and driest quarters. Additionally, bioclimatic data also provide information about annual trends, such as the average annual temperature and total precipitation.

Accordingly, based on the availability of metrological data and recommendations from the literature sources (Thakuri et al., 2019; Kariyawasam et al., 2021), the researcher employs four seasons of climate data. Namely, the annual mean temperature of the wet and dry seasons and the annual mean precipitation of the wet and dry seasons were used. The meteorological data were collected from seven (Gorgora, Addis Zemen, Bahir Dar, Deke Estifanos, Zenzelma, Dera Hamusit and Zege) stations surrounding the study area for 30 years.

Table 3.7 Bioclimate variables and their description

Variable	Description
Annual Mean Temperature of Dry Season	Mean annual monthly mean air temperatures averaged over 30 years for the dry season
Annual Mean Temperature of Wet Season	Mean annual monthly mean air temperatures averaged over 30 years for the wet season
Annual Mean Precipitation of Dry Season	Mean annual monthly precipitation of dry season averaged over 30 years
Annual Mean Precipitation of Wet Season	Mean annual monthly precipitation of wet season averaged over 30 years

3.3.2.3. Field data

To answer research questions related to the management of water hyacinth, interviews were conducted with seven key informants. The interview was semi-structured and handled by the researcher. The semi-structured interview was selected due to its advantage of providing open-ended and in-depth responses. The interviewee was purposively selected based on their participation and engagement, and expert level on the management of water hyacinth. The interview was conducted with experts from government organizations, development agents, and elders who had connections to the management and conservation activities in Lake Tana.

On the other hand, 458 presence data of water hyacinth were obtained to train the machine learning models in species distribution modeling. In species distribution modeling, background (absence) data are also important to make the model more accurate. Therefore, pseudoabsence points (458) are created randomly without overlapping with the presence points. From the total GCP (presence) points that were collected from the field and randomly created as pseudoabsence points, 70% were used to train the model. The remaining 30% of the GCP points were used to validate the model. In collecting the GCP, a portable Global Positioning System (Garmin eTrex 10) with 5 m accuracy receivers was used, and data were collected by five field data collectors. The data collectors were trained before the actual data collection.

3.3.3. Software packages

In this study, the following software packages were used.

Table 3.8 Software and its major application area

NO.	Software	Major Application Area
1	Microsoft Excel 2016	For recording GCPs, and statistical computations
2	ArcGIS Pro 2.8	Image preprocessing and mapping
3	Google Earth Engine	For LULC change detection, processing data and statistical computation
4	R Studio 4.3.1	For species distribution modeling, processing data and statistical computation
5	SNAP 7.0	Sentinel image processing

3.4. Data processing

3.4.1. Image preprocessing

Distortions in remotely sensed images are caused by changes in illumination, weather, and sensor noise (Mukarugwiro et al., 2019). As a result, preprocessing helps to enhance image quality and data. Hence, the downloaded Sentinel-2A and Sentinel-1SAR images were projected to Zone 37, World Geodetic System 1984, and Universal Transverse Mercator. Atmospherically, haze removal image preprocessing was carried out. After the images were atmospherically corrected, digital numbers were converted into reflectance (top of atmosphere).

On the other hand, the Sentinel-2A image bands with different spatial resolutions of 10 m, 20 m, and 60 m were resampled to 10 m using the nearest neighbor resampling method. Additionally, climate data collected from metrological stations were resampled to a 10-meter spatial resolution and interpolated using the ordinary kriging method. Kriging interpolation has the advantage of providing statistical reports on the level of accuracy (Dersseh et al., 2019b). The remote sensing image and climatic variables gathered from different sources were clipped to the study area. After clipping, the image was mosaiced to create multilayer raster images.

3.4.2. Image classification

The classification of land use/land cover was based on the pixel-based image classification method. A supervised classification technique was applied. Due to its robustness and accuracy among classification algorithms (Mukarugwiro et al., 2019), a machine learning-based random forest classification was applied.

3.4.3. Model selection for water hyacinth

According to Durgesh and Lekha (2010), Naimi et al., (2011), Li and Wang (2013), and Rezaei and Sengül (2019), models that are commonly applied in species distribution modeling and employ pseudoabsence and presence data include random forest (RF), boosted regression tree (BRT), support vector machine (SVM), and ensemble models (the combination of models). Therefore, as this study depends on occurrence data (presence and pseudoabsence) as well as a large number of predictor variables, the study employed four machine learning models (RF, BRT, SVM, and Ensemble). The research implements a tenfold bootstrap resampling method in RStudio with the "sdm" package. Bootstrap resampling is selected due to its ability to mitigate the issue of small sample sizes, robustness and estimation of uncertainties (Vaughan and Ormerod, 2005). In the ensemble modeling approach, a weighted averaging method was applied.

3.5. Data analysis and interpretation

3.5.1. Remote sensing data analysis and interpretation

The spatial data collected from the field and variables that are used in modeling can be affected by spatial autocorrelation (when variables are collected at close places but their values are not independent of one another) (Elith and Leathwick, 2009) and multicollinearity (where several independent variables in a model are correlated) (Miller, 2010). Therefore, to avoid geographic clustering of sample points, the GCPs were collected at 500-meter intervals (Shiferaw et al., 2019), and Moran's *I* method was applied to test spatial autocorrelation. According to Ghodousi et al., (2020), Moran's *I* is a statistical measure used to assess the spatial pattern exhibited by a set of features and their associated attributes. It helps determine scattered, clustered, or random patterns. The range of values for Moran's *I* is -1 to 1. In this range, a negative value indicates the dispersion of GCPs. A value of 0 suggests random clustering, while a value of positive indicates clustering or grouping. Therefore, values of 0 and negative were accepted, as the data were free from spatial autocorrelation.

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} wi, j^{Z} j^{Z} j}{\sum_{i=1}^{n} Z_i^{\,2}}$$

where *Zi* is the deviation of an attribute for feature *I* from its mean *x–Xi*, *Wi,j* is the spatial weight between feature *i* and *j*, *n* is equal to the total number of features, and *S*0 is the aggregate of all spatial weights:

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}$$

The 1-score for the statistic is computed as:

$$z_1 = \frac{I - E[I]}{\sqrt{V[I]}}$$

$$E[I] = -\frac{1}{n-1}$$

$$V[I] = E[I^2] - E[I]^2$$

On the other hand, to avoid a correlation among predictor/explanatory variables, a multicollinearity test was carried out. Correlation coefficients were computed for each pair of covariates. In this study, variables were dropped from the model when the correlation coefficient r was > 0.8 (Singh et al., 2020).

3.5.2. Model performance and validation

Threshold-based species distribution models have been commonly evaluated through AUC/ROC, TSS, kappa coefficient, sensitivity, and specificity (Wilkinson et al., 2021).

- **AUC/ROC**

The ROC curve is a graphical representation that compares model sensitivity and specificity. Sensitivity is the percentage of accurately identified water hyacinth occurrences, while specificity refers to the percentage of correctly identified non-water hyacinth occurrences. The ROC curve helps assess the model's performance in distinguishing between water hyacinth and non-water hyacinth areas by plotting the trade-off between sensitivity and specificity (Gobeyn et al., 2019).

$$\text{Sensetivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

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

where TP (true positive) is the number of correctly identified water hyacinth occurrences, TN is the number of correctly identified areas without water hyacinth, FP is the number of areas incorrectly classified as having water hyacinth when they are free of it, and FN is the number of water hyacinth occurrences that are missed or erroneously classified as absent. The model produces excellent results when the value is near 1. A value greater than 0.7 denotes the availability of the model. The value is 0.5 when it is randomly distributed (Chen et al., 2021).

- **Kappa Coefficient**

Interrater agreement was measured statistically, excluding coincidental agreements. A confusion matrix was calculated as

$$\text{Kappa} = \frac{\text{TP} + \text{TN} - \frac{[(\text{TP} + \text{FN} * (\text{TP} + \text{FP}) + (\text{FP} + \text{TN}) * (\text{FN} + \text{TN})]}{\text{N}}}{\text{N} - \frac{[(\text{TP} + \text{FN}) * (\text{TP} + \text{FP}) + (\text{FP} + \text{TN}) * (\text{FN} + \text{TN})]}{\text{N}}}$$

where N is the total number of testing samples. Kappa values of > 0.81, 0.61-0.80, and 0.41-0.60 are rated as excellent, good and medium, respectively (Senay et al., 2013).

- **Threshold**

The maximum value at which the specificity intersects with sensitivity. Values over the threshold show the presence of water hyacinth, while values below the threshold show its absence (Chen et al., 2021).

- **True Skill Statistics (TSS)**

The TSS indicates agreement between the observed and predicted values (Somodi et al., 2017). As the TSS result is higher and close to 1, the model is best in comparison to other models.

$\text{TSS} = \text{TPR} + \text{TNR} - 1$

where TPR= TP/TP+FN

TNR=TN/TN+FP

3.5.3. Accuracy assessment for land use/land cover classification

Accuracy assessment involves comparing a classified image with another reliable and accurate data source, often referred to as ground truth data. To assess the accuracy of the classified map by machine learning RF classifiers, a separate test dataset was used, and its results were presented in the form of a confusion matrix. The RF machine learning model was trained with 70% of the GCP, and the remaining 30% was used to test the accuracy level. Classification accuracy was assessed using the kappa coefficient, producer, user, and overall accuracy. According to Asmare et al., (2020), for supervised classification, a minimum of 30 points or features per class are needed. Therefore, more than 150 GCPs were used in the post-classification accuracy assessment. According to Bharatkar and Patel (2013), the classification accuracy of an image is calculated through the formula indicated in Appendix 2.

3.5.4. Qualitative Data Analysis

Data that were collected from key informants through semi-structured interviews were recorded, translated and transcribed. After the data were transcribed and translated, they were organized chronologically and thematically. A qualitative data analysis, particularly a narrative method, was applied. In doing so, secondary data that were gathered from the literature and reports were used as an auxiliary to manifest the key informant interview results.

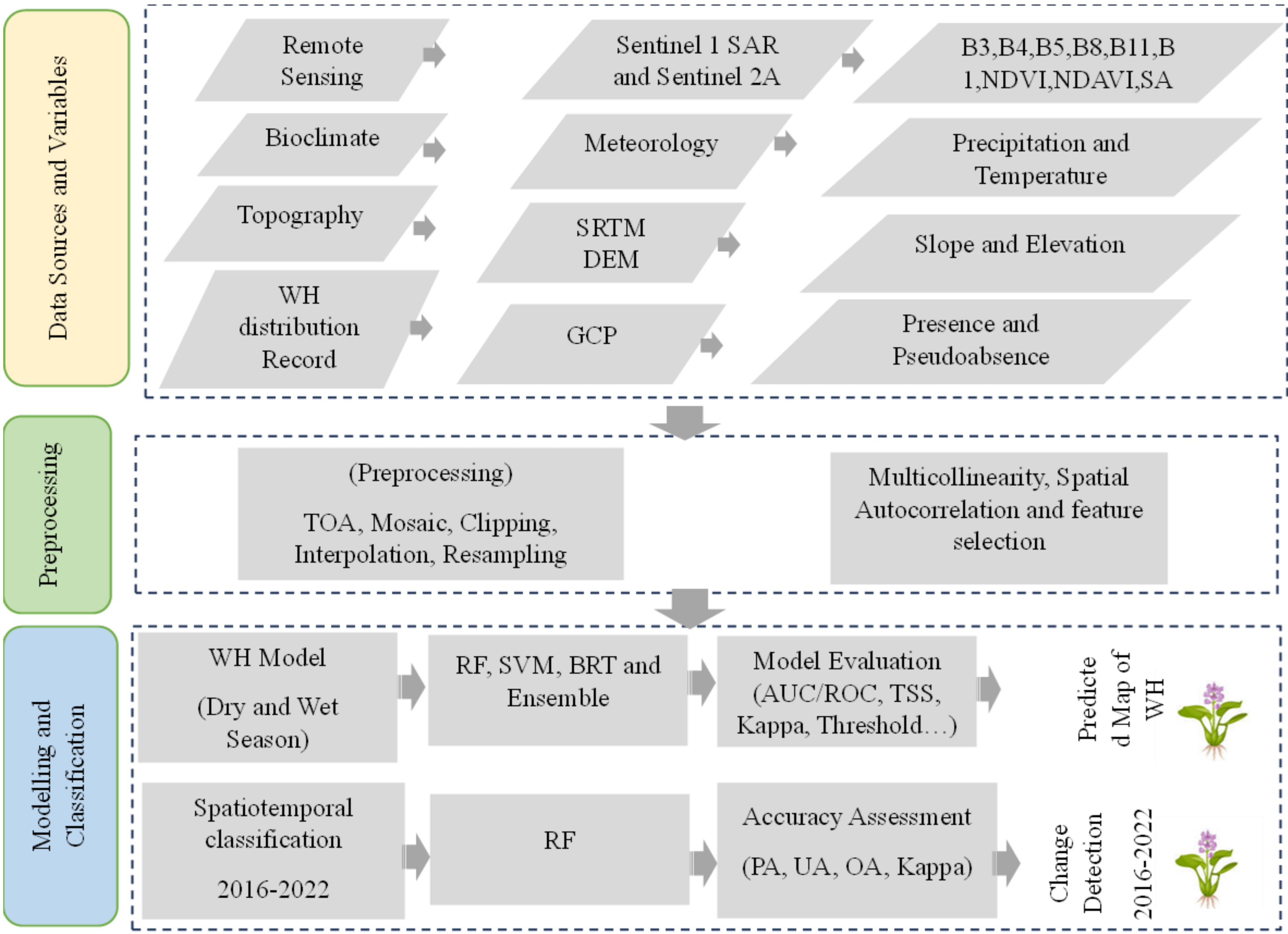


Figure 3.2 Methodology flow chart

CHAPTER FOUR

4. RESULTS AND DISCUSSION

4.1. Spatial autocorrelation and multicollinearity of predictor variable

The results (Table 4.1) indicate that there is positive but close to zero spatial autocorrelation in the dataset (Moran's index 0.011 and z score 1.23), but the strength of this correlation is relatively weak and not statistically significant ($p > 0.05$). Hence, training sample data are randomly distributed and have no significant spatial autocorrelation.

Table 4.1 Moran's I index results

Moran's Index	Variance	Z score	P value
0.011	0.014	1.23	0.22

In addition, a total of 20 predictor variables, including 10 vegetation indices and 6 bands derived from Sentinel-2A and 2 climatic variables and 2 bands from Sentinel-1 SAR, were independently tested based on the dataset to check collinearity and select important variables for modeling. Accordingly, from 20 variables, a total of 11 variables are selected. However, B4, B8, B11, MSAVI, GNDVI, MNDWI, RENDVI, WAVI and EVI were excluded from the list of variables because those variables had shown a greater degree of collinearity or a high correlation coefficient ($r > 0.8$). In addition, independent modeling was performed with excluded variables to check if they had a better probability of prediction, and the result shows that variables that have greater collinearity performed somewhat equally well compared to the selected variables.

4.2. Spatial distribution of water hyacinth using machine learning models

4.2.1. Performances of machine learning model

In this study, RF, SVM, BRT and ensemble machine learning models were used to detect the geographic distribution of water hyacinth. An independent test was performed for two seasons since the geographic coverage proliferates in the wet season (Jun-September) and decreases in the dry season (March-May). Machine learning model performance was tested using AUC, ROC, COR, TSS, Kappa, Sensitivity, and Specificity parameters that have been commonly used in SDM.

