RISK MAPPING OF VISCERAL LEISHMANIASIS INFECTIONS IN WEST POKOT, KENYA: CHARACTERISATION OF LOCAL ENVIRONMENTAL RISK FACTORS

SANDRA CHEPKEMBOI KOSGEI
Enschede, The Netherlands, June 2024

Thesis submitted to the Faculty of Geo-information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Spatial Engineering

SUPERVISORS:
Dr. ir. P.W.M. Augustijn
Dr. S. Amer

THESIS ASSESSMENT BOARD:
Prof.dr. R. Zurita Milla (chair)
Drs. B.J. Köbben

Disclaimer

This document describes work undertaken as part of a programme of study at the Faculty of Geo-information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

VL is a fatal neglected tropical disease, and the number of infections in Kenya has increased in recent years. Surveillance efforts in endemic areas can be improved through risk maps and knowledge of environmental risk associated with the vector and guide the development and placement of vector management tools.

The anthroponotic transmission of VL is affected by the proximity to sandfly breeding habitats, population densities of the vector, abundance of plant sugar sources for the sandflies in the surroundings and presence of ample blood sources for sandfly females including host individuals. Understanding these dynamics requires detailed characterisation of the environment, particularly at a fine spatial resolution. This study aimed to model and predict VL in West Pokot at a fine spatial scale.

We identified vector habitats for sandfly *P. martini* as termite mounds, animal sheds and the banks of seasonal rivers. Using very high-resolution worldview imagery we trained a deep learning model that was able to clearly distinguish animal sheds. However, the method was not successful with termite mounds.

We extracted environmental variables at very high resolution. After modelling, NDVI had the highest contribution. We were unable to incorporate humidity, high-resolution rainfall data, and acacia trees which are crucial for vector survival.

We simulated potential dispersal points for infection cases in Kacheliba using the BAM framework to generate input data. We ran a maxent model to predict risk for Visceral Leishmaniasis and the best score was an AUC of 0.805.
.

ACKNOWLEDGEMENT

I would like to express my deepest gratitude and appreciation to my supervisors, Dr Ellen-Wien and Dr Amer for their guidance, flexibility, and support throughout my research process. Their expert advice, constructive feedback and encouragement were pivotal for my thesis and improved the outcome of my work.

I would also like to thank Norbert Van Dijk for his input and advise during my thesis phase. His feedback and valuable guidance helped me refine my research. Additionally, I would like to thank the staff at Faculty ITC, my thesis board for their expert knowledge and my classmates from spatial engineering. I am happy that I shared the journey with them and for all the help throughout my two-year master's.

Lastly, I would like to thank God, my friends and family. The support from my family has been instrumental for me to complete the course. Their supporting words inspired me to keep going.

Contents

List of Figures

List of tables

ACRONYMS

VL	-	Visceral Leishmaniasis
ENM	-	Ecological Niche Model
SDM	-	Species Distribution Model
CNN	-	Convolutional Neural network
R-CNN`	-	Region-based Convolutional Neural Network
DEM	-	Digital Elevation model
KNBS	-	Kenya National Bureau of Statistics
BAM	-	Biotic Abiotic Movement
NDVI	-	Normalized Difference Vegetation Index
GEE	-	Google Earth Engine
LST	-	Land Surface Temperature.
RF	-	Random Forest
BAM	-	Biotic Abiotic Movement
VIF	-	Variance Inflation Factor
AUC	-	Area Under the receiver operating Characteristic curve

1. Introduction

1.1. Background

Visceral leishmaniasis (VL) is a parasitic infection caused by *Leishmania donovani or Leishmania infantum,* and its transmission vector is the female sandfly (Van Dijk et al. 2023). It is a neglected tropical disease (NTD), and an infected patient will present symptoms such as fever, splenomegaly, weight loss, anaemia, coughing and body weakness. If not treated, VL can result in death (Ministry of Health Kenya 2017). Globally, VL infections are concentrated in a few countries, with 95% of cases occurring in just ten countries. It is endemic to parts of East Africa (Kenya, Uganda, Somalia, Sudan, and Ethiopia), and the region has a history of severe epidemics (D. Elnaiem, 2011).

In Kenya, endemic areas are low-lying, semi-arid, and often remote, with poor access to health facilities. Outbreaks put pressure on the already fragile existing healthcare facilities in these areas. Additionally, VL infection cases are underreported due to inadequate surveillance systems (Mewara et al. 2022). Populations at risk also experience poverty and insecurity from constant conflict (Alvar et al. 2021). One of these regions is West Pokot. It is located to the West of Kenya on a semi-arid plateau. The region is inhabited by the Pokot community, which lives in clustered communities, grows crops, and keeps livestock (Mueller et al. 2014).

VL is anthroponotic in Africa, meaning humans are the reservoirs of the *Leishmania donovani* parasite (Alves et al. 2018). The transmission begins when a female phlebotomine sandfly becomes infected with the parasite while feeding on an infected person's blood. Subsequently, the infection is transmitted when the infected sandfly bites and injects the *Leishmania* parasite into the next human host (Ministry of Health Kenya 2017). The main vector sandflies in East Africa are *Phlebotomus orientalis* and *Phlebotomus martini* (Mueller et al. 2014), and the principal vector in Kenya is *P. martini* (Van Dijk et al. 2023). Control and management of VL in Kenya is still challenging, and current methods have not been effective as case numbers remain persistent (Mewara et al. 2022).

In previous control studies carried out in endemic areas in Kenya, the risk of contracting the disease increased with exposure to the sandfly vector (Kolaczinski et al. 2008). Vector dynamics such as density, feeding and resting behaviour affect transmission. Further, the density of these vectors is affected by ecological factors such as land cover, vegetation, and climatic conditions (Ministry of Health Kenya 2017). Since transmission zones are affected by the ecological niche of the sandfly, knowledge of the sandfly habitat, behaviour, and favourable environmental conditions can be useful in predicting and controlling VL occurrence.

Sandflies have various habitats, such as cracks and crevices of soils, caves, termite mounds, human habitats, animal sheds, and *Acacia* trees and each species has its preferences(Hassaballa, Torto, et al. 2021). The sandfly *P. martini* is thought to lay its eggs in termite mounds on ventilation openings (Van Dijk et al. 2023). In field etymological studies of sandflies in Baringo County Kenya, the sandfly *P. martini* was collected in animal sheds and termite mounds. Additionally, the female sandfly requires a blood meal for the maturation of its eggs and the primary source for this was cattle, but also, humans, and dogs, which may explain why it was found in animal sheds. The sandfly also feeds on plants in the *Fabaceae* family, of which *Acacia* is the most dominant (Hassaballa et al., 2021a), which may explain the link between sandflies and acacia trees.

The behaviour of sandflies is affected by environmental changes brought about by changing weather patterns and consequently affects their distribution, development, and interaction with the protozoa *Leishmania* (Capucci et al. 2023). In Ethiopia, *P. martini* was found at low altitudes ranging between 500- 1800m asl (Aklilu et al. 2023). In other studies, for VL in India and China (different vector species responsible for transmission), the predictor variables identified were humidity, NDVI, high temperatures, and rainfall (Jiang et al. 2021; Sardar et al. 2020).

The risk of VL infection is determined by interactions between humans and the vector sandfly, and some socioeconomic activities contribute to these interactions. Pastoralism is practised in the Pokot community, and boys and young men herd cattle (Mueller et al. 2014). During dry seasons, herders and their livestock take shade from the sun under big ever-green trees, which also serve as breeding sites for the vectors(Abdullahi et al. 2022). Working or playing near acacia trees or termite mounds during the dry season increases the risk of infection (Alvar et al. 2021). The migration of pastoralists and refugees with their livestock in endemic areas can spread infection to previously unaffected areas (Mewara et al. 2022).

Several modelling approaches have been used to describe the spatiotemporal distribution of diseases and identify key environmental variables related to the spread of VL. Species distribution models and spatial analysis techniques offer a valuable approach to identifying environmental patterns of *Leishmaniasis* vectors. This information is particularly useful in areas where data collection is limited (Rajabi et al. 2016).

Predictive models and species distribution models estimate the relationships between species occurrence at a location and the environmental characteristics of those locations (Elith et al. 2011). These methods range from statistical, such as generalised linear models (GLM) and generalised additive models (GAM), to machine learning methods, such as Maxent, random forest (RF), support vector machines (SVM) and boosted regression trees (BRT) (Grimmett, Whitsed, and Horta 2020).

A defining factor for models is the type of input data they use. In instances where species data has been collected, and presence and absence have been established, GLM, GAM, or an ensemble of regression trees i.e., RF are used. Data collected from systematic biological

surveys to establish absence and presence are rare, and most species records are available as presence-only data. Maxent is one of the models that uses presence-only data (Elith et al. 2011).

Machine learning models have been found to have a higher predictive performance than standard statistical models. They are especially suitable for complex ecological interactions between explanatory variables but are prone to overfitting (Chollet Ramampiandra et al. 2023). Methods like RF and SVM are non-parametric and can model non-linear relationships between predictor variables, but they have more errors because they learn noise. Maxent performs slightly better than other algorithms, especially with presence-only data(Grimmett et al. 2020).

1.2. Research Problem

VL is a fatal neglected tropical disease, and the number of infections in Kenya have increased in recent years (Mewara et al. 2022). As suggested by (Alvar et al. 2021) surveillance efforts in endemic areas can be improved through risk maps and knowledge of environmental risk associated with the vector and guide the development and placement of vector management tools. The central research problem is to identify these ecological and environmental conditions that influence the dispersal and distribution of the vector, as well as the transmission dynamics, and in turn, effectively predict the risk. Findings from this research can contribute to elimination strategies in place.

The transmission of VL is a complex socio-ecological system, and identifying risk factors presents a wicked problem. Wicked problems are complex issues with uncertain knowledge and low stakeholder consensus (Balint et al. 2011). This problem can further be described as an analytically complex problem whereby the problem is clear, but the solutions are not obvious (Alford et al. 2017). Risk maps are often developed through modelling approaches integrating climatic, environmental, and socioeconomic variables. Factors influencing sandfly distribution vary from region to region; it is unknown which variables are relevant for West Pokot County.

The anthroponotic transmission of VL is affected by the proximity to sandfly breeding habitats, population densities of the vector, abundance of plant sugar sources for the sandflies in the surroundings and presence of ample blood sources for sandfly females including host individuals (Kirstein et al. 2018). Understanding these dynamics requires detailed characterisation of the environment, particularly at a fine spatial resolution. This study aimed to model and predict VL in West Pokot at a fine spatial scale. With the absence of systematic field data collection of sandfly samples, the choice of modelling approach was confined to methods that use presence only VL occurrence as the input data.

1.2.1. Objectives

This study aims to identify the environmental risk factors associated with Visceral Leishmaniasis (VL) transmission in West Pokot and develop a model to predict risk areas for the spread of VL. To achieve this goal, the specific objectives are defined below.

1. To map vector microhabitats from satellite imagery.
2. To determine the key environmental variables and their contribution to the prediction of disease occurrence of VL.
3. To model the identified relationships to predict the geographic distribution risk of VL in West Pokot.

1.2.2. Research Questions

The corresponding research questions to the objectives above are.

1. Where are the vector habitats, and how can they be identified from high-resolution satellite imagery?

2. What are the relevant environmental variables, what is their contribution to the occurrence of VL infections and can they be used to develop an accurate and reliable model for forecasting the risk?

3. What is the predictive accuracy of a machine learning model in forecasting the risk of VL based on the combination of vector ecology and environmental factors in West Pokot, and how can this model be applied to identify risk zones?