4.2.1.1.Performance of species distribution modeling for the wet season

Table 4.2 Performance of models, wet season

Model	AUC	COR	TSS	Deviance	Threshold	Sensitivity	Specificity	Kappa
RF	0.93	0.76	0.77	0.68	0.61	0.87	0.89	0.76
BRT	0.91	0.73	0.75	0.92	0.52	0.86	0.86	0.71
SVM	0.9	0.7	0.71	0.8	0.54	0.9	0.81	0.71
Ensemble	0.9	0.71	0.73	0.89	0.58	0.87	0.85	0.72

Table 4.2 shows the performance of the four models using evaluation parameters. The ensemble model is the combination of the three models (RF, SVM, and BRT) using the weighted average of their prediction. According to the table, all the four models perform reasonably well, with AUC values ranging from 0.9 - 0.93. RF has the highest AUC value of 0.93, while SVM and Ensemble have the lowest AUC value of 0.9.

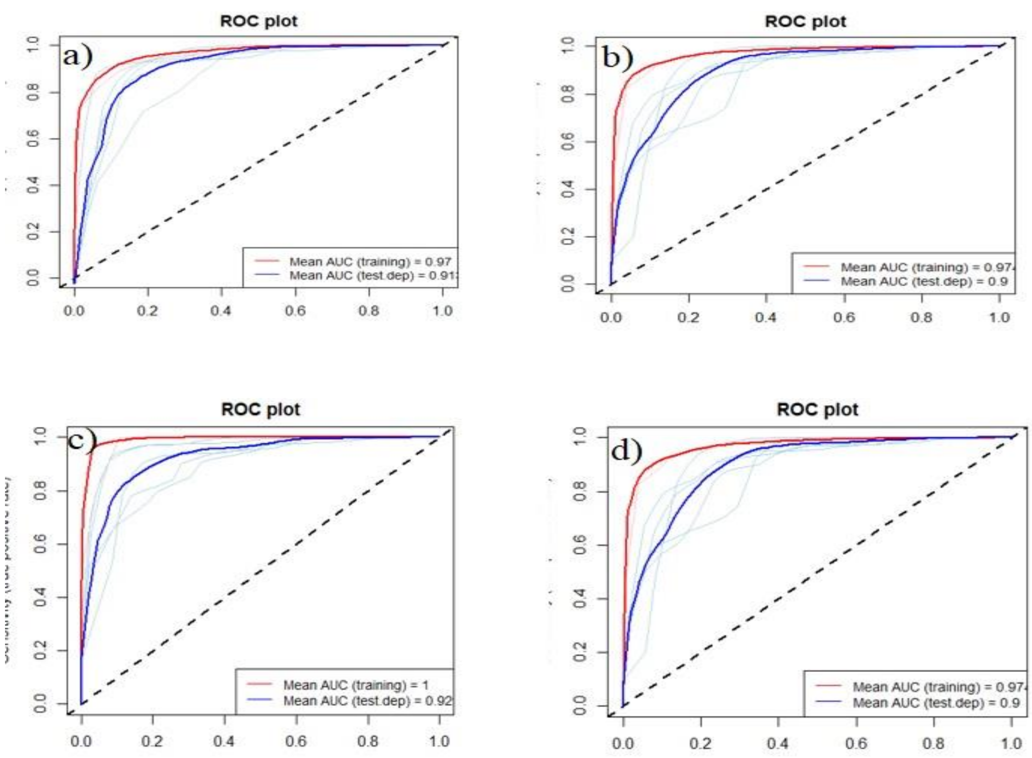


Figure 4.1 Receiver operator characteristic (ROC) curve: a) BRT, b) SVM, c) RF, and d) Ensemble for the wet season

In Figure (4.1) of the ROC curve, AUC values for training and test data are represented by red and blue curves, respectively. The vertical line shows the sensitivity, and the horizontal line shows the specificity. Figure (4.1c) of ROC reveals that RF performs better, with the highest ROC value of 0.93, followed by BRT, with a value of 0.91. The SVM and Ensemble models perform equally well, with a value of 0.9 in the test dataset. Even though the performance of RF is the highest, the performances of the BRT, SVM and Ensemble models are also significant and higher. According to (Chen et al., 2021), a larger ROC value is an indicator of better model fit, and a value close to 1 signifies an excellent performance of the model. An AUC value below 0.7 is an indicator of poor model fit. In a similar study conducted by Graham et al. (2008), an AUC value of > 0.75 was considered acceptable for use in management activities and related planning. As a result, those models that have high AUC/ROC values, and their prediction is within the margin of acceptance.

In terms of the other performance metrics, Ensemble and RF have the lowest deviance values of 0.52 and 0.68, respectively, indicating a good model fit. Meanwhile, BRT (0.92) and Ensemble (0.89) have the highest deviance values, indicating comparatively less fit compared to the above two models. Looking at the sensitivity and specificity values, all models have reasonably high values, ranging from 0.86 - 0.9 for sensitivity and 0.81 - 0.89 for specificity. The identification of the true positive rate is the highest for SVM with a value of 0.9, followed by RF and Ensemble with an equal value of 0.87. On the other hand, true negatives that are correctly predicted by the model from Table 4.2 show that the RF model has the highest (0.89), and SVM discloses the lowest (0.81). Overall, the results of specificity and sensitivity indicate that all models can correctly identify a large proportion of actual positive and negative values. The results are also consistent with Allouche et al., (2006), as the study takes the value of 0.8 sensitivity and specificity in the modeling of species distribution.

Cohen's Kappa coefficient, a measure of agreement between the observed and predicted values that accounts for chance agreement, ranges from -1 to 1. Table 4.2 shows kappa values ranging from 0.71 to 0.76. RF had the highest Kappa value of 0.76, which means that the model performed better and that there was agreement between the predicted and observed water hyacinth. BRT and SVM have comparatively low but equal Kappa values of 0.71. Nevertheless, the Kappa values of BRT and SVM are low compared with those of the RF and Ensemble models, and the performance of those two models in the agreement between the observed and predicted water hyacinth is significant.

TSS helps to evaluate the performance of modeling. Accordingly, based on the table results indicated above, the TSS value ranges from 0.71 to 0.77. The RF model has a greater TSS value of 0.77, indicating higher predictive performance than the other models. However, the SVM model performs relatively less, with a value of 0.71. According to Yoon and Lee (2022), a TSS value > 7.0 is an indicator of better performance of a model. Consequently, despite a variation in the value of TSS, the four models performed well.

A threshold value that is used to classify water hyacinth as absent or present ranges from 0.52-0.61. The RF model has a high threshold value (0.61), followed by Ensemble (0.58), and the lowest is scored by BRT (0.52).

4.2.1.2. Performance of machine learning models for the dry season

Table 4.3 Performance of Models, Dry Season (2023)

Model	AUC	COR	TSS	Deviance	Threshold	Sensitivity	Specificity	Kappa
RF	0.95	0.81	0.82	0.57	0.57	0.90	0.92	0.82
BRT	0.92	0.74	0.73	0.86	0.52	0.96	0.77	0.73
SVM	0.91	0.70	0.71	0.84	0.51	0.90	0.81	0.71
Ensemble	0.92	0.75	0.75	0.77	0.53	0.92	0.83	0.75

Table 4.3 reveals the performance results of the four models for the dry season (2023). The AUC value ranges from 0.91 to 0.95. The RF model achieved the highest AUC value of 0.95, indicating its ability to effectively discriminate between true positives and true negatives. The SVM model achieved a lower AUC value (0.91). The Ensemble model achieved an AUC score of 0.92, which is similar to that of the BRT model and shows a low level of deviance (0.77). The highest COR (0.81) and TSS (0.82) values were also achieved by the RF model, indicating a strong correlation with the data and accurate predictions of both presence and absence.

In terms of sensitivity, the BRT model (0.96) and specificity, the RF model (0.92) outperformed the other models, which portrays that the two models predict the occurrence and absence values more accurately. The Kappa scores for all the models were relatively high and ranged from 0.71 –

0.82. The RF model had the highest Kappa value (0.82). The Kappa performance of SVM is the lowest among the four models, with a value of 0.71. However, this does not mean that the prediction is poor and insignificant. A Kappa value of 0.60-0.8 is frequently used in species distribution modeling, as it shows a good performance level (Senay et al., 2013). On the other hand, the threshold value of classifying water hyacinth as a binary class of absent or present ranged from 0.51-0.57. The RF model has the highest (0.57) threshold level, and SVM has the lowest (0.51). However, there is a very narrow difference between the RF and SVM models (0.06).

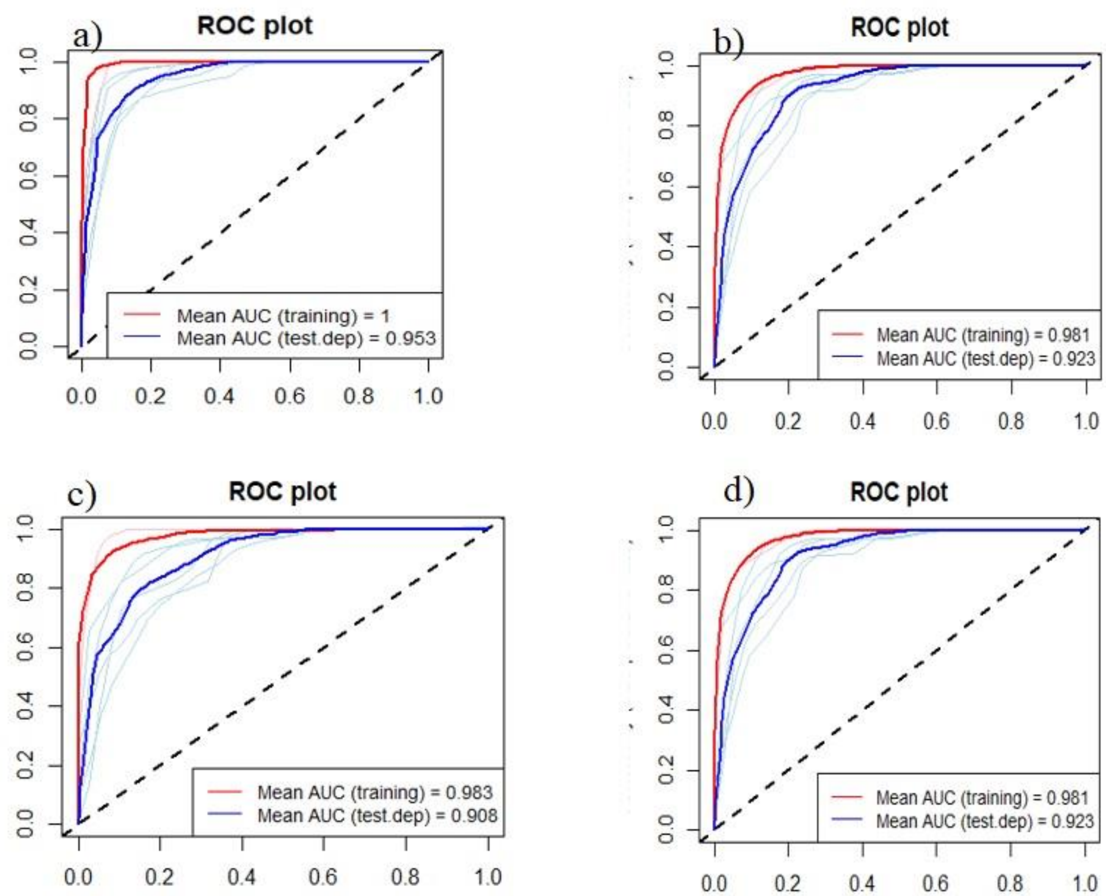


Figure 4.2 Receiver operator characteristic (ROC) curve: a) RF, b) BRT, c) SVM, and d) Ensemble for the dry season (2023)

Figure 4.2 indicates the ROC curve of the four machine learning models using a bootstrap resampling method. The ROC value of the RF model is the highest (0.95), followed by the Ensemble and BRT at equal weight (0.92) in the dry season. The SVM model also performs well, with a value of 0.91. Overall, despite a disparity in the level of performance among models, all of the models performed well using the ROC parameter. The results of this study are also consistent with those of Chen et al., (2021).

Imprecisely, the AUC/ROC, COR and TSS performance of the RF model is the highest and lowest deviance value, and it shows consistency across the two seasons. SVM and BRT performed well in identifying the true positive rate during the wet and dry seasons, respectively. In terms of identifying the true negative rate and kappa, the RF model outperformed the other models in both seasons. Regarding the threshold level, there was no significant difference, and the value ranged from 0.51-0.61 for both seasons. The RF model had the highest threshold that demarcated the occurrence and nonoccurrence of water hyacinth. Overall, the RF model is the best model that shows better performance in all parameters of model evaluation across the two seasons. The results of this study are also consistent with those of Bayable et al., (2023). The study found that the RF model accurately classified water hyacinth in all seasons with 97% overall accuracy and a 0.90 kappa value from Sentinel-2A images.

4.2.2. Predictor variable importance

The study employs a total of 11 variables for two seasons. The variables are broadly grouped into Sentinel-2A bands and vegetation indices and Sentinel-1SAR bands and climatic data. In both seasons and models, all of the variables are important for the prediction of water hyacinth at different levels.

4.2.2.1.Importance of predictor variables for the wet season

Figure 4.3 below discloses the level of importance of predictor variables. The key predictor variables are presented with their percent contribution. The variables chosen to train the model showed different levels of predictive performance in each machine learning model. The figure has four sections, and the x-axis of each section shows the percentage of importance, while the y-axis shows the predictor variables. The predictor variables include B12, B5, B3, annual average temperature of the wet season (Twet), annual average precipitation of the wet season (PPwet), NDVI, SAVI, NDAVI, NDWI, VV, and VH.

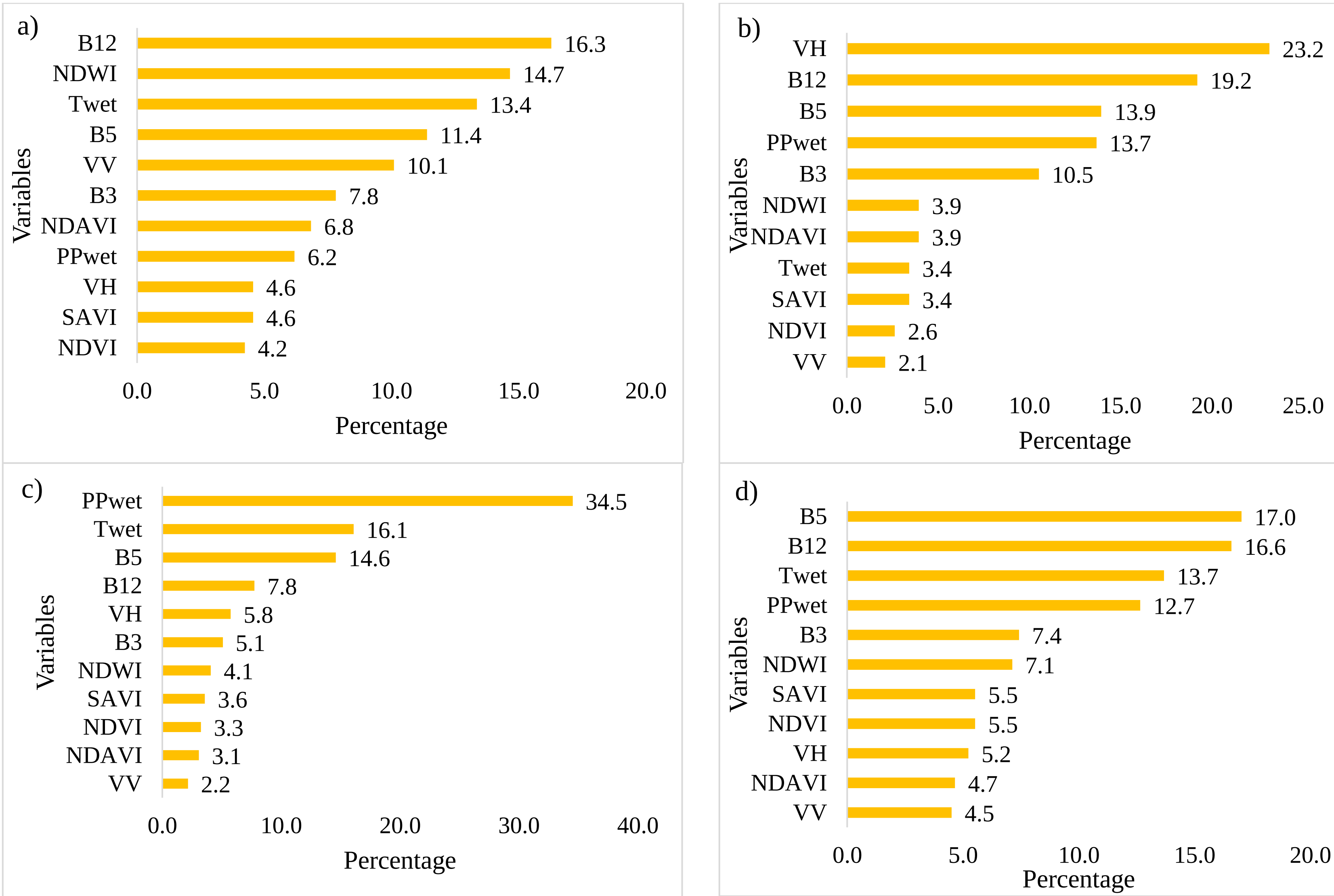


Figure 4.3 Predictor variable importance a) RF model b) BRT model c) SVM model and d) Ensemble model, wet season (2022)

Based on the above Figure (4.3a) of variable importance, 66% of the prediction in the RF model is performed by B12, NDWI, Twet, B5 and VV. Among those five variables, B12 (16.3%) is the most highly important variable, followed by NDWI (14.7%), Twet (13.4%), B5 (11.4%), and VV (10.1%), in decreasing order. On the other hand, NDVI, SAVI and VH performed less, with an aggregate value of 13.4%. The Sentinel-2A datasets that include spectral indices and bands were the most important dataset, followed by Sentinel-1 SAR indices. In this model, the importance of climate variables is limited, but unignorable.

Regarding the BRT model (Figure 4.3b), 80.5% of the prediction was made by five variables in decreasing order VH (23.2%), B12 (19.2%), B5 (13.9%), PPTwet (13.7%), and B3 (10.5%). The least performed variables are NDVI and VV, with percentages of 2.6 and 2.1, respectively. Variable VH from the Sentinel-1 SAR dataset is the most important variable in the predication BRT model. A study by Janssens et al., (2022) also assures that a C-band (wavelength of ~5.5465763 cm) SAR imaging instrument carried by the ESA Sentinel-1 satellites enables the identification of water hyacinth through haze and cloud conditions.

According to Figure 4.3c, the most key variables in the SVM model prediction are PPwet and Twet of the wet season, which account for more than 50%. In addition, nearly 33% of the prediction was made by Sentinel-2A bands: B5 (14.6%), B12 (7.8%), B3 (5.1%) and VH (5.8%). Although there is variation in the importance level of predictor variables, the degree of difference between the least performed variable (VV) and the 4th ranked variable B12 is not significant. In this model, climatic variables (temperature and precipitation of the wet season) are the most crucial variables.

Similar to other machine learning models, the degree of importance of variables in the Ensemble model is quite different. The study found that approximately 60% of the model prediction is derived from B5, B12, Twet and PPwet, with values of 17%, 16.6%, 13.7% and 12.7%, respectively. Hence, Sentinel-2A bands and climatic variables outweigh other variables in the case of the Ensemble model. Although the importance of those spectral bands and climatic variables is high, NDAVI (4.7%) and VV (4.5%) variables are the least important predictors of water hyacinth.

Figure (4.3) also shows that some predictor variables are consistently important across different models, such as B12, B5, Twet and PPwet. These variables have a strong influence regardless of the modeling method. On the other hand, some predictor variables are less important but nonnegligible across different models in the wet season, such as SAVI and VV. These variables have a relatively weak influence on the distribution of water hyacinth.

Regarding the dataset, approximately 35.5% and 30.3% of the RF model are predicted by Sentinel-2A bands and vegetation indices, respectively. Similarly, the highest 55.7% and 21.5% of the BRT model and 41% and 26.3% of the ensemble model predictions are derived from Sentinel-2A bands and climatic variables, respectively. Contrary to the RF, Ensemble and BRT models, in the case of the SVM model, climatic variables predict higher (50.6%), followed by Sentinel-2A spectral bands (27.4%). The results of the study revealed that in all four models, the prediction capability of Sentinel-2A bands, particularly B12, which is SWIR, and B5 (red edge) are the highest. The highest reflectance value of water hyacinth in the spectral domain of SWIR1 and 2 and NIR becomes very important to model the spatial distribution of the weed.

4.2.2.2. Importance of predictor variables for the dry season

Similar to the wet season, the degree of importance of variables for the dry season is calculated using percentages. The list of variables is similar to that of the wet season except for the difference in season. The month from January to May 2023 is considered for dry season modeling. Likewise, the level of importance of variables is different for those selected models depending on the nature of the model.

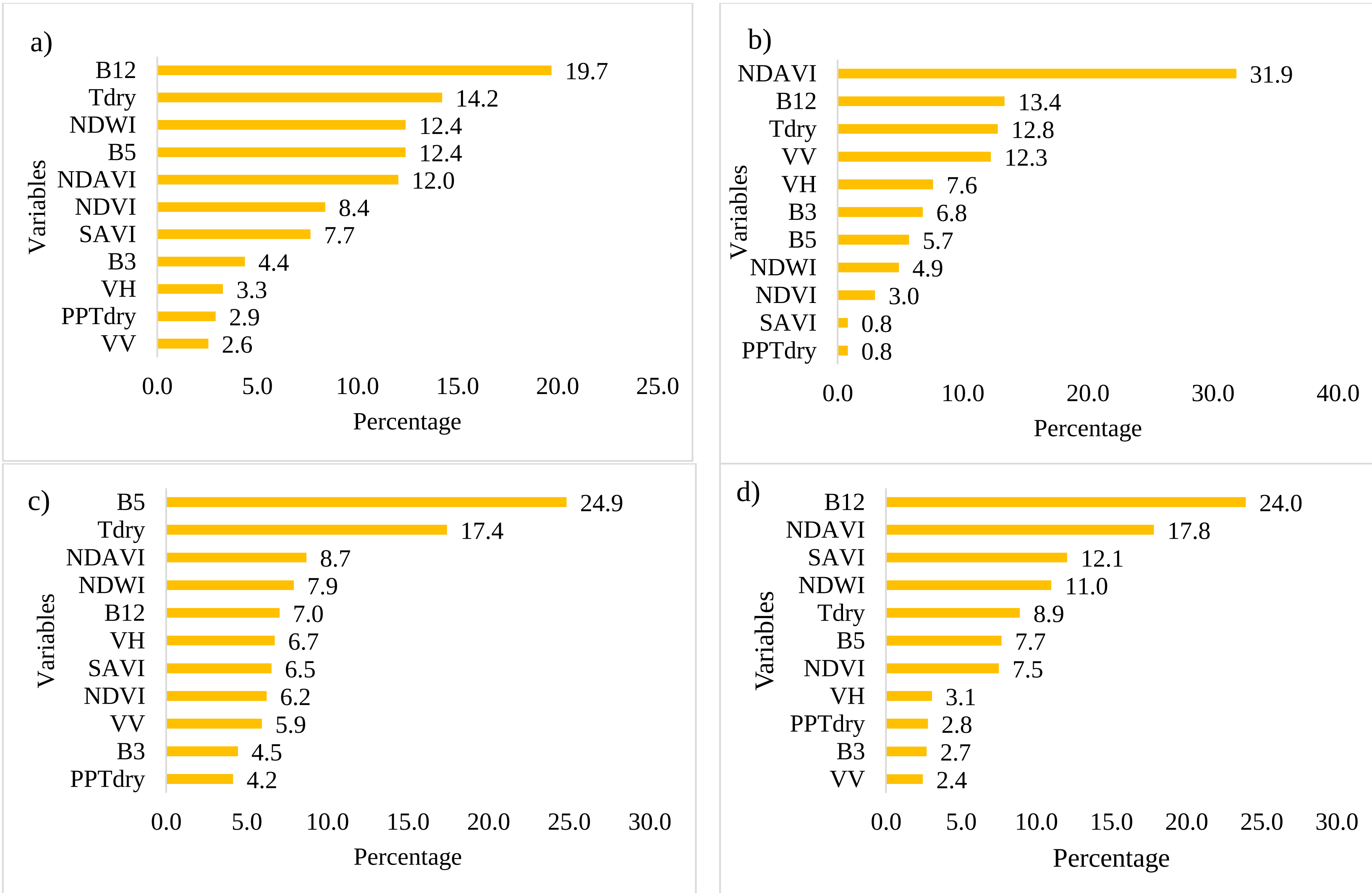


Figure 4.4 Predictor variable importance a) RF model b) BRT model c) SVM model and d) Ensemble model, Dry season (2023)

Figure 4.4 shows a key variable that predicts the spatial distribution of water hyacinth across four machine learning models during the dry season. According to the first RF model (4.4a), approximately 70% of the prediction was contributed by B12 (19.7%), annual average temperature of dry season (Tdry 14.2%), NDWI (12.4%), B5 (12.4%) and NDAVI (12%). The VV of Sentinel-1SAR and the annual precipitation of the dry season (PPTdry) contributed less to the model, with values of 2.6% and 2.9%, respectively. This indicates that SWIR1 and SWIR2 of the Sentinel-2A bands contributed immensely to modeling.

Figure (4.4b) shows the relevance of the variables in the BRT model. The model assigns the highest importance to NDAVI (approximately 31.9%), followed by B12 (approximately 13.4%) and VV (approximately 12.3%). This indicates that over 55% of the prediction is contributed by three variables, namely, NDAVI, B12 and VV. The precipitation of the dry season (PPTdry) and SAVI are the least predictive variables, with an equal value of 0.8%. More than 60% of the SVM model prediction was derived from four variables (see Figure 4.4c). The SVM model assigns the highest importance to B12 (approximately 25%), followed by Tdry (approximately 17.4%), NDAVI (approximately 8.7%) and NDWI (approximately 7.9%). In this model, the precipitation of the dry season (PPTdry) is the least important variable, with a value of 4.2%. The two most important variables, B12 and mean annual temperature of the dry season (Tdry), also had the highest positions in the RF model.

The ensemble model assigns the highest importance to B12 (approximately 24%), followed by NDAVI (17.8%), SAVI (12.1%), and NDWI (11%). Those most important variables account for approximately 65% of the total prediction. On the other hand, VV, B3 and PPTdry become the least important variables in the prediction of this model. Figure 4.4 also shows that some predictor variables are consistently important across different models, such as B12, NDAVI and Tdry. Except for the BRT models, B12 and NDAVI are the most relevant variables in the RF, SVM and ensemble models. These variables have a strong influence on the modeling dry season. On the other hand, some predictor variables are less important across different models, such as the annual precipitation of the dry season (PPTdry) from the climatic variable and VV from the Sentinel-1 SAR datasets. Regarding the dataset, the aggregate of Sentinel-2A index variables that comprise NDVI, NDAVI, SAVI and NDWI consistently show a high degree of predictive importance in the RF, BRT and Ensemble models, with percentages of 40.5%, 40.6% and 48.4%, respectively. Next to Sentinel-2A indices, Sentinel-2A bands that comprise B3, B5 and B12 are the second most relevant datasets used in the modeling process. In contrast, Sentinel-2A bands are the leading predictor variable in the SVM model, with a value of 36.3%, followed by indices. From the list of bands, B12 and B5 are the most important variables in the prediction. The other two datasets, climate and Sentinel-1 SAR bands, also significantly contributed values of 17% and 5.8% in the RF model, 13.6% and 19.9% in BRT, 21.6% and 12.7% in SVM, and 11.7% and 6% in the ensemble model, respectively. This demonstrates that climatic variables (precipitation and temperature) and Sentinel-1 SAR bands help to detect the geographic dispersal of water hyacinth.

Generally, the predictive variable importance of the models in the two seasons shows that B12, B5, the annual average temperature of the seasons (Twet and Tdry), the annual precipitation of the wet season (PPwet), and NDAVI are the most consistent variables. The results indicate that Sentinel-2A bands, particularly shortwave infrared (SWIR1 and 2) and red-edge band (B5) and vegetation indices as well as climatic variables are the most significant datasets. A study by Thamaga and Dube (2019) concluded that red edge bands (B5-7), NIR, SWIR 1 and SWIR 2 of Sentinel-2A showed a unique capacity for differentiating water hyacinth from other neighboring aquatic plants. This further elaborated that in both seasons, the weed can be distinguished spectrally from other types of land use/land cover, such as shrubland, forests, riparian vegetation, and plantations, primarily in the red edge (1-3), NIR, SWIR 1 and SWIR 2 regions of the spectrum. Another study by Mucheye et al., (2022) found that the reflectance of water hyacinth becomes much higher in the red edge and NIR spectra. The high reflectance is associated with the internal spongy leaf structure of water hyacinth. Mouta et al., (2023) also outlined that the shortwave infrared spectral region of Sentinel-2A is more sensitive to vegetation variation and able to distinguish water hyacinth. However, this study excludes NIR (B8) and SWIR1 (B11) of Sentinel-2A from the variables list, as it shows a greater degree of correlation ($r > 0.8$) with the red-edge band (B5) and B12, respectively. According to Dogliotti et al., (2018), the greater reflectance characteristics in the red-edge, NIR, SWIR 1 and SWIR 2 bands are also associated with the cellular structure of the leaves, contrasting reflectance properties of leaves and the surrounding water, pigments, and chlorophyll content.

Although the spectral bands of Sentinel-2A have a distinctive capacity, vegetation indices have also played a pivotal role. The study reveals that vegetation indices of NDAVI and water index

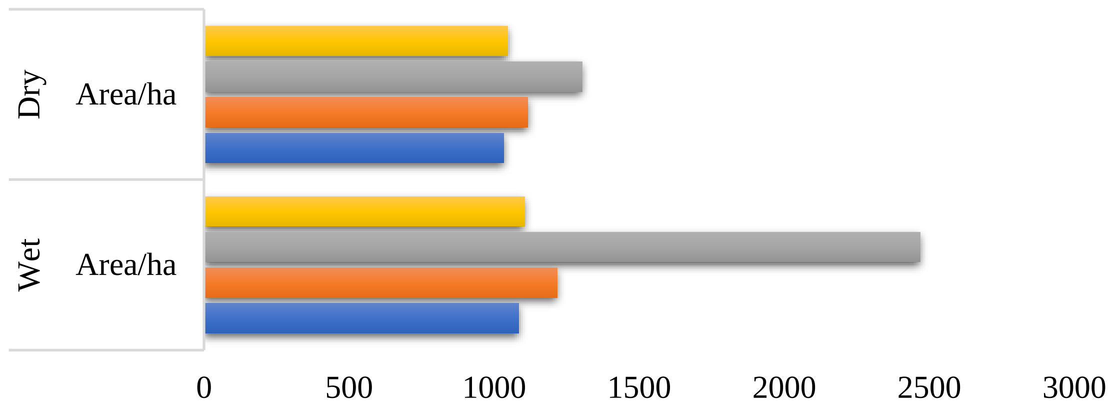
NDWI exhibit a great deal of performance in detecting water hyacinth spatial extent. The findings of this study are consistent with those of Damtew et al., (2021). The study found that the vegetation indices of NDAVI and NDVI showed a better performance in distinguishing aquatic microphytes such as *Cyperus, Echinochloa, and Phragmites* in Lake Zeway. However, in this study, the performance of NDVI was lower than that of NDAVI in both seasons. NDVI is ineffective for detecting submerged and partially submerged vegetation at greater depths, and tidal flooding of water bodies may alter the reflectance characteristics. NDVI is used for broader detection of terrestrial green vegetation. However, in the case of aquatic vegetation such as water hyacinth, NDAVI has a better detection capacity. By utilizing the blue band and near-infrared, NDAVI provides a better indicator of aquatic vegetation biomass and density in turbid waters compared to traditional NDVI. It has been used effectively to map and monitor aquatic macrophytes, algae and other hydrophytes (Villa et al., 2013; Ma and Zhou, 2018; Brooks et al., 2019; Gerardo and de Lima, 2022b). In addition, the value of NDWI (-0.8-0.04) is used to delineate water bodies and water hyacinth (Gerardo and de Lima, 2022a).