2. Case Study Area and Datasets

2.1. Case Study Area

West Pokot is situated in the Northwest of Kenya along the international border of Kenya and Uganda and lies between 34° 47'and 35° 49' E and Latitude 1° and 2° N. The county is approximately 9169 km^2 and has a population of 621,241 people as per the last census of 2019. It has a rural population, with the urban population making up only 5% of the total. Administratively, the county has 4 sub-counties, 20 wards, 16 divisions, 65 locations, and 224 sublocations. The main economic activities are agriculture and livestock, and the main community is the Pokot people (West Pokot n.d.).

The region has diverse topographic features. The North and Northeastern regions are characterised by dry plains and a low altitude of below 900m. At the same time, on the Southeastern side lies the Cherangany hills with an altitude of almost 3370 meters. The high-altitude areas are highly suitable for agriculture, while the medium-altitude regions, ranging from 1,500 to 2,100 meters, receive limited rainfall and are mainly used for pastoral activities. The low-lying areas are in Alale, Kacheliba, and Kongelai divisions (West Pokot n.d.) which are mostly arid areas with low agricultural potential and are pastoral zones (Obwocha et al. 2022).

The chosen study site is in the Kacheliba division within the Pokot North sub-county of West Pokot. We selected the study area because of the prevalence of VL in the area and on the availability of termite mound data. The data description for the termite mound data is in the section 2.3.1. and these locations and the case study area are shown in Figure 1.

The Kacheliba area is characterised by its low elevation and receives an average annual rainfall of about 400 mm. The region typically experiences average temperatures of 28°C. It has a savannah climate and two rainy seasons between March to June and in October and November. The average monthly weather data are shown in Figure 2.

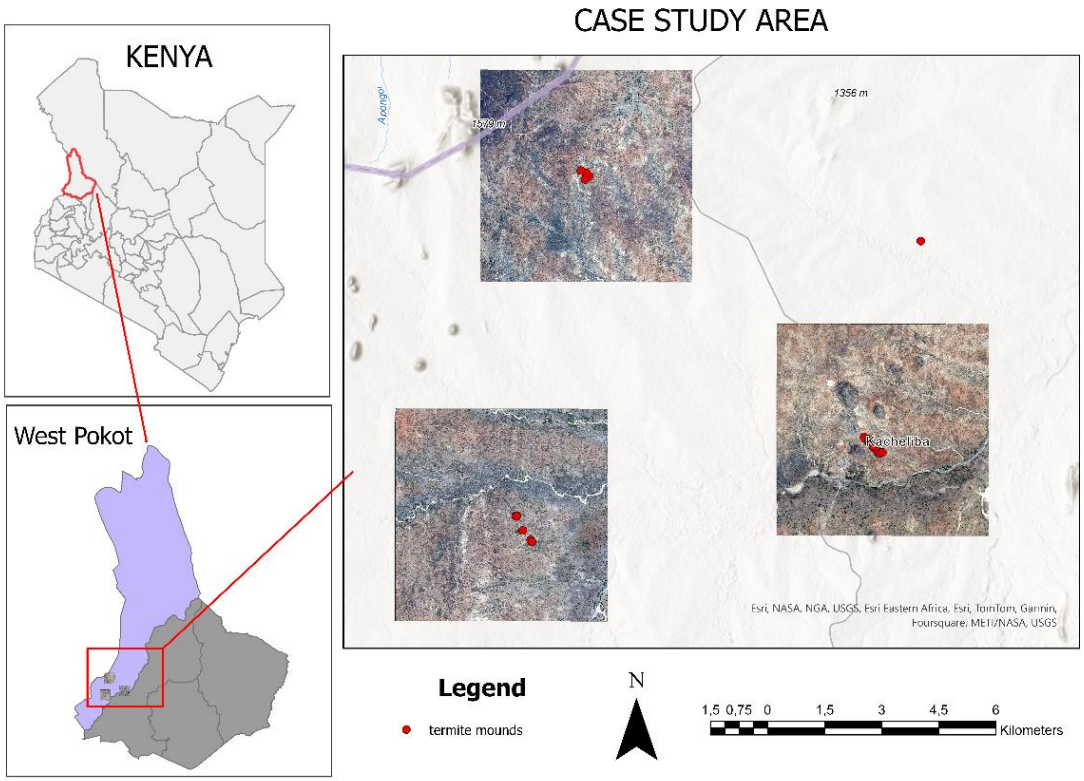

Figure 1: Case study area showing the position of West Pokot in Kenya and highlighting the location of Pokot North sub-county (in purple) where the Kacheliba division is located and termite mound locations.

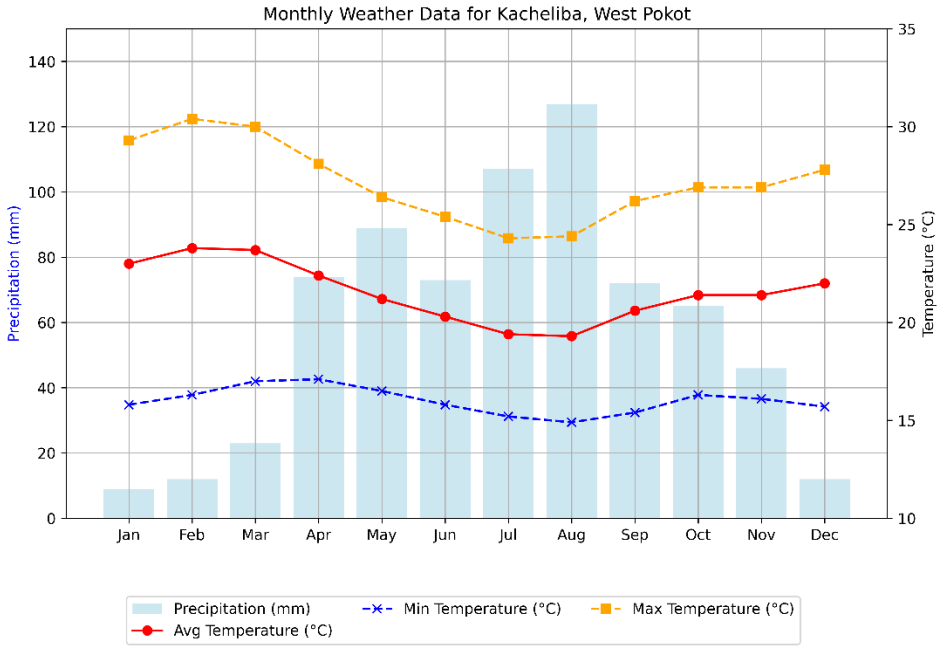

Figure 2: Weather and climate conditions in Kacheliba from (Climatedata.org n.d.)