This study also found that Sentinel-1 SAR variables, particularly the VH band, significantly predicted water hyacinth in the BRT and SVM models for the two seasons. The VV and VH bands help to distinguish between open water with low backscatter and vegetated water bodies with high backscatter (Hardy et al., 2019). Simpson et al., (2022) demonstrated the use of Sentinel-1 SAR data to detect water hyacinth. The VV and VH polarizations of Sentinel-1 SAR data exhibit a greater accuracy. According to the results, a threshold value of 3 dB is used to separate the infested and clean pixels of water bodies.

As indicated above, climatic variables, particularly temperature and precipitation in the wet season, were key variables. Thakuri et al. (2019) also found that the average temperature of the wettest quarter, the warmest quarter, and the coldest months are the significant variables that contribute the most to the distribution of water hyacinth. Thamaga and Dube (2019) also found that during the wet season, extreme temperature and high rainfall conditions coupled with runoff that carries nutrients are the most suitable conditions for proliferation. Dersseh et al., (2019b) found that temperatures of 22-26°C for the wet season and 20-26°C for the dry season are the most conducive for the quick spread and rapid growth in Lake Tana.

4.2.3. Spatial coverage of water hyacinth

A threshold level is used to classify the area as absent or present, and the extent of water hyacinth is presented in Figure 4.5.



	Wet Area/ha	Dry Area/ha
Ensemble	1105.4	1046.8
BRT	2469.3	1304.5
SVM	1217.4	1116.7
RF	1085.4	1033.9

Figure 4.5 Spatial coverage and distribution of water hyacinth

Figure 4.5 shows the spatial coverage and distribution of water hyacinth using the RF, BRT, SVM and Ensemble models in 2023. As the performance of the models is quite different, the estimation of spatial distribution by the models is also variant. The coverage values for the Ensemble, BRT, SVM and RF models during the wet season are 1105.4 ha, 2429.3 ha, 1217.4 ha, and 1085.4 ha, respectively. The best-performing RF model predicted that 1.5% (1085.4 ha) of the study area was infested, followed by Ensemble (1.6%), SVM (1.9%), and BRT (2.4%). The results show that the areal coverage and spatial distribution of the four models are nearly similar. Based on BRT model prediction, the highest proportion (2429.3 ha or 2.4%) of the lake is occupied compared to others.

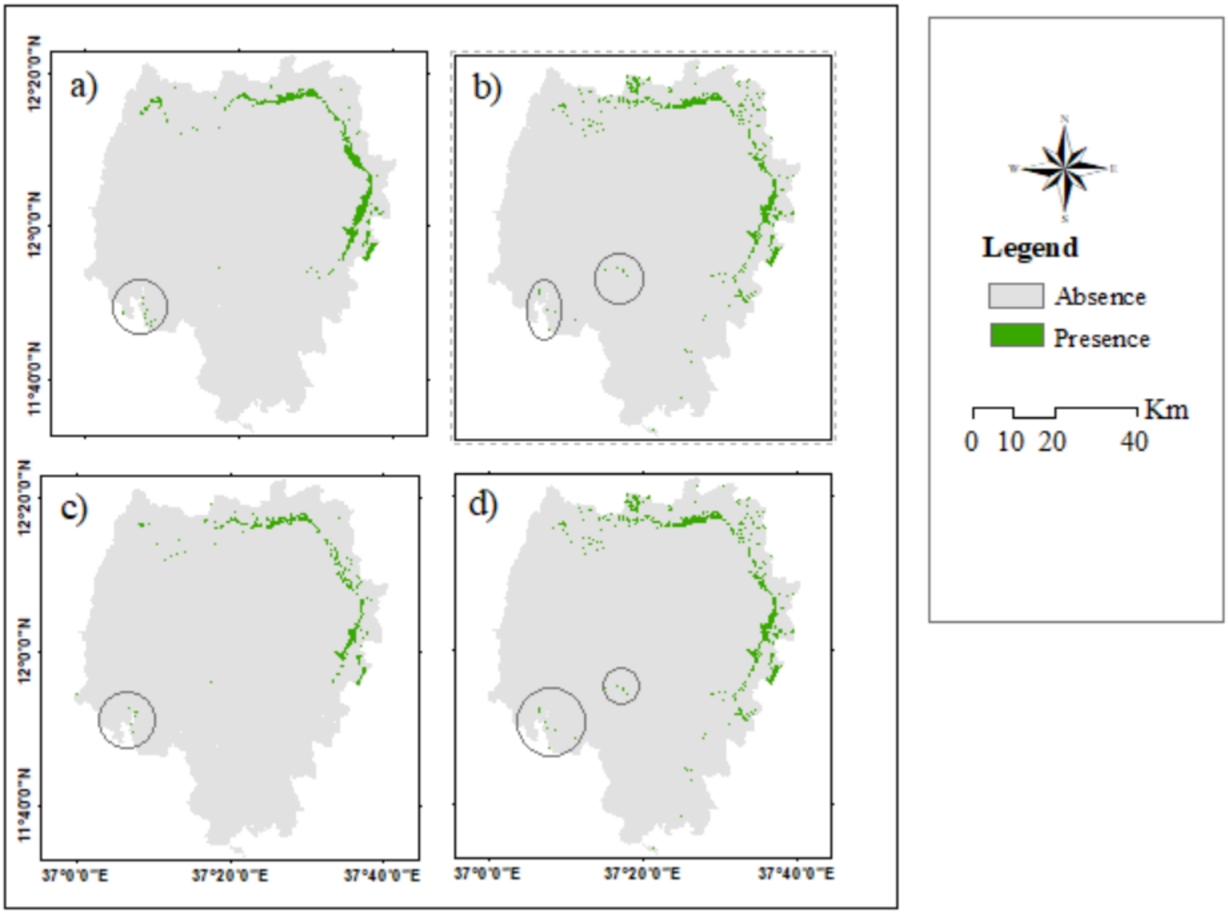


Figure 4.6 Absence and presence of water hyacinth for the wet season: a) RF, b) SVM, c) BRT and d) Ensemble

Figure 4.6 is a map that shows the geographic infestation of water hyacinth of the four models based on a threshold level. The encircled portion of the study area is an absent area where water hyacinth infestation is not observed on the ground but is wrongly predicted by the models. In particular, the models wrongly predicted the westernmost tip of the study area as highly infested by water hyacinth. According to Figure (4.5), the total area that is infested by water hyacinth predicted by the BRT model is approximately 2429.3 ha. This indicates that there are a large number of pixels that are wrongly predicted as water hyacinth, which is an indicator of the overestimation of the model. There is a 1343.9-ha difference between the best-performing RF model and the BRT model. On the other hand, the RF, SVM, and Ensemble models also wrongly predict the eastern edge of the study area, as it has been infested by water hyacinth (see 4.6 a, c, d figures that are encircled). Overall, the northern, northeastern and eastern lakeshores are highly invaded areas.

Regarding the dry season, Figure 4.5 above shows that a total area of 1033.9 ha, 1116.7 ha, 1304.5 ha and 1046.8 ha of land is infested by water hyacinth using the prediction of the RF, SVM, BRT and Ensemble models, respectively. The study found that the RF SVM, BRT and Ensemble models predicted that 0.8%, 1.2%, 0.9%, and 0.8% of the study area was infested, respectively. The results show that the infested area decreases during the dry season. However, it is highest in the wet season. For instance, there is a 51.5-ha, 1164.8-ha, 100.3 ha and 58.6 ha decrement in the infestation level using the RF, BRT, SVM and Ensemble models, respectively.

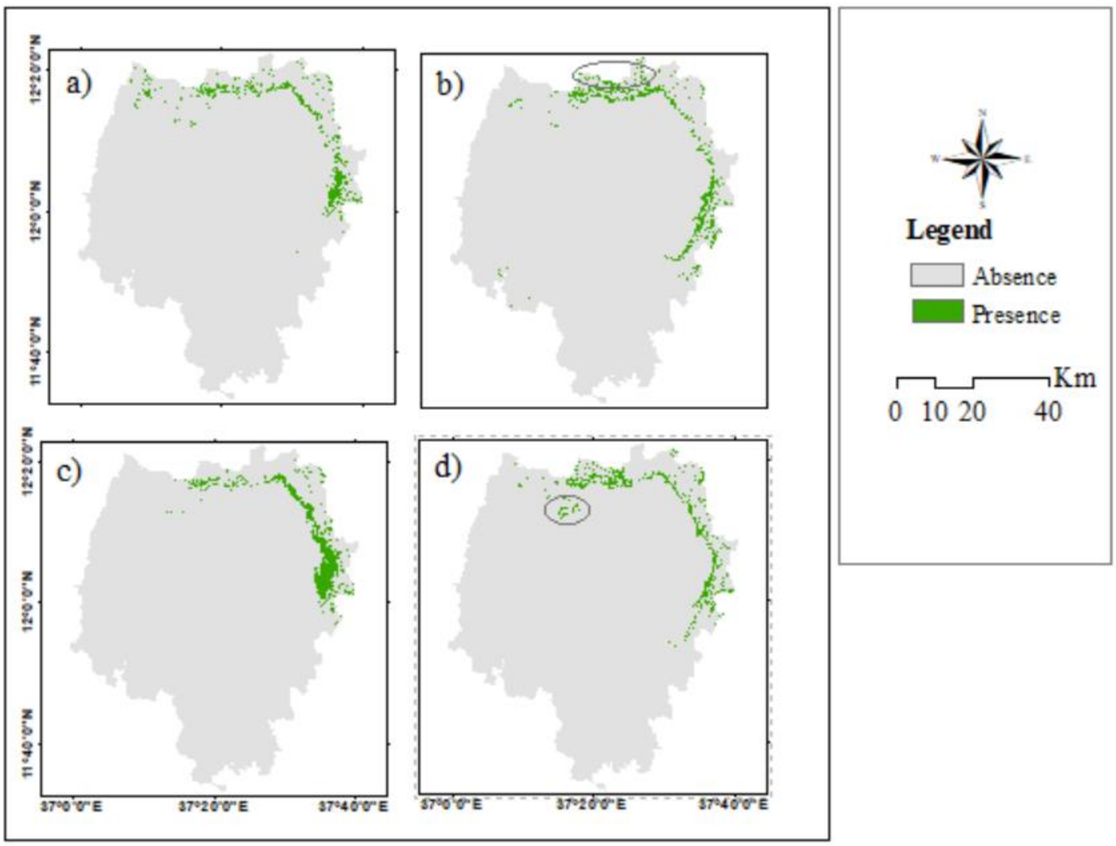


Figure 4.7 Absence and presence of water hyacinth for the dry season: a) RF, b) SVM, c) BRT and d) Ensemble

Figure 4.7 shows the spatial coverage predicted by the four models. As shown in the figure, almost all of the models uniformly predict the spatial distribution of water hyacinth. Nevertheless, the models have yet to predict areas that have no water hyacinth as it is infested (see Figure 4.7, the encircled models). Specifically, the northern tip of the study areas was wrongly predicted. According to the Ensemble model, approximately 0.8% of the area is infested by water hyacinth.

However, some portion of the study area is wrongly predicted as affected water hyacinth (see Figure 4.7d). In this study, the ensemble model was underestimated and the BRT model overestimated the spatial presence of water hyacinth. As shown in the figure, the northeastern and eastern parts of the lake are highly infested regions.

Overall, the study found that the RF model is the best model that consistently shows spatial infestation across the two seasons, followed by the SVM and ensemble models. However, the BRT model is not consistent, and it over- and underestimated the spatial coverage. The misprediction of presence in the unoccupied area is related to the presence of dense vegetation that has spectral reflectance similar to water hyacinth. The results of the study are also aligned with Bayable et al., (2023).

The study found that the spatial coverage was different in the two seasons. As shown in Figures 4.7, 4.6 and 4.5, the spatial coverage becomes highest during the wet season and lowest during the dry season. The highest biomass and distribution of weeds occurred around the northern, eastern and northeastern lakeshores. This result is comparable to that of Dersseh et al., (2019b), who found that coastal eutrophication, lake topography and wind direction are the main reasons for the high density and expansion in the region. Additionally, the deposition of sediments from the major rivers of Gumara, Rib and Megech leads to high infestation levels. A study by Thamaga and Dube (2019) found that the areal coverage and infestation level increased in the wet season. Another study by Mucheye et al., (2022) and Bayable et al., (2023) identified that from July to October, the water hyacinth becomes high and proliferates in Lake Tana. The biomass started to decrease after February and was lowest during the dry season. An interview results with an expert from the ARLTWBPDA also reveal that the density and spatial coverage of the weed is highest during the wet season (from June to mid-October) and lowest during the dry season (February to May). Thamaga and Dube (2019) concluded that high concentrations can be associated with increased growth rates, which are facilitated by nutrient inputs from nearby agricultural areas as a result of enhanced river flows and runoff. Conversely, the decrease in runoff and nutrient replenishment leads to reduced nutrient movement and consequently creates less favorable conditions in the dry season. In addition, the interview results show that the wet season provides favorable environmental conditions for water hyacinth growth. The increased rainfall and higher humidity created a moist environment that is suitable for the germination of seeds and the growth of their vegetative parts. Additionally, the warmer temperatures during the wet season also contribute to accelerated growth. Furthermore, flooding events during the wet season can lead to the displacement of water hyacinth plants, allowing them to spread to new areas. Floodwaters can carry water hyacinth fragments downstream, facilitating their dispersal and colonization of new habitats. The ability of water hyacinth to float and form dense mats on the water surface enhances its chances of survival and spread during flooding events.

4.3. Spatiotemporal distribution of water hyacinth

4.3.1. Accuracy assessment result

Table 4.4 Accuracy assessment results for the dry and wet seasons

Year	Overall Accuracy		Kappa	
	Wet	Dry	Wet	Dry
2016	0.91	0.93	0.88	0.9
2018	0.94	0.94	0.92	0.91
2020	0.93	0.95	0.92	0.93
2022	0.94	0.94	0.91	0.92

The table above (4.4) shows the accuracy assessment results for RF machine learning classification. The table includes the overall accuracy and Kappa coefficients for 2016, 2018, 2020, and 2022. In the given table, the overall accuracy ranges from 0.91 to 0.94 for the wet season. These values suggest that the classification model achieved a high level of accuracy. The Kappa coefficients range from 0.88 to 0.92. Kappa coefficient values closer to 1 indicate a higher level of accuracy. Therefore, the kappa coefficients in the table suggest a strong agreement between the observed land use/land cover class and the expected class.