2.2. VL in West Pokot

West Pokot County continues to have the highest number of leishmaniasis cases in the country, with over six hundred patients affected annually and the burden is still heavy (Shanzu 2023). VL infection cases were obtained from Kacheliba Hospital, located in the study area, which is a regional centre for treating visceral leishmaniasis (Van Dijk et al. 2023). A total of 1949 cases were documented at the hospital from 2018 to 2022.

In this period, infections have been on the rise and the year 2022 had the highest number of occurrences. Peaks of infection mostly in March, June, and October/November (Figure 3). Sandfly populations are usually at their peak during the wet seasons (March, August, and October to December) and these increase human–vector interactions (Koskei et al. 2024).

The median age for VL was 10 years and men contributed to 69% of the infections. The burden of VL is high in these two subgroups of the population. The incidences of 2019 and 2022 were termed as outbreaks (Mewara et al. 2022), and the persistent increase shows that control is still a challenge.

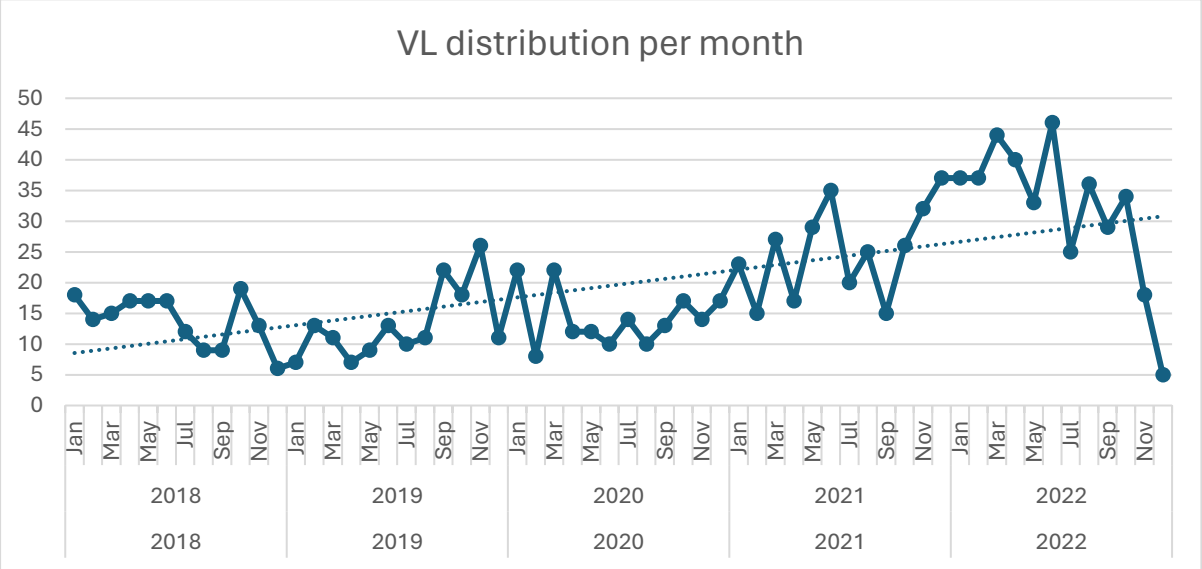

Figure 3: Temporal trends In VL in West Pokot showing an increase in the occurrence over the years with peaks from March to June, October, and November

We integrated the occurrence data with a sublocation boundary shapefile to assess its spatial distribution. The spatial patterns of VL in the period between 2018 to 2022 are highlighted in Figure 4. Most of the disease cases are concentrated in the eastern part of the county, although infections are widespread and recur in most sublocations. Comparing the maps from 2020 to 2022, it is evident that the disease is gradually expanding from the east to other regions during this period. In our study area, though the cases are not as high as in the East, there has always been constant disease cases presence.

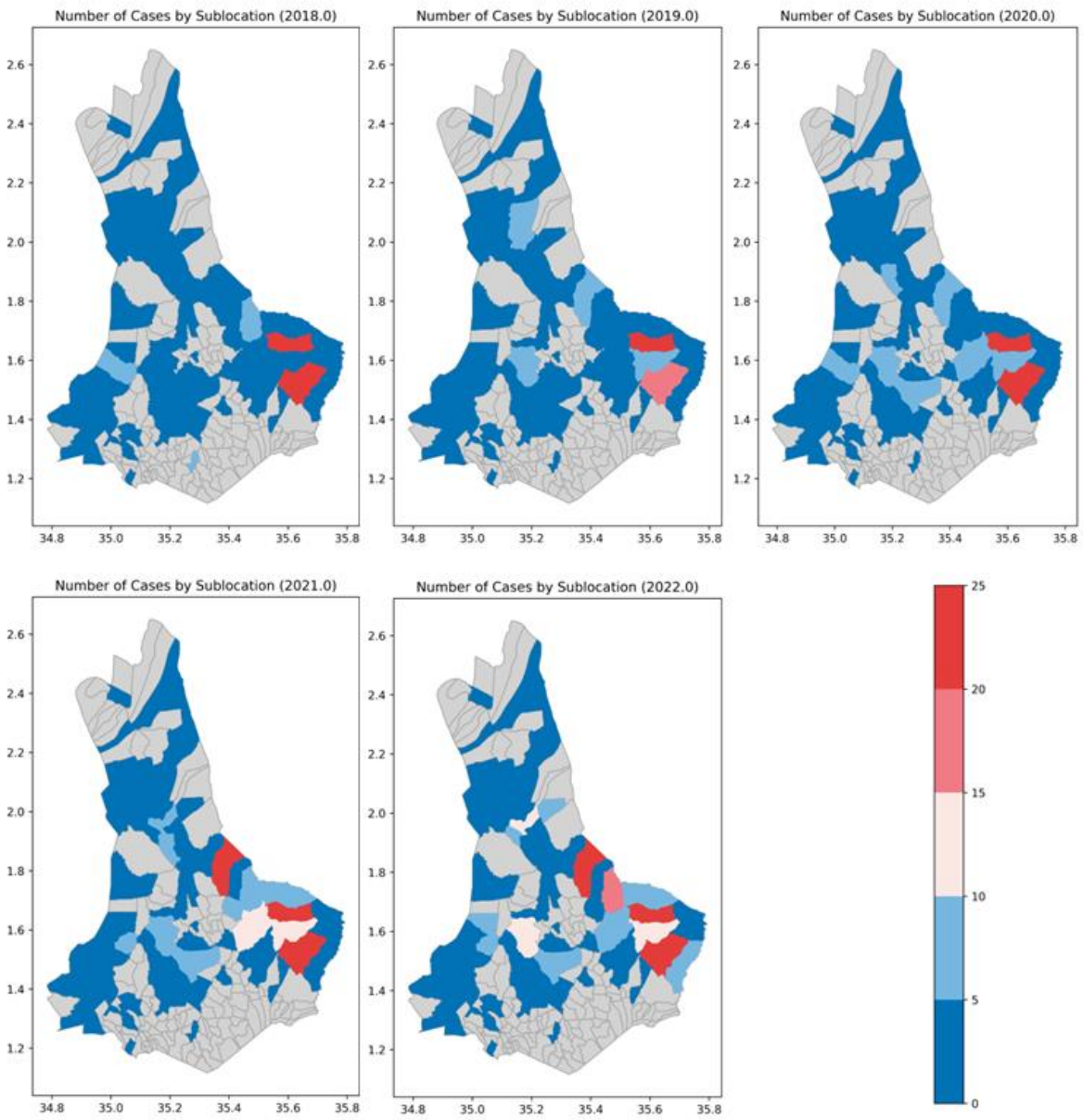

Figure 4: Annual cases of VL in West Pokot from 2018 to 2022

2.3. Datasets

2.3.1. Termite Mound Field Data

The primary goal of the first research question was to develop a method for mapping vector habitats, including termite mounds, using high-resolution imagery. To achieve this, we utilised GPS location data of termite mounds collected in the Kacheliba Division. During fieldwork conducted with (Van Dijk et al. 2023) for his research, 30 termite mound locations were randomly sampled, and their coordinates were recorded using GPS. The locations of these mounds are shown in Figure 1.

2.3.2. Disease Occurrence.

Patient information for visceral Leishmaniasis (VL) was obtained from records from Kacheliba Sub-County Hospital, West Pokot County, Kenya, between 2018 and 2021. The data contained the details of the patient's sex, age, month of infection and location of residence. The data was cleaned and geocoded. In the hospital records, the finest resolution was at the village level. Villages are administrative regions that comprise a cluster of houses and compounds (manyattas) (Mueller et al. 2014).

The initial method involved assigning villages to GPS coordinates collected in West Pokot by (Mueller et al. 2014) in 2007. Additionally, infection cases were geocoded to enumeration boundaries used by the KNBS in the 2019 census. These boundaries are typically designed to encompass settlements, and, in nomadic areas, they are based on walking distances, and they are like village boundaries.

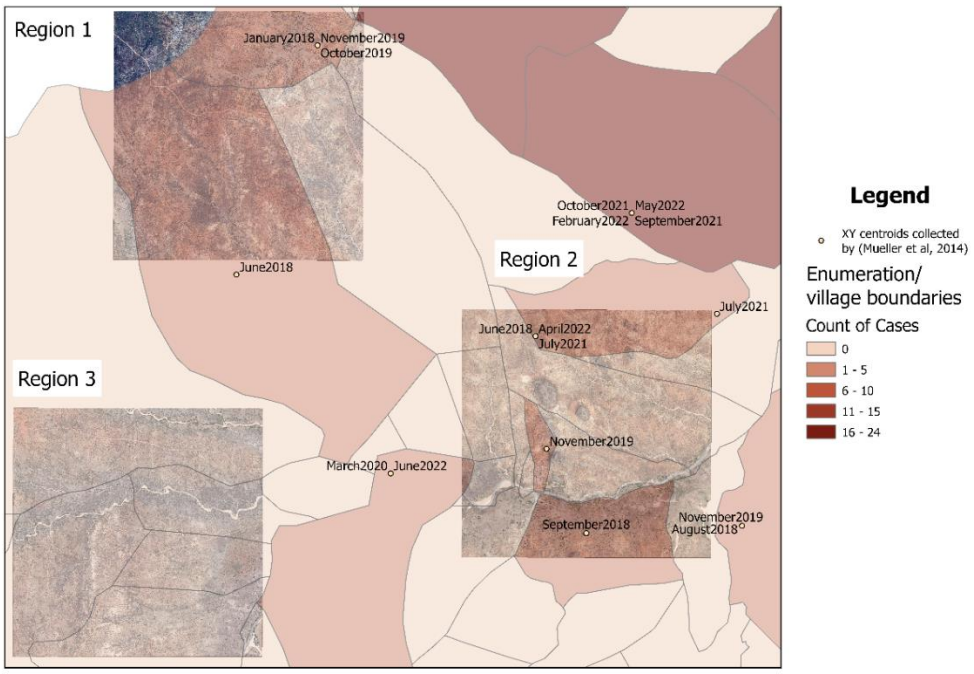

Figure 5: Location of geocoded disease cases labelled to their date of occurrence.

All cases occurring in a village are placed in one geographic coordinate, minimising the sample size for the modelling and only five of these locations fall in the study area. In presence-absence distribution models, training data with fewer presence samples can reduce predictive accuracy (Collart and Guisan 2023; Jiménez-Valverde 2020). It was necessary to take extra steps to overcome this pitfall.