From Table (4.4), we can see that the overall accuracy and the kappa coefficient are both high. The highest value was obtained in 2020, with an overall accuracy of 0.95 and a kappa coefficient of 0.93. This means that 95% of the pixels are correctly classified. The lowest values were obtained in 2016, with an overall accuracy of 0.93 and a kappa coefficient of 0.9. Overall, the accuracy assessment result indicates that the classification method is reliable and accurate for detecting water hyacinth and water bodies in Lake Tana.

4.3.2. Feature importance

The study employs a total of 16 variables across Sentinel-2A bands, indices, Sentinel-1SAR bands and topography. The variables have shown different levels of performance in different years and seasons. The variation in importance level is associated with the GCPs that are used to train the model.

According to Figure 4.8, the most important variables across the dry season of the study period are NDAVI, SAVI, B5, B8, B4 and NDVI. However, topographic variables such as slope and elevation as well as SAR bands also significantly contribute to the classification. NDAVI, SAVI and B5 were the most relevant variables for 2022 and 2020. B8, elevation and B4 (2018) and NDVI and B5 (2016) were highly important variables for the classification. Figure 4.9 shows that B5, B11, B12, VH, NDAVI and NDWI are among the most important covariates used in the classification of the wet season. In addition, slope, SAVI and VV variables also have high contributions. B11 and B5 (2022), VH, SAVI and NDWI (2020), B5 and B12 (2018), and VH and slope (2016) are the variables that are highly important in the classification of the RF model.

The result indicates that, when compared to other vegetation types or land use/land cover, water hyacinth can be identified by its spectral reflectance patterns in these bands and indices. The most important variable, NDAVI, combines the reflectance values in the red and NIR bands to highlight the presence and distribution of water hyacinth. The study also shows that the VH polarization is particularly useful for discriminating water hyacinth from land use/land cover types due to its specific backscatter response at a central frequency of 5.404 GHz.

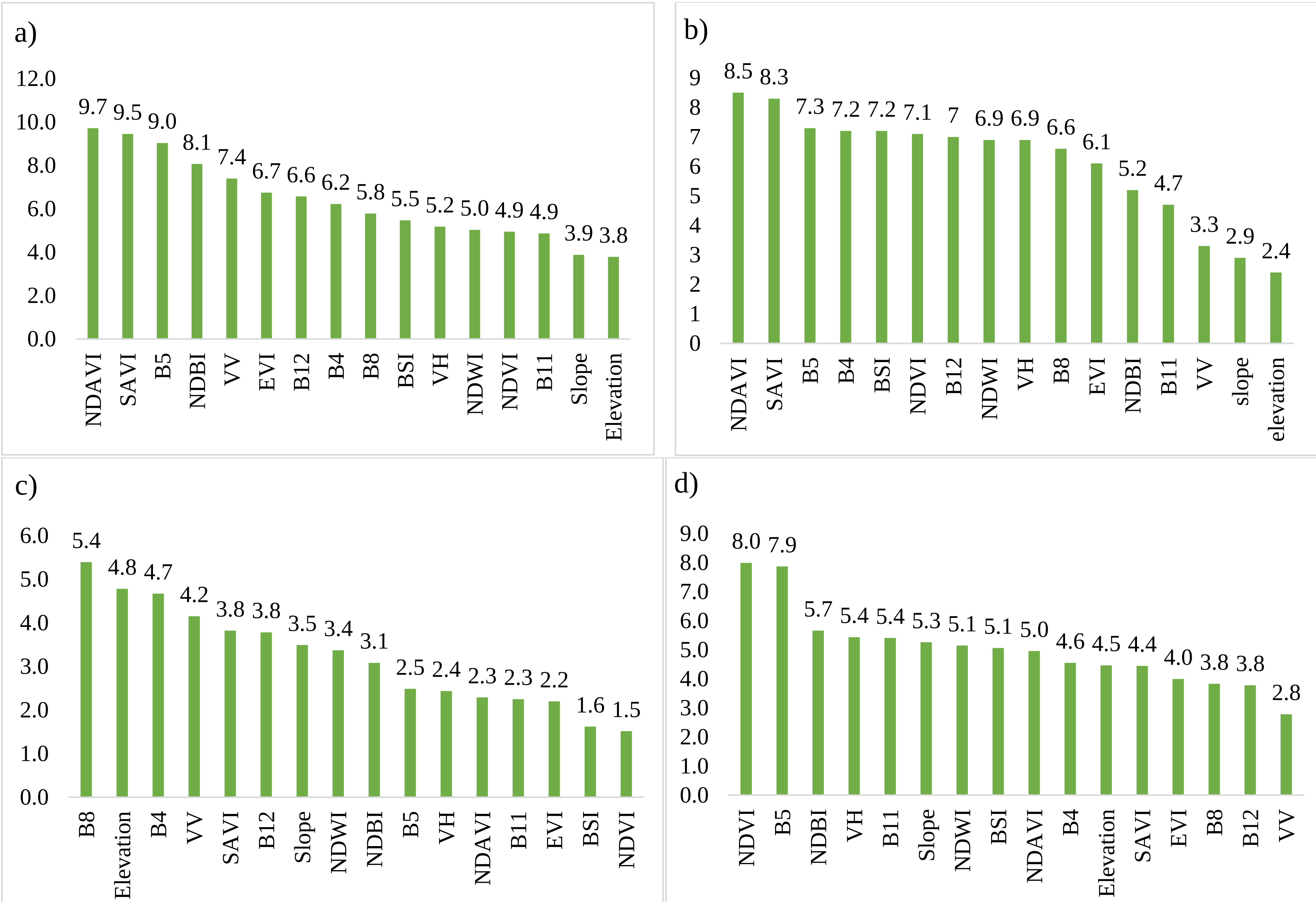


Figure 4.8 Feature importance for the dry season a) 2022, b) 2020, c) 2018 and d) 2016

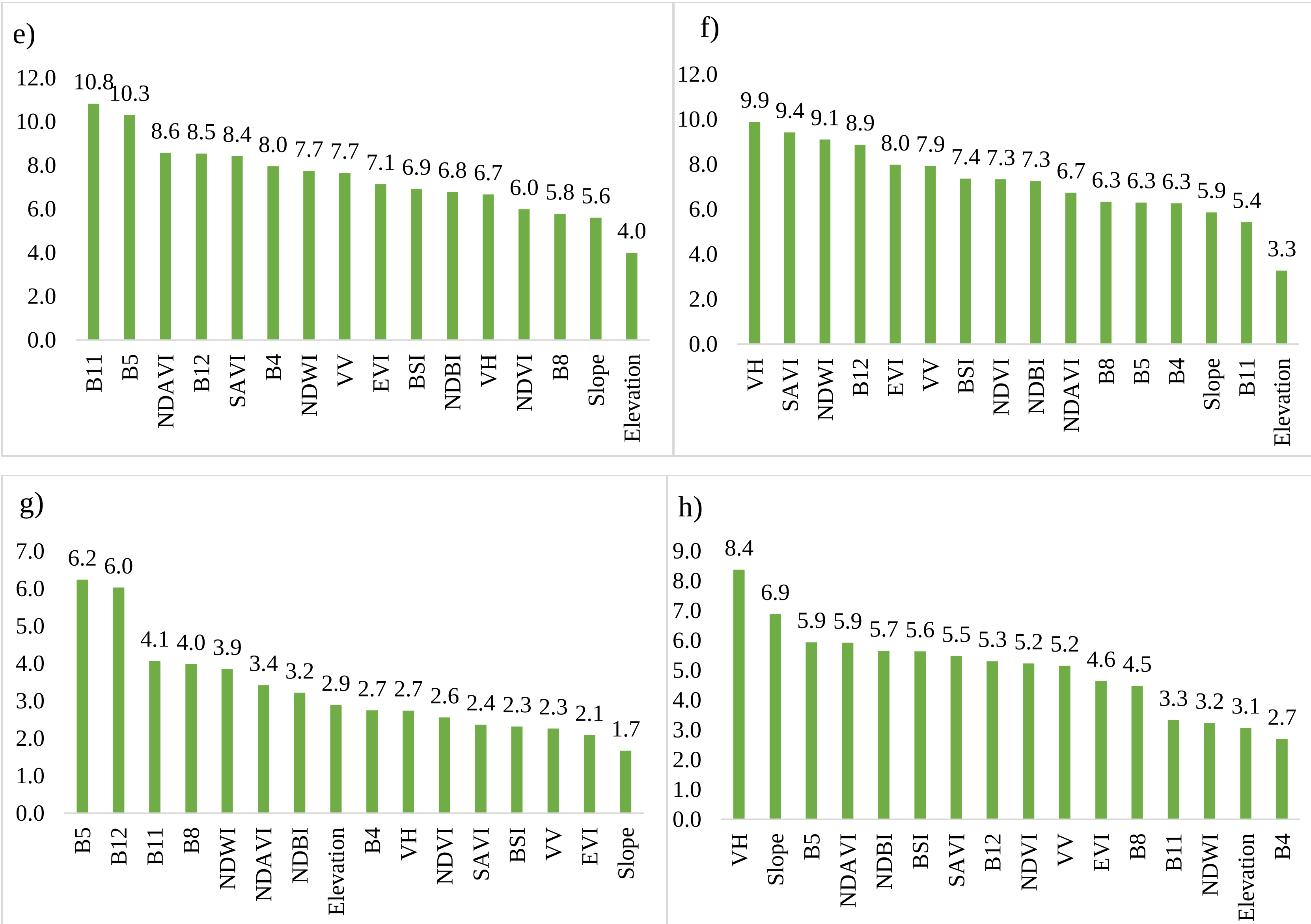


Figure 4.9 Feature importance for the wet season e) 2022, f) 20202, g) 2018 and h) 2016

4.3.3. The spatiotemporal distribution of water hyacinth for the wet season

Table 4.5 Spatial coverage of water hyacinth (2016-2022), wet season

Year	Area/ha	%
2016	686.5	2.2
2018	1851	6.0
2020	1396.7	4.6
2022	1436.5	4.8

Table 4.5 shows the spatiotemporal coverage over a period of four years: 2016, 2018, 2020, and 2022. The data include the area covered in ha and the percentage of the total lake area that the water hyacinth occupies. In 2016, it covered an area of 686.5 ha, which accounted for 2.2% of the total lake area. By 2018, the coverage had increased significantly, reaching 1851 ha, representing 6% of the lake area. In 2020, the areal coverage decreased slightly to 1396.7 ha, accounting for 4.6% of the total lake area, and in 2022, the coverage increased again to 1436.5 ha, representing 4.8% of the lake area.

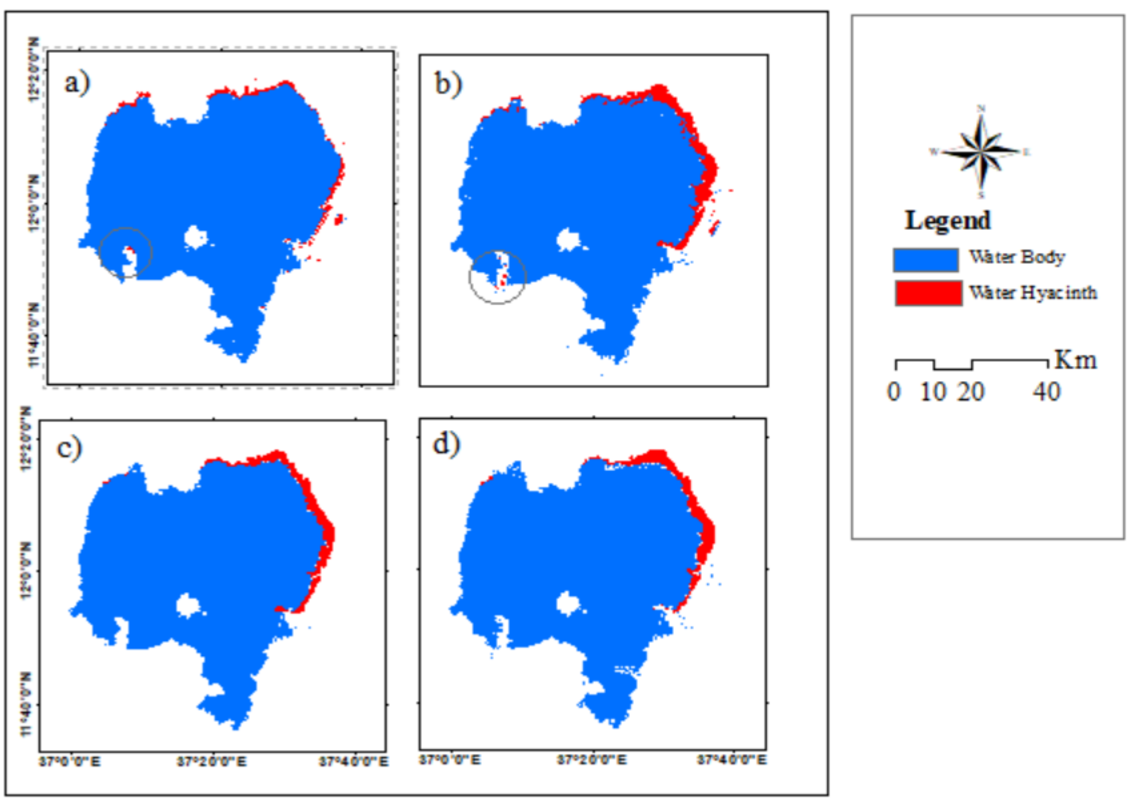


Figure 4.10 Spatiotemporal distribution of water hyacinth wet season a) 2016 b) 2018 c) 2020 d) 2022

4.3.4. The spatiotemporal distribution of water hyacinth for the dry season

Table 4.6 Spatial coverage of water hyacinth (2016-2022), dry season

Year	Area/ha	%
2016	650.4	2.1
2018	1259	4.2
2020	1305.7	4.3
2022	1216.5	4.0

According to Table (4.6), the coverage increased from 2.1% (650 ha) in 2016 to 4.3% (1305.7 ha) in 2020, which is a significant change over a short period of time. However, a slight decrease in the infested area was recorded from 2020 to 2022, with a value of 4%. A total of 89.2 ha of land was cleared from water hyacinth between 2020 and 2022.

According to Tables 4.5 and 4.6, the water hyacinth infestation increased from 2016 to 2018 but decreased slightly from 2020 to 2022 in both seasons. In addition, the coverage is much higher in the wet season than in the dry season. For example, in 2018, it covered 1259 ha (4.2%) of the lake in the dry season but 1851 ha (6.0%) in the wet season. This means that there was a difference of 592 ha (1.8%) between the two seasons. A study by Bayable et al., (2023) also confirmed that area coverage significantly decreased (62.5%) from winter to spring but significantly increased (81.7%) from wet to autumn from 2021 to 2022. From the tables, it is possible to see that water hyacinth coverage peaked in the 2018 wet season and then declined considerably in 2020 and 2022. The results of this study are consistent with those of Dersseh et al., (2020). The study found that in 2015, 2016, 2017, 2018, and 2019, the maximum water hyacinth coverage was 278.3, 613.6, 1108.7, 2036.5, and 2504.5 ha, respectively.