2.3.3. Buildings Data

Buildings often function as indicators of human settlement, representing a more detailed resolution than villages. In this study, building footprint polygons were obtained from the Google Earth Engine catalogue (Sirko et al. 2021). The dataset is a comprehensive collection comprising 1.8 billion building detections, encompassing Sub-Saharan Africa. Each polygon is accompanied by a confidence score indicating the reliability of the detection and most of the buildings in this dataset were accurate for the case study area.

2.3.4. Remote sensing data

Very high-resolution imagery

The study utilised WorldView-2 satellite images acquired on 6 March 2023, covering the southern part of Kacheliba in West Pokot County. Worldview-2 images have 8 bands, are pan-sharpened to a spatial resolution of 0.5m and were employed to extract vector microhabitats in the region. Additionally, high-resolution satellite imagery and WorldView-2 satellite images with 3m resolution were used to complement the analysis. The data coverage is shown in Figure 1.

Environmental Datasets

Environmental factors identified in the literature were temperature, NDVI, precipitation, land use and land cover, soil, elevation, and humidity (Aklilu et al. 2023; Jiang et al. 2021; Sardar et al. 2020). The datasets, derived variables, and resolution information are summarised in Table 1.

Table 1: Environmental Datasets used and their specifications.

Data source	Variable	Spatial res (m)	Data values	Temporal properties
ALOS PALSAR DEM	Slope and altitude	12.5	continuous	static
Sentinel 2 Imagery	Maximum median and minimum NDVI	10	continuous	2018-2022
Landsat- OLI	LST	30	Continuous	2018-2022
Worldview -2 imagery	Land cover	0.5	Categorical	static
ISDA soil grids	Clay content	30	continuous	static
Bioclim	precipitation	927	continuous	static

3. Methodology

This chapter provides an overview of the methods used to achieve the objectives. All processes used in this study are discussed further in subsequent sections and are summarised in Figure 6 below.

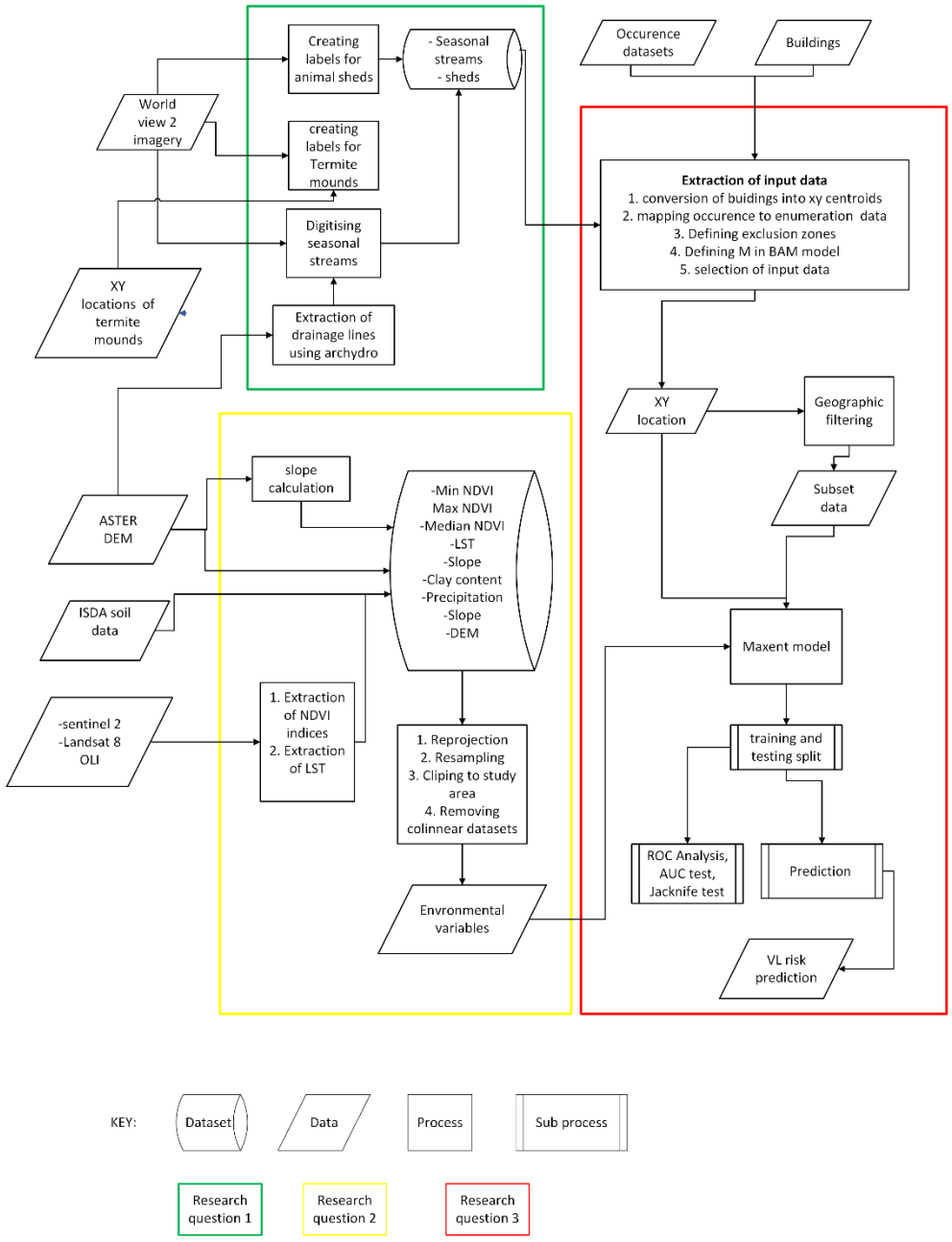

Figure 6: Workflow adapted in the research.

3.1. Vector Habitat Extraction

In previous studies, *P. martini* was observed to occur in abundance in termite mounds and animal sheds. The chances of encountering the sandfly vector were five times higher in termite mounds and animal sheds compared to indoor spaces. Additionally, the sandfly prefers to feed on acacia trees more than other tree species (Hassaballa, Sole, et al. 2021). The sandfly vector favours more humid habitats (Mueller et al. 2014), and riverbank fissures and crevices of seasonal rivers create ideal humid zones preferred by the vector for hiding and breeding (Abdullahi et al. 2022).

A significant ecological adaptation of sand flies is their selective use of habitats (Hassaballa, Sole, et al. 2021), and sandflies in general are weak at flying with reports indicating that adults typically fly one hundred meters or less from their larval habitats (Cecílio et al. 2022). Without systematically collected field data for sandflies, we hypothesised that animal sheds, termite mounds and seasonal river crevices are the vector's niche and can be used as indicators of sandfly presence.

3.1.1. Animal Sheds

Animal Enclosures (sheds) in the savannahs of Kenya can be visually detected on high-resolution imagery because of spectral contrast with surrounding land and a visible fence around the animal enclosure (Vrieling et al. 2022). They have a distinctive colour of manure and a continuous fence around them, and they vary in size, usually between 15-25m in diameter (Tyrrell et al. 2022) as shown in Figure 7 below.

Figure 7: Examples of animal enclosures visible on high-resolution imagery; A is adapted from (Tyrrell et al. 2021) *showing animal enclosures in the south of Kenya, B and C are in the study area as seen from Google Earth Engine and Worldview-2 imagery respectively.*

The availability of sub-meter, very high-resolution satellite imagery enables the delineation and extraction of boundaries. The methods of extracting boundaries can be categorised into edge-based or region-based. Edge detection methods such as Sobel, canny or Scharr extract boundaries of all isolated objects and the boundaries are usually not closed (Cheng et al. 2020). Region-based methods group homogenous pixels, and sometimes, locating linear and visible edges in the segmented image is impossible (Cao et al. 2023).

Recently, deep learning algorithms have been used for automatic edge and contour detection, and they have been instrumental in learning advanced data representation for feature extraction, classification, and segmentation (Cao et al. 2023). Mask R-CNN, a state-of-the-art CNN, for instance, has been used together with very high-resolution imagery to detect ships, automatically detect artic ice-wedge polygons, and extract farm boundaries (Cheng et al. 2020; Feng et al. 2019; Zhang et al. 2018). So far, no study has used it to extract animal enclosures in the savannah and this research trained and extracted polygon boundaries for animal sheds.

3.1.1.1. Feature extraction

To develop an object detection algorithm, the first step was the generation of labelled data. Shed polygons were manually digitised from the pan-sharpened image labelled 'B' in Figure 8, based on size, visual characteristics, and the presence of a clear boundary. A total of 151 polygons were digitised and using the ArcGIS Pro tool ‘create labels for deep learning model’ (ESRI n.d.), they were converted into labels for the deep learning model. This resulted in 569 labels, which served as the input to train a Mask R-CNN model.

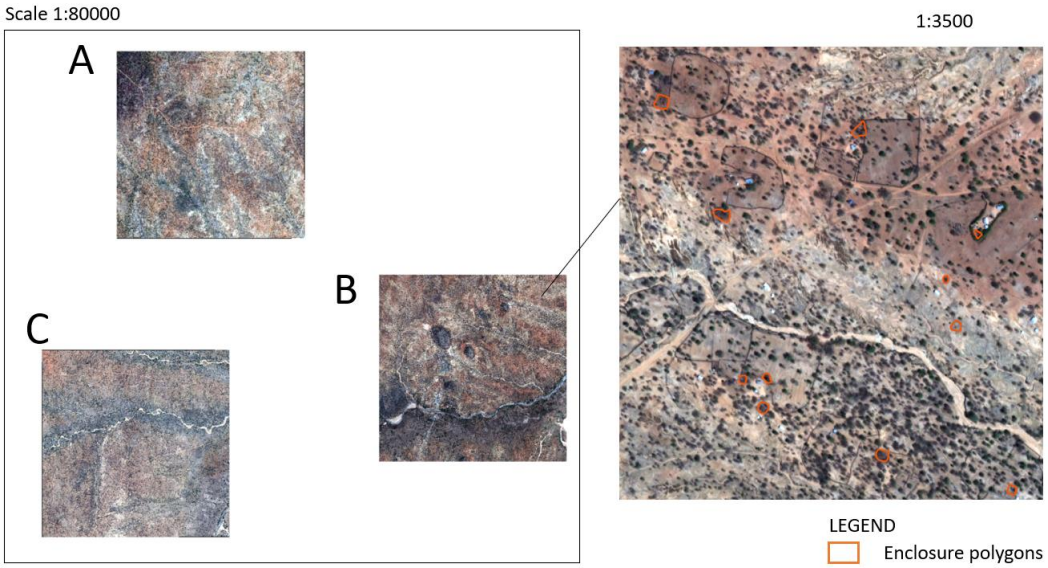

Figure 8: Location of train and test areas. Image B was used to generate labels (from digitising polygons) to train the model and it was tested on images A and C

Afterwards, we trained a model with the labels generated from the previous step. Mask R-CNN models are developed on top of a faster R-CNN. The architecture of a faster R-CNN works in two stages. The first stage consists of two networks a backbone and a regional proposal network. The selected backbone was Resnet-50 which is a convolutional neural network that is 50 layers deep. Once these two networks run, a set of proposals are formed. These proposals are regions in the feature map that contain the object. In the second stage, the network predicts bounding boxes and object classes for the regions in stage one. The size of these regions is fixed using RoI pooling or RoIAlign method. This is illustrated in Figure 9.

We trained an instance detection Mask R-CNN model with a Resnet-50 backbone. We set up the training procedure to run the model twenty times (20 epochs) and to stop training when the model was no longer improving, to save time and processing power.