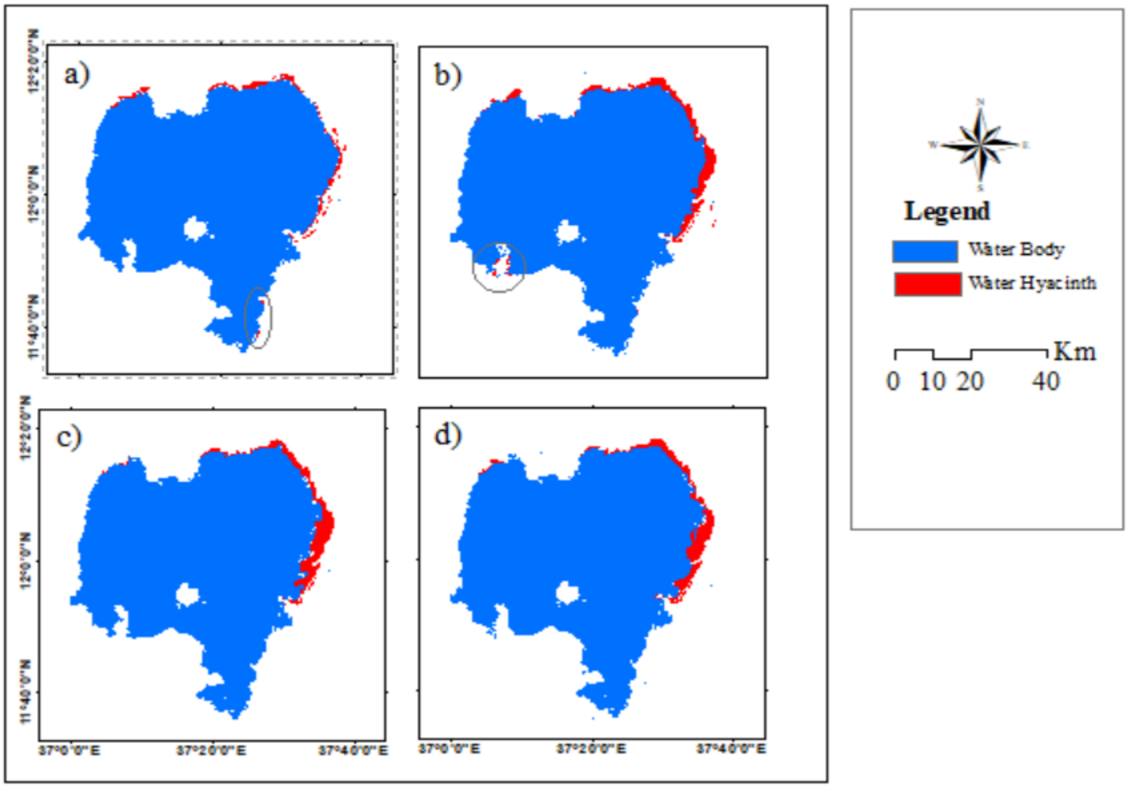


Figure 4.11 Spatiotemporal distribution of the water hyacinth dry season a) 2016 b) 2018 c) 2020 d) 2022

Interview results revealed that water hyacinth dynamics are influenced by seasonal climate variability. According to Dersseh et al., (2020), lake-level fluctuations and seasonal climate variability or rainfall patterns affect the extent of coverage. The increase in the water level allowed the weed to spread, expanding its range of habitats from the shallows to the flooded region. Low water levels reduce the available area for the weed and expose it to higher temperatures and solar radiation, which may affect its survival and reproduction.

According to Worqlul et al., (2020) and Abebe et al., (2023), seasonal climate variations, particularly those related to lake level and water temperature, have an impact on water hyacinth infestation. The quick spread in the rainy season is related to lake level rise and nutrient concentration. On the other hand, when the lake level reduces and the water temperature rises, water hyacinth typically declines.

According to the interview results, the decline in coverage of water hyacinth is related to human interventions, such as mechanical and removal of the weed by humans. The local communities together with the government have been trying to manage the rapid infestation by using mechanical removal that involves the use of machinery and equipment to physically remove from water bodies. Manual removal involves manually pulling out the plants from the water to reduce the spatial coverage. When the coverage of water hyacinth diminishes, the local farmer plows the land and produces crops with the application of fertilizer. As a result of these practices, once again during the rainy season, the rate of infestation increases rapidly due to the higher nutrient level resulting from the application of fertilizer from the surrounding farmland. This result is consistent with Dersseh et al. (2020). The study employs satellite images to estimate coverage in Lake Tana from 2011 to 2019. They found that the areal coverage increased from 20 km^2 in 2011 to 40 km^2 in 2019, with a peak of 50 km^2 in 2018. They also observed that water hyacinth was mainly distributed along the northeastern shore of the lake. Another study by Worqlul et al. (2020) used high-resolution PlanetScope satellite images to examine the dynamics of spatial coverage from August 2017 to July 2018. They found that geographic coverage increased from 4.3 km^2 in August 2017 to 23.4 km^2 in April 2018 and then decreased to 14.5 km^2 in July 2018. A recent study by Abebe et al., (2023) found that between 2011 and 2019, the annual water hyacinth area considerably expanded. For instance, from 2011 to 2019, the lake's surface area increased by approximately 1603 ha with the rapid proliferation of water hyacinth.

4.4. Management of water hyacinth in Lake Tana

According to the key informant interview results, different management efforts have been carried out since the time of infestation in 2011. The management practices have been aimed at controlling the spread of the plant and reducing its negative impacts on the lake's ecosystem and local communities. The management practices have involved a combination of mechanical and physical control methods, as well as awareness-raising campaigns to promote community participation in the management of the lake. The most widely used management strategies are manual removal by the labor force, mechanical use of machinery and floating harvest. According to key informant interview results with experts and local peoples, physical control or manual removal is the most important and highly practiced water hyacinth management practice in Lake Tana. Since 2012, a significant number of individuals have been actively participating in the campaign to remove weeds. The Land Administration and Use Bureau of Amhara Region (LAUBAR) reports that, from 2013 to 2017, 212,779, 45,097, 242, 086, 98,755, and 184,161 people participated in manually clearing water hyacinth from the lake. According to Enyew et al., (2020), from 2012 to 2018, more than 800,000 human laborers participated. The majority of the people who participated in manual harvesting removal campaigns are farmers and rural communities living nearby villages. The people used their own hands, wooden sticks, reed boats, and motorized boats to harvest water hyacinth. The manual removal of weeds has been focused on leveled terrains and shallow water sections. According to the local community, because the water hyacinth weed is spreading so quickly and manual removal of the weed exposes people to insects and snakebites, it is becoming increasingly difficult to remove the plant by hand. Moreover, the plant's rapid growth and ability to reproduce quickly make it challenging to completely eradicate weeds manually. It also requires a significant amount of labor and time. Experts also disclosed that the availability of resources, such as manpower, equipment, and funding, poses a significant challenge for sustained manual removal efforts. Limited resources hinder the effectiveness and efficiency of removal campaigns.

Figure 4.12 Manual removal of water hyacinth by the local community.

Source: Bahir Dar University, 2020

Another method of managing water hyacinth that has been implemented in Lake Tana is the mechanical method. Harvesting machinery has been employed to manage water hyacinth. According to EFWPDA reports, seven harvesting machines have been purchased with the help of private businesses, nonprofit organizations, and public colleges. The harvesting machine has been operating in places where dense mats of water hyacinth are largely observed, such as the Fogera area.

Figure 4.13 Mechanical harvesting of water hyacinth

Source: Charitable Organization for Integrated Tana Basin Development, 2020

However, problems including the muddy and clay nature of the lake shore, the harvesting equipment breaking down quickly and needing infrequent maintenance, and the expensive cost of the technology have made it difficult to mechanically control weeds with machines or floating harvesters. Machinery used for water hyacinth removal requires regular maintenance and repair to ensure its optimal performance. In remote areas or areas with limited resources, obtaining spare parts and skilled technicians for maintenance and repair is a major challenge. In addition, according to the Bahir Dar University report (2020), the effort of mechanical removal of water hyacinth using machinery has been inefficient and influenced by various factors, such as road accessibility, depth of the lake, maintenance, operational cost, and choice of dumping site. Furthermore, the interview results reveal that in both manual and mechanical removal, there is no specific disposal site that leads to regeneration of the weed in the lake zone. It has been disposed of on the lake shore and farming land. The presence of disposed and accumulated water hyacinth in agricultural land has resulted in a reduction in available farming land for the local community. Consequently, this has led to a decrease in agricultural productivity. Another management method of controlling water hyacinth quick expansion is a biological method. According to the key informant interview results, a biological method using weevils at the experimental level has not been tested thus far. It has been tested in the laboratory but not largely applied in the lake. The rearing of the weevils was performed in the laboratory, and approval was not obtained from the concerned body.

CHAPTER FIVE

5. SUMMARY CONCLUSIONS AND RECOMMENDATIONS

5.1. Summary

Water hyacinth is an invasive aquatic plant that poses a threat to the biodiversity and ecosystem of Lake Tana, Ethiopia. This study aimed to model the spatial distribution and temporal dynamics of water hyacinth using machine learning models and remote sensing data. The study used Sentinel 2A, Sentinel 1 SAR, and bioclimate variables as inputs for four machine learning models: random forest, support vector machine, boosted regression tree, and ensemble. The study also evaluated the water hyacinth management system in the lake. The results showed that random forest was the best model for water hyacinth detection, achieving high accuracy in all parameters. The study also revealed that water hyacinth coverage varied seasonally, with the highest area in the wet season and the lowest in the dry season. The study revealed that manual and machine harvesting method is the most widely used management techniques applied in the study area. This study demonstrated the potential of machine learning and remote sensing for water hyacinth monitoring and management in Lake Tana. The study suggested that the water hyacinth management system can be supported with the help of geospatial technology.

5.2. Conclusions

This study aimed to model the distribution of water hyacinth in Lake Tana, Ethiopia, using machine learning models, including random forest (RF), support vector machine (SVM), boosted regression tree (BRT), and ensemble models, and detect the spatiotemporal coverage for 2016, 2018, 2020 and 2022. The study utilized a comprehensive set of explanatory variables, including vegetation indices (NDVI, NDWI, NDAVI, SAVI), Sentinel-1 SAR (VV and VH), Sentinel-2A bands (B4, B5, B12, and B8), and precipitation and temperature data from wet and dry seasons. Based on the findings of the study, the following major conclusions are drawn.

Remote sensing is essential for water hyacinth distribution modeling since it offers valuable data and information that can be used to understand and predict how species are distributed in different habitats and environmental conditions. In this study, Sentinel-1 SAR and Sentinel-2A high-resolution optical and radar images were used to identify water hyacinth, measure vegetation density, and track changes in spatial coverage over time. It is proven that Sentinel-1 SAR and

Sentinel-2A have an immense advantage in water hyacinth detection and the measurement of spatiotemporal dynamics in Lake Tana, Ethiopia.

The results of the study confirmed that the RF model outperformed the other machine learning models across multiple evaluation metrics, including TSS, AUC/ROC, sensitivity, specificity, kappa, and COR. AUC values of 0.93 and 0.95, sensitivity values of 0.87 and 0.9, specificity values of 0.89 and 0.92, and kappa values of 0.76 and 0.82 were obtained by the RF model in the wet and dry seasons, respectively. This indicates that RF is a robust and reliable model for predicting the distribution of water hyacinth in Lake Tana. The inclusion of a diverse range of explanatory variables, such as vegetation indices, SAR data, and climatic factors, contributed to the accuracy and effectiveness of the model. The explanatory variables of B12 (16.3% and 19.7%), NDWI (14.7% and 12.4%), mean annual temperature (13.4% and 14.2%, and B5 (11.4% and 12.4%) were the most relevant in the prediction in the wet and dry seasons, respectively. Additionally, the analysis indicated that the SWIR region of the sentinel imagery, along with the NIR band, played a significant role in predicting the presence of water hyacinth in the model. The findings of the study highlighted that in all four models, the SWIR and NIR from Sentinel 2A, along with the Sentinel-1A SAR VH band and climatic variables, are crucial. Specifically, B12, B5, the annual average temperature, annual precipitation, and NDAVI consistently played crucial roles in predicting water hyacinth in the study area.

The study revealed spatial variations in the coverage across the two seasons. Water hyacinth has the highest coverage during the wet season. This finding shows that water hyacinth proliferation is influenced by seasonal variations, with wet conditions favoring its growth. The growth rate of the weed was found to be higher during the wet season due to nutrient supplies from surrounding farming areas, enhanced river flows and runoff, flooding, conducive environmental conditions, and a humid environment.

According to the results from the land use/land cover classification, the overall accuracy for 2016-2022 ranged from 0.91-0.94 in the dry season and 0.93-0.95 in the wet season. B3, B5, B11 and B12, VH, elevation, NDAVI and NDWI were the most relevant features in the land use/land cover classification of the study area. The findings demonstrated that the spatial coverage increased from 2016 to 2018 and decreased from 2020 to 2022. The highest infestation was recorded during 2018, with 6% (1851 ha) and 4.2% (1259 ha), and the lowest infestation was recorded in 2016, with 2.2%

(686.5 ha) and 2.1% (650.4 ha) of the lake in the wet and dry seasons, respectively. The decrease in spatial coverage is associated with manual and mechanical harvesting of the weed from the infested areas. In addition, the highest biomass and distribution of weeds were detected around the northern, eastern, and northeastern shores of the lake.

Regarding management, the study demonstrates that manual removal by the local community and harvesting through machinery were the main water hyacinth management strategies held in Lake Tana and its surroundings

5.3. Recommendations

Based on the findings of the study, the following recommendations are forwarded.

- It is crucial to continue monitoring the distribution and expansion of water hyacinth in Lake Tana using geospatial techniques. Regular assessments will facilitate the timely detection and response to new infestations, enabling prompt management actions.
- Given the observed seasonal variations in water hyacinth distribution, it is crucial to implement monitoring and management strategies that account for these fluctuations. Targeted interventions during the wet season can help prevent the rapid spread of water hyacinth and minimize its impact on aquatic ecosystems.
- Further research could focus on refining the set of explanatory variables used in the modeling process. Exploring additional vegetation indices, SAR parameters, and climatic variables may enhance the accuracy and precision of water hyacinth distribution models. In addition, employing high spatial, spectral and temporal resolution satellite imagery may induce more accurate and reliable predictions. Furthermore, employing other machine learning models and deep learning models can helps to accurately predict the spatial distribution of water hyacinth.
- Water Hyacinth infestation is also observed in the Rift valley region. Therefore, it is better if the models can be tested in those water bodies and thereby assist the management practices carryout.

References

Abdikan, S., Sanli, F. B., Ustuner, M., & Calò, F. (2016). Land cover mapping using sentinel-1 SAR data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *41*, 757-761.

Abebe, W. B., & Minale, A. S. (2017). Land use and watershed management practices in Lake Tana Basin. *Social and ecological system dynamics: characteristics, trends, and integration in the Lake Tana Basin, Ethiopia*, 479-521.

Abebe, W. B., G/Michael, T., Leggesse, E. S., Beyene, B. S., & Nigate, F. (2017). Climate of Lake Tana Basin. *Social and Ecological System Dynamics: Characteristics, Trends, and Integration in the Lake Tana Basin, Ethiopia*, 51-58.

Abebe, T., Awoke, B. G., & Nega, W. (2023). Spatiotemporal patterns of water hyacinth dynamics as a response to seasonal climate variability in Lake Tana, Ethiopia. *Applied Water Science*, *13*(9), 170.

Abera, M. W. (2017). Impact of water hyacinth, Eichhornia crassipes (Martius)(Pontederiaceae) in Lake Tana Ethiopia: a review. *J Aquac Res Dev*, *9*, 520.

Andrew, M. E., & Ustin, S. L. (2009). Habitat suitability modelling of an invasive plant with advanced remote sensing data. *Diversity and Distributions*, *15*(4), 627-640.

Anteneh, W., Tewabe, D., Assefa, A., Zeleke, A., Tenaw, B., & Wassie, Y. (2015). Water hyacinth coverage survey report on Lake Tana biosphere reserve. *In Technical Survey Report Series 2*.

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of applied ecology*, *43*(6), 1223-1232.

Asmare, E. (2017). Current Trend of Water Hyacinth Expansion and Its Consequence on the Fisheries around North Eastern Part of Lake Tana. *Ethiopia. J Biodivers Endanger Species*, *5*, 189.

Asmare, T., Demissie, B., Nigusse, A. G., & GebreKidan, A. (2020). Detecting spatiotemporal expansion of water hyacinth (Eichhornia crassipes) in Lake Tana, Northern Ethiopia. *Journal of the Indian Society of Remote Sensing*, *48*, 751-764.

Ayalew, M. G. (2014). Spatial coverage of water hyacinth infestation around Lake Tana, Ethiopia. *Earth Sci Int J*, *1*(1), 1-5.

Bahir Dar University. (2020). Water Hyacinth: A threat to Lake Tana.

Barreto, R., Charudattan, R., Pomella, A., & Hanada, R. (2000). Biological control of neotropical aquatic weeds with fungi. *Crop Protection*, *19*(8-10), 697-703.

Bayable, G., Cai, J., Mekonnen, M., Legesse, S. A., Ishikawa, K., Imamura, H., & Kuwahara, V. S. (2023). Detection of Water Hyacinth (Eichhornia crassipes) in Lake Tana, Ethiopia, Using Machine Learning Algorithms. *Water*, *15*(5), 880.

Belayneh, A., Workie, W. S., & Gashaye, D. (2022). Current Trend of Water Hyacinth Expanation and Investigation of Possible Cause for Water Hyacinth Using Remote Sensing in the Case Study of Lake Tana, Ethiopia.

Berihun, D. (2017). Lake Tana Subbasin's Economy and the Role of Natural Resources. Social and Ecological System Dynamics: Characteristics, Trends, and Integration in the Lake Tana Basin, Ethiopia, 347-355.

Bharatkar, P. S., & Patel, R. (2013). Approach to accuracy assessment tor RS image classification techniques. *International Journal of Scientific & Engineering Research*, *4*(12), 79-86.

Birara, H., Pandey, R. P., & Mishra, S. K. (2018). Trend and variability analysis of rainfall and temperature in the Tana basin region, Ethiopia. *Journal of Water and Climate Change*, 9(3), 555-569.

Bogale, A. (2020). Review, impact of land use/cover change on soil erosion in the Lake Tana Basin, Upper Blue Nile, Ethiopia. *Applied Water Science*, 10(12), 235.

Bradley, B. A. (2014). Remote detection of invasive plants: a review of spectral, textural and phenological approaches. *Biological invasions*, *16*, 1411-1425.

Brooks, C. N., Grimm, A. G., Marcarelli, A. M., & Dobson, R. J. (2019). Multiscale collection and analysis of submerged aquatic vegetation spectral profiles for Eurasian watermilfoil detection. *Journal of Applied Remote Sensing*, *13*(3), 037501-037501.

Buermann, W., Saatchi, S., Smith, T. B., Zutta, B. R., Chaves, J. A., Milá, B., & Graham, C. H. (2008). Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeography*, *35*(7), 1160-1176.

Cai, J., Jiao, C., Mekonnen, M., Legesse, S. A., Ishikawa, K., Wondie, A., & Sato, S. (2023). Water hyacinth infestation in Lake Tana, Ethiopia: A review of population dynamics. *Limnology*, *24*(1), 51-60.

Center, T. D., & Dray Jr, F. A. (2010). Bottom-up control of water hyacinth weevil populations: do the plants regulate the insects? *Journal of Applied Ecology*, *47*(2), 329-337.

Chen, J., Chen, S., Fu, R., Wang, C., Li, D., Jiang, H., ... & Mei, Y. (2021). Simulation of water hyacinth growth area based on multi-source geographic information data: An integrated method of WOE and AHP. *Ecological Indicators*, *125*, 107574.

Cheruiyot, E. K., Mito, C., Menenti, M., Gorte, B., Koenders, R., & Akdim, N. (2014). Evaluating MERIS-based aquatic vegetation mapping in Lake Victoria. *Remote Sensing*, *6*(8), 7762-7782.

Cho, H. J., Kirui, P., & Natarajan, H. (2008). Test of multi-spectral vegetation index for floating and canopy-forming submerged vegetation. *International journal of environmental research and public health*, *5*(5), 477-483.

Cilliers, C. J., Hill, M. P., Ogwang, J. A., & Ajuonu, O. (2003). Aquatic weeds in Africa and their control. In *Biological control in IPM systems in Africa* (pp. 161-178). Wallingford UK: CABI Publishing.

Damtew, Y. T., Verbeiren, B., Awoke, A., & Triest, L. (2021). Satellite imageries and field data of macrophytes reveal a regime shift of a tropical lake (Lake Ziway, Ethiopia). *Water*, *13*(4), 396.

Damtie, Y. A., & Mengistu, D. A. (2022). Water hyacinth (Eichhornia crassipes (Mart.) Solms) impacts on land-use land-cover change across northeastern Lake Tana. *Journal of the Indian Society of Remote Sensing*, *50*(6), 975-986.

Damtie, Y. A., Berlie, A. B., Gessese, G. M., & Ayalew, T. K. (2022). Characterization of water hyacinth (Eichhornia crassipes (Mart.) Solms) biomass in Lake Tana, Ethiopia. *All Life*, *15*(1), 1126-1140.

Damtie, Y. A., Mengistu, D. A., & Meshesha, D. T. (2021). Spatial coverage of water hyacinth (Eichhornia crassipes (Mart.) Solms) on Lake Tana and associated water loss. *Heliyon*, *7*(10).

Datta, A., S. Maharaj, G. N. Prabhu, D. Bhowmik, A. Marino, V. Akbari, S. Rupavatharam et al. "Monitoring the spread of water hyacinth (*Pontederia crassipes*): Challenges and future developments. *Front Ecol Evol*." (2021).

Dechassa, N., & Abate, B. (2020). Current status of water hyacinth (Eichhornia crassipes) in Ethiopia: achievements, challenges and prospects: a review. *J. Environ. Earth Sci*, *10*(12), 1-13.

Demissie, S., & Abebe, A. (2017). Gender and Rural Livelihood in the Lake Tana Basin. Social and Ecological System Dynamics: Characteristics, Trends, and Integration in the Lake Tana Basin, Ethiopia, 293-311.

Dersseh, M. G., Kibret, A. A., Tilahun, S. A., Worqlul, A. W., Moges, M. A., Dagnew, D. C., ... & Melesse, A. M. (2019a). Potential of water hyacinth infestation on lake Tana, Ethiopia: a prediction using a GIS-based multi-criteria technique. *Water*, *11*(9), 1921.

Dersseh, M. G., Melesse, A. M., Tilahun, S. A., Abate, M., & Dagnew, D. C. (2019b). Water hyacinth: review of its impacts on hydrology and ecosystem services—lessons for management of Lake Tana. *Extreme hydrology and climate variability*, 237-251.

Dersseh, M. G., Tilahun, S. A., Worqlul, A. W., Moges, M. A., Abebe, W. B., Mhiret, D. A., & Melesse, A. M. (2020). Spatial and temporal dynamics of water hyacinth and its linkage with lake-level fluctuation: Lake Tana, a sub-humid region of the Ethiopian highlands. *Water, 12*(5), 1435.

Dogliotti, A. I., Gossn, J. I., Vanhellemont, Q., & Ruddick, K. G. (2018). Detecting and quantifying a massive invasion of floating aquatic plants in the Río de la Plata turbid waters using high spatial resolution ocean color imagery. *Remote Sensing*, *10*(7), 1140.

Durgesh, K. S., & Lekha, B. (2010). Data classification using support vector machine. *Journal of theoretical and applied information technology*, *12*(1), 1-7.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution, and systematics*, *40*, 677-697.

Enyew, B. G., Assefa, W. W., & Gezie, A. (2020). Socioeconomic effects of water hyacinth (Eichhornia crassipes) in lake tana, north western Ethiopia. *PloS one*, *15*(9), e0237668.

Feilhauer, H., He, K. S., & Rocchini, D. (2012). Modeling species distribution using niche-based proxies derived from composite bioclimatic variables and MODIS NDVI. *Remote Sensing*, *4*(7), 2057-2075.

Firehun, G. (2017). *Management of water hyacinth (Eichhornia crassipes [Mart.] Solms) using bioagents in the Rift Valley of Ethiopia* (Doctoral dissertation, Wageningen University, Wageningen, the Netherlands).

Gaikwad, R. P., & Gavande, S. (2017). Major factors contributing growth of water hyacinth in natural water bodies. *International Journal of Engineering Research*, *6*(6), 304-306.

Gao, B. C. (1995). Normalized difference water index for remote sensing of vegetation liquid water from space. In *Imaging spectrometry* (Vol. 2480, pp. 225-236). SPIE.

Gerardo, R., & de Lima, I. P. (2022a). Assessing the potential of Sentinel-2 data for tracking invasive water hyacinth in a river branch. *Journal of Applied Remote Sensing*, *16*(1), 014511-014511.

Gerardo, R., & de Lima, I. P. (2022b). Monitoring duckweeds (Lemna minor) in small rivers using sentinel-2 satellite imagery: application of vegetation and water indices to the Lis River (Portugal). *Water*, *14*(15), 2284.

Gezie, A., Assefa, W. W., Getnet, B., Anteneh, W., Dejen, E., & Mereta, S. T. (2018). Potential impacts of water hyacinth invasion and management on water quality and human health in Lake Tana watershed, Northwest Ethiopia. *Biological Invasions*, *20*, 2517-2534.

Ghodousi, M., Sadeghi-Niaraki, A., Rabiee, F., & Choi, S. M. (2020). Spatial-temporal analysis of point distribution pattern of schools using spatial autocorrelation indices in Bojnourd city. *Sustainability*, *12*(18), 7755.

Ghoussein, Y., Abou Hamdan, H., Fadel, A., Coudreuse, J., Nicolas, H., Faour, G., & Haury, J. (2023). Biology and ecology of Pontederia crassipes in a Mediterranean river in Lebanon. *Aquatic Botany*, 103681.

Gitelson, A. A., Merzlyak, M. N., & Chivkunova, O. B. (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry and photobiology*, *74*(1), 38-45.

Gobeyn, S., Mouton, A. M., Cord, A. F., Kaim, A., Volk, M., & Goethals, P. L. (2019). Evolutionary algorithms for species distribution modelling: A review in the context of machine learning. *Ecological Modelling*, *392*, 179-195.

Gottschalk, T. K., Huettmann, F., & Ehlers, M. (2005). Thirty years of analysing and modelling avian habitat relationships using satellite imagery data: a review. *International Journal of Remote Sensing*, *26*(12), 2631-2656.

Graham, C. H., Elith, J., Hijmans, R. J., Guisan, A., Townsend Peterson, A., Loiselle, B. A., & NCEAS Predicting Species Distributions Working Group. (2008). The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, *45*(1), 239-247.

Greenfield, B. K., Blankinship, M., & McNabb, T. J. (2006). Control costs, operation, and permitting issues for non-chemical plant control: case studies in the San Francisco Bay-Delta Region, California. *Journal of Aquatic Plant Management*, *44*(1), 40-49.

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological modelling*, *135*(2-3), 147-186.

Hallman, T. A., & Robinson, W. D. (2020). Comparing multi-and single-scale species distribution and abundance models built with the boosted regression tree algorithm. *Landscape ecology*, *35*, 1161-1174.

Hardy, A., Ettritch, G., Cross, D. E., Bunting, P., Liywalii, F., Sakala, J., ... & Thomas, C. J. (2019). Automatic detection of open and vegetated water bodies using Sentinel 1 to map African malaria vector mosquito breeding habitats. *Remote Sensing*, *11*(5), 593.

He, K. S., Bradley, B. A., Cord, A. F., Rocchini, D., Tuanmu, M. N., Schmidtlein, S., ... & Pettorelli, N. (2015). Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation*, *1*(1), 4-18.

Hernandez, P. A., Franke, I., Herzog, S. K., Pacheco, V., Paniagua, L., Quintana, H. L., ... & Young, B. E. (2008). Predicting species distributions in poorly-studied landscapes. *Biodiversity and conservation*, *17*, 1353-1366.

Hestir, E. L., Brando, V. E., Bresciani, M., Giardino, C., Matta, E., Villa, P., & Dekker, A. G. (2015). Measuring freshwater aquatic ecosystems: The need for a hyperspectral global mapping satellite mission. *Remote Sensing of Environment*, *167*, 181-195.

Hill, M. P., & Coetzee, J. A. (2008). Integrated control of water hyacinth in Africa 1. *EPPO bulletin*, *38*(3), 452-457.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of environment*, *83*(1-2), 195-213.

Huete, A. R. (1988). A Soil-Adjusted Vegetation Index (SAVI). Remote Ssensing of Environment, 25, 295-309.

Ingwani, E., Gumbo, T., & Gondo, T. (2010). The general information about the impact of water hyacinth on Aba Samuel Dam, Addis Ababa, Ethiopia: Implications for ecohydrologists. *Ecohydrology & Hydrobiology*, *10*(2-4), 341-345.

Janssens, N., Schreyers, L., Biermann, L., Van Der Ploeg, M., Bui, T. K. L., & Van Emmerik, T. (2022). Rivers running green: water hyacinth invasion monitored from space. *Environmental Research Letters*, *17*(4), 044069.

Jiang, H., Wang, M., Hu, H., & Xu, J. (2021). Evaluating the performance of Sentinel-1A and sentinel-2 in small waterbody mapping over urban and mountainous regions. *Water*, *13*(7), 945.

Julien, M. H. (2000). Biological control of water hyacinth with arthropods: a review to 2000. In *Aciar Proceedings* (pp. 8-20). ACIAR; 1998.

Julien, M. (2008). Plant biology and other issues that relate to the management of water hyacinth: a global perspective with focus on Europe 1. *EPPO bulletin*, *38*(3), 477-486.

Kariyawasam, C. S., Kumar, L., & Ratnayake, S. S. (2021). Potential distribution of aquatic invasive alien plants, Eichhornia crassipes and Salvinia molesta under climate change in Sri Lanka. *Wetlands Ecology and Management*, *29*, 531-545.

Keebine, G. L. (2019). Mapping and monitoring the spatial distribution of Eichhornia crassipes (water hyacinth) in the Hartbeespoort dam, South Africa, using remote sensing datas (*Doctoral dissertation, Faculty of Science, University of the Witwatersrand, Johannesburg*).

Kothari, C.R. (2012). Research Methodology: Methods and Techniques. New Age International, New Delhi.

Kumlachew, Y. Z., Tilahun, S. A., Cherie, F. F., Akale, A. T., Kibret, E. A., Alemie, N. A., & Animut, M. (2023). Quantifying flow rate using stage-discharge rating curve and Scs runoff equation on upland watershed of Lake Tana Sub Basin, Ethiopia. *Sustainable Water Resources Management*, 9(2), 47.

Laranjeira, C. M., & Nadais, G. (2008). Eichhornia crassipes control in the largest Portuguese natural freshwater lagoon 1. *EPPO bulletin*, *38*(3), 487-495.

Li, X., & Wang, Y. (2013). Applying various algorithms for species distribution modelling. *Integrative zoology*, *8*(2), 124-135.

Liaw, A., & Wiener, M. (2002). Classification and regression by random Forest. *R news*, *2*(3), 18-22.

Luan, J., Zhang, C., Xu, B., Xue, Y., & Ren, Y. (2020). The predictive performances of random forest models with limited sample size and different species traits. *Fisheries Research*, *227*, 105534.

Ma, Z., & Zhou, G. (2018). Canopy Modeling of Aquatic Vegetation: Construction of Submerged Vegetation Index. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, 1253-1258.

Makhanu, K. S. (1997). Impact of water hyacinth on Lake Victoria.

Mallya, G., Mjema, P., & Ndunguru, J. (2001). Water hyacinth control through integrated weed management strategies in Tanzania. *Proc. ACIAR. Biological and Integrated Control of Water hyacinth*, *102*, 120-122.

Masifwa, W. F., Twongo, T., & Denny, P. (2001). The impact of water hyacinth, Eichhornia crassipes (Mart) Solms on the abundance and diversity of aquatic macroinvertebrates along the shores of northern Lake Victoria, Uganda. *Hydrobiologia*, *452*, 79-88.

Mequanent, D., Mingist, M., Getahun, A., & Anteneh, W. (2021). Impact of irrigation practices on Gilgel Abay, Ribb and Gumara fisheries, Tana Sub-Basin, Ethiopia. Heliyon, 7 (3), e06523.

Miller, J. (2010). Species distribution modeling. *Geography Compass*, *4*(6), 490-509.

Mouta, N., Silva, R., Pinto, E. M., Vaz, A. S., Alonso, J. M., Gonçalves, J. F., ... & Vicente, J. R. (2023). Sentinel-2 Time Series and Classifier Fusion to Map an Aquatic Invasive Plant Species along a River—The Case of Water-Hyacinth. *Remote Sensing*, *15*(13), 3248.

Mucheye, T., Haro, S., Papaspyrou, S., & Caballero, I. (2022). Water Quality and Water Hyacinth Monitoring with the Sentinel-2A/B Satellites in Lake Tana (Ethiopia). Remote Sens. 2022, 14, 4921.

Mukarugwiro, J. A., Newete, S. W., Adam, E., Nsanganwimana, F., Abutaleb, K. A., & Byrne, M. J. (2019). Mapping distribution of water hyacinth (Eichhornia crassipes) in Rwanda using multispectral remote sensing imagery. *African Journal of Aquatic Science, 44*(4), 339-348.

Mund, J. P., Murach, D., & Parplies, A. (2014). Monitoring and quantification of floating biomass on tropical water bodies. *Geospatial Innov. Soc*, 67-76.

Naimi, B., Skidmore, A. K., Groen, T. A., & Hamm, N. A. (2011). Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. *Journal of biogeography*, *38*(8), 1497-1509.

Nang'alelwa, M. (2008). The environmental and socio-economic impacts of Eichhornia crassipes in the Victoria Falls/Mosi-oa-Tunya World Heritage Site, Livingstone, Zambia 1. *EPPO bulletin*, *38*(3), 470-476.

New, M., Lister, D., Hulme, M., & Makin, I. (2002). A high-resolution data set of surface climate over global land areas. *Climate research*, *21*(1), 1-25.

Noble, W. S. (2006). What is a support vector machine? *Nature biotechnology*, *24*(12), 1565-1567.

Osborne, P. E., Alonso, J. C., & Bryant, R. G. (2001). Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. *Journal of applied ecology*, *38*(2), 458-471.

Otieno, D., Nyaboke, H., Nyamweya, C. S., Odoli, C. O., Aura, C. M., & Outa, N. O. (2022). Water hyacinth (Eichhornia crassipes) infestation cycle and interactions with nutrients and aquatic biota in Winam Gulf (Kenya), Lake Victoria. Lakes & Reservoirs: Research & Management, 27(1), e12391.

Ouma, Y. O., Shalaby, A., & Tateishi, R. (2005). Dynamism and abundance of water hyacinth in the Winam Gulf of Lake Victoria: evidence from remote sensing and seasonal-climate data. *International journal of environmental studies*, *62*(4), 449-465.

Pádua, L., Antão-Geraldes, A. M., Sousa, J. J., Rodrigues, M. Â., Oliveira, V., Santos, D., ... & Castro, J. P. (2022). Water hyacinth (Eichhornia crassipes) detection using coarse and high-resolution multispectral data. *Drones*, *6*(2), 47.

Pan, X., Geng, Y., Zhang, W., Li, B., & Chen, J. (2006). The influence of abiotic stress and phenotypic plasticity on the distribution of invasive Alternanthera philoxeroides along a riparian zone. *Acta Oecologica*, *30*(3), 333-341.

Parra, J. L., Graham, C. C., & Freile, J. F. (2004). Evaluating alternative data sets for ecological niche models of birds in the Andes. *Ecography*, *27*(3), 350-360.

Pejchar, L., & Mooney, H. A. (2009). Invasive species, ecosystem services and human well-being. *Trends in ecology & evolution*, *24*(9), 497-504.

Phiri, G., & Navarro, L. (Eds.). (2000). *Water Hyacinth in Africa and the Middle East: A Survey of Problems and Solutions*. International Development Research Centre.

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote sensing of environment*, *48*(2), 119-126.

Rezaei, R., & Şengül, H. (2018). Development of generalized additive models (GAMs) for Salmo rizeensis endemic to north-eastern streams of Turkey. *Turkish Journal of Fisheries and Aquatic Sciences*.

Rezene, F. (2005). Water hyacinth (Eichhornia crassipes): A review of its weed status in Ethiopia. *Arem*, *6*, 105-111.

Robles, W., Madsen, J. D., & Wersal, R. M. (2015). Estimating the biomass of waterhyacinth (Eichhornia crassipes) using the normalized difference vegetation index derived from simulated landsat 5 TM. *Invasive Plant Science and Management*, *8*(2), 203-211.

Rodrigues, M., & De la Riva, J. (2014). An insight into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling & Software*, *57*, 192-201.

Rollins, M. G., Keane, R. E., & Parsons, R. A. (2004). Mapping fuels and fire regimes using remote sensing, ecosystem simulation, and gradient modeling. *Ecological Applications*, *14*(1), 75-95.

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ*, *351*(1), 309.

Senay, S. D., Worner, S. P., & Ikeda, T. (2013). Novel three-step pseudo-absence selection technique for improved species distribution modelling. *PloS one*, *8*(8), e71218.

Shiferaw, H., Bewket, W., & Eckert, S. (2019). Performances of machine learning algorithms for mapping fractional cover of an invasive plant species in a dryland ecosystem. *Ecology and evolution*, *9*(5), 2562-2574.

Shiferaw, W., Demissew, S., & Bekele, T. (2018). Invasive alien plant species in Ethiopia: ecological impacts on biodiversity a review paper. *Int J Mol Biol*, *3*(4), 171-178.

Simpson, M. D., Akbari, V., Marino, A., Prabhu, G. N., Bhowmik, D., Rupavatharam, S., ... & Hunter, P. D. (2022). Detecting water hyacinth infestation in kuttanad, india, using dual-pol Sentinel-1 SAR imagery. *Remote Sensing*, *14*(12), 2845.

Simpson, M., Marino, A., Prabhu, G. N., Bhowmik, D., Rupavatharam, S., Datta, A., ... & Maharaj, S. (2020). Monitoring water hyacinth in Kuttanad, India using Sentinel-1 SAR data. In *2020 IEEE India Geoscience and Remote Sensing Symposium (InGARSS)* (pp. 13-16). IEEE.

Singh, G., Reynolds, C., Byrne, M., & Rosman, B. (2020). A remote sensing method to monitor water, aquatic vegetation, and invasive water hyacinth at national extents. *Remote Sensing*, *12*(24), 4021.

Somodi, I., Lepesi, N., & Botta‐Dukát, Z. (2017). Prevalence dependence in model goodness measures with special emphasis on true skill statistics. *Ecology and evolution*, *7*(3), 863-872.

Stave, K., Goshu, G., & Aynalem, S. (2017). *Social and ecological system dynamics*. Springer Verlag.

Tasew, A. Y., & Wendimagegnehu, W. W. (2022). Occurrence, Diversity and Potential of Fungal Pathogens Associated with Water Hyacinth (Eichhornia crassipes [Mart.] Solms) for Sustainable Biological Management in Ethiopia: A Review. *Adv Environ Stud*, *6*(1), 497-503.

Téllez, T. R., López, E. M. D. R., Granado, G. L., Pérez, E. A., López, R. M., & Guzmán, J. M. S. (2008). The water hyacinth, Eichhornia crassipes: an invasive plant in the Guadiana River Basin (Spain). *Aquatic Invasions*, *3*(1), 42-53.

Tewabe, D. (2015). Preliminary survey of water hyacinth in Lake Tana, Ethiopia. *Global Journal of Allergy*, *1*(1), 013-018.

Tewabe, D., Asmare, E., Zelalem, W., & Mohamed, B. (2017). Identification of impacts, some biology of water hyacinth (Eichhornia crassipes) and its management options in Lake Tana, Ethiopia. *Net Journal of Agricultural Science*, *5*(1), 8-15.

Thakuri, S. S., Shrestha, P., Deuba, M., Shah, P., Bhandari, O. P., & Shrestha, S. (2019). Potential habitat modeling of water hyacinth in lakes of Nepal using Maxent algorithm. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*, 103-110.

Thamaga, K. H., & Dube, T. (2018a). Testing two methods for mapping water hyacinth (Eichhornia crassipes) in the Greater Letaba river system, South Africa: discrimination and mapping potential of the polar-orbiting Sentinel-2 MSI and Landsat 8 OLI sensors. *International journal of remote sensing*, *39*(22), 8041-8059.

Thamaga, K. H., & Dube, T. (2018b). Remote sensing of invasive water hyacinth (Eichhornia crassipes): A review on applications and challenges. *Remote Sensing Applications: Society and Environment*, *10*, 36-46.

Thamaga, K. H., & Dube, T. (2019). Understanding seasonal dynamics of invasive water hyacinth (Eichhornia crassipes) in the Greater Letaba river system using Sentinel-2 satellite data. *GIScience & Remote Sensing*, *56*(8), 1355-1377.

Tibebe, D., Kassa, Y., Melaku, A., & Lakew, S. (2019). Investigation of spatio-temporal variations of selected water quality parameters and trophic status of Lake Tana for sustainable management, Ethiopia. *Microchemical Journal*, 148, 374-384.

Twongo, T., Bugenyi, F., & Wanda, F. (2009). The potential for further proliferation of water hyacinth in Lakes Victoria, Kyoga nad Kwania and some urgent aspects for research'. African Journal of Tropical Hydrobiology and Fisheries, 6(1).

Van Wyk, E., & Van Wilgen, B. W. (2002). The cost of water hyacinth control in South Africa: a case study of three options. *African Journal of Aquatic Science*, *27*(2), 141-149.

Vaughan, I. P., & Ormerod, S. J. (2005). The continuing challenges of testing species distribution models. *Journal of applied ecology*, *42*(4), 720-730.

Verma, R., Singh, S. P., & Raj, K. G. (2003). Assessment of changes in water-hyacinth coverage of water bodies in northern part of Bangalore city using temporal remote sensing data. *Current Science*, 795-804.

Villa, P., Bresciani, M., Braga, F., & Bolpagni, R. (2013). Mapping aquatic vegetation through remote sensing data: A comparison of vegetation indices performances. In *Proceedings of the 6th EARSeL Workshop on Remote Sensing of the Coastal Zone, Matera, Italy* (pp. 10-15).

Villamagna, A. M., & Murphy, B. R. (2010). Ecological and socio-economic impacts of invasive water hyacinth (Eichhornia crassipes): a review. *Freshwater biology*, *55*(2), 282-298.

Wang, H., Li, Y., Zeng, S., Cai, X., Bi, S., Liu, H., ... & Zhang, Y. (2022). Recognition of aquatic vegetation above water using shortwave infrared baseline and phenological features. *Ecological Indicators*, *136*, 108607.

Wang, Z., & Yan, S. H. (2017). Impact of water hyacinth on aquatic environment in phytoremediation of eutrophic lakes. *Water Hyacinth: Environmental Challenges, Management and Utilization*, 204-249.

Wanyonyi, N. B. (2019). Monitoring The Distribution of Water Hyacinth, Using Remotely Sensed Data: Case Study of Lake Victoria, Kenya (Doctoral dissertation, University of Nairobi).

Wilkinson, D. P., Golding, N., Guillera-Arroita, G., Tingley, R., & McCarthy, M. A. (2021). Defining and evaluating predictions of joint species distribution models. *Methods in Ecology and Evolution*, *12*(3), 394-404.

Wimalarathne, H. D. A., & Perera, P. C. D. (2019). Potentials of water hyacinth as livestock feed in Sri Lanka.

Worqlul, A. W., Ayana, E. K., Dile, Y. T., Moges, M. A., Dersseh, M. G., Tegegne, G., & Kibret, S. (2020). Spatiotemporal dynamics and environmental controlling factors of the Lake Tana water hyacinth in Ethiopia. *Remote Sensing*, *12*(17), 2706.

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International journal of remote sensing*, *27*(14), 3025-3033.

Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of sensors*, *2017*.

Yan, S. H., Song, W., & Guo, J. Y. (2017). Advances in management and utilization of invasive water hyacinth (Eichhornia crassipes) in aquatic ecosystems–a review. *Critical reviews in biotechnology*, *37*(2), 218-228.

Yitbarek, M., Belay, M., & Bazezew, A. (2019). Determinants of manual control of water hyacinth expansion over the Lake Tana, Ethiopia. *AFRREV STECH: An International Journal of Science and Technology*, *8*(1), 1-14.

Yoon, S., & Lee, W. H. (2023). Application of true skill statistics as a practical method for quantitatively assessing CLIMEX performance. *Ecological Indicators*, *146*, 109830.

Yu, H., Cooper, A. R., & Infante, D. M. (2020). Improving species distribution model predictive accuracy using species abundance: Application with boosted regression trees. *Ecological Modelling*, *432*, 109202.

Zhang, Z. H., & Guo, J. Y. (2017). Biology of water hyacinth. In *Water Hyacinth* (pp. 15-43). CRC Press.

Zhang, J., & Li, S. (2017). A review of machine learning based Species' distribution modelling. In *2017 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII)* (pp. 199-206). IEEE.

Zhang, T., Su, J., Liu, C., Chen, W. H., Liu, H., & Liu, G. (2017). Band selection in sentinel-2 satellite for agriculture applications. In *2017 23rd international conference on automation and computing (ICAC)* (pp. 1-6). IEEE.

Appendix 1

Key informant interview checklist

1. Sex_______
2. Age_______
3. Name______
4. Position______
5. Work Experience ____________
6. Institution ____________________ Address Tel: ___________________

Objectives: The interview is intended to acquire information required in the study entitled "Water hyacinth distribution modeling using geospatial technology: the case of Lake Tana, Ethiopia", where you have been selected to help provide the information. Your experience, knowledge and suggestions are very important.

Therefore, I am kindly requesting to answer questions pertaining to this study.

<u>Questions</u>

1. What type of water hyacinth management techniques have been applied?
2. What challenges have you experienced in managing water hyacinth?
3. What have been the most successful strategies for managing water hyacinths?
4. What has been the biggest obstacle to the successful management of water hyacinths?
5. Why is water hyacinth infestation high during the wet season?
6. Do you think the management practices reduce the rate of expansion?

Appendix 2

Classification Accuracy

I. Overall Accuracy

Overall accuracy (%) = (Correctly classified pixels/Total number of GCPs)

II. User Accuracy

User accuracy (%) = (Correctly classified pixels/Classified total pixels)

III. Producer Accuracy

Producer's accuracy (%) = (Correctly classified pixels/Reference total pixels)

IV. Kappa Coefficient

The kappa coefficient (K) can be computed as follows:

$K = Po – Pc/1 − Pc$

$P0$ = proportion of units that agree, = overall accuracy

Pc = proportion of units for expected chance agreement

If the kappa value is < 0.4, the classification is poor; if it is between 0.4-0.75, a good kappa value; and if K > 0.75, an excellent kappa value.