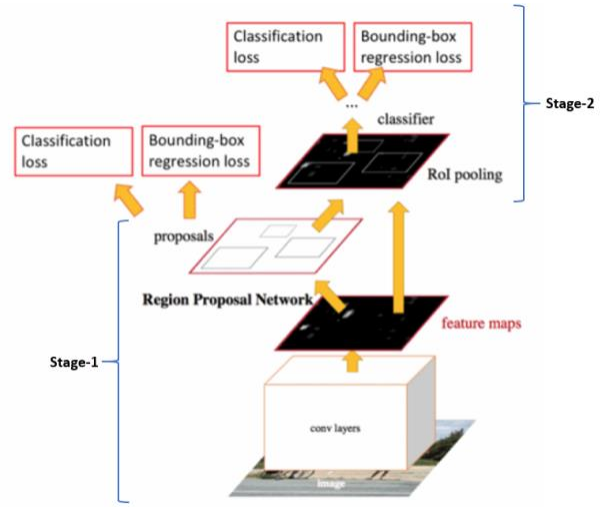

Figure 9: Faster RCNN architecture showing the 2 stages (ESRI n.d.).

3.1.2. Termite Mounds

The same extraction method using a deep learning model and satellite imagery was to be applied to the extraction of termite mounds. XY locations were collected in the field, and they were to serve as training data to delineate labels. Termite mounds, characterised by their crevices and occasional large openings, typically rise from 1 to 5 meters in height. To identify artillery craters in Ukraine (Duncan et al. 2023), used very high-resolution Worldview-2 imagery data and a U-Net model. We hypothesised that termite mounds, like artillery craters, would be recognisable on high with unique visual characteristics, such as lighting, shadows, and angles. However, each termite mound appeared different, and generating training labels for the locations was impossible.

3.1.3. Seasonal Riverbanks

The study area has a prominent main river alongside several seasonal tributaries, though these tributaries are challenging to identify since the Worldview-2 satellite images are for the dry season. An alternative for this is drainage feature extraction from a digital elevation model. To identify and extract river networks in the study area, the ALOS PALSAR digital elevation model (DEM) with 12.5 m resolution, was downloaded from the Alaska Satellite Facility (ASF) Distributed Active Archive Center (DAAC). Stream delineation was conducted using the Archydro tool on ArcGIS Pro which uses a DEM for watershed delineation and stream network generation by calculating flow direction and accumulation (ESRI 2013).

From a visual inspection to assess the correctness of the output, the drainage networks extracted did not align accurately when overlaid with very high-resolution satellite imagery as seen in Figure 10. With the guide extracted drainage network from the DEM, a new drainage network was digitised to eliminate the positional errors that arose from the extracted network.

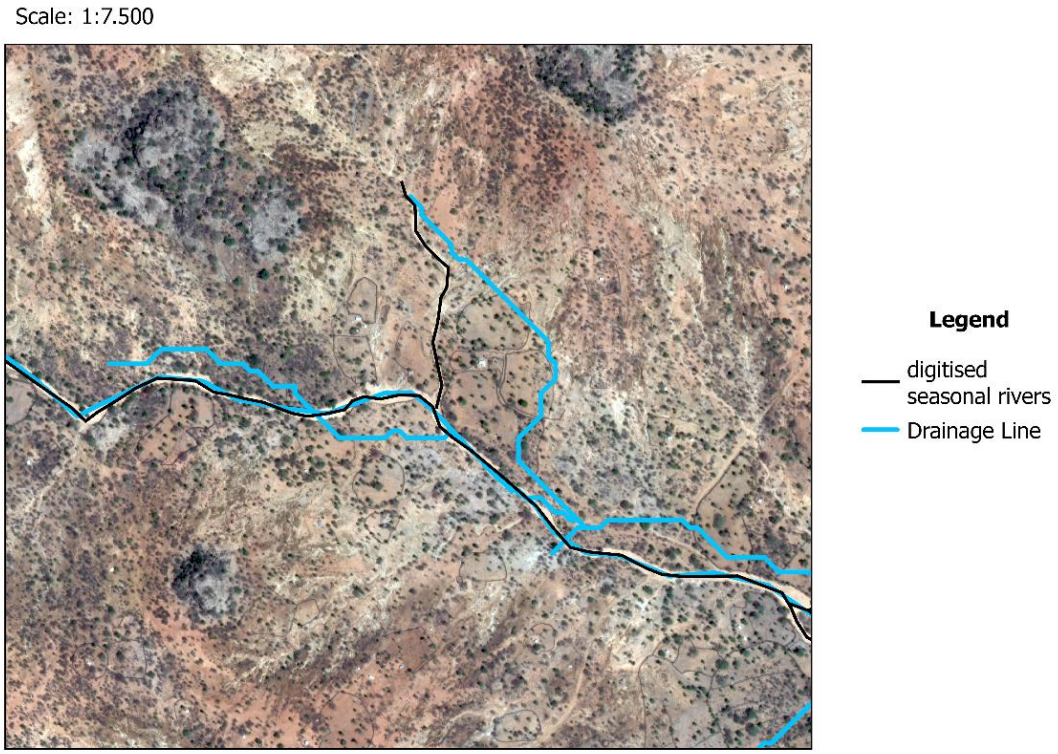

Figure 10: Comparison of the output of the methods used to extract seasonal rivers. The drainage lines were obtained using stream delineation with Archydro tools and they served as a guide for digitising seasonal rivers.

3.2. Environmental Variables Derivation

3.2.1. NDVI

Normalized Difference Vegetation Index (NDVI), an index derived from the red and near-infrared spectral bands of satellite imagery and measures green vegetation abundance. We utilized the European Space Agency's (ESA) Sentinel-2 Level 2A surface reflectance data to calculate the maximum, mean, and minimum NDVI indices for the period between 2018 and 2022. The satellite image has a high resolution of 10m. These indices were generated using built-in tools on the Google Earth Engine (GEE), a cloud-based platform for geospatial analysis.

3.2.2. Temperature

In previous ecological studies, Land Surface Temperature (LST) has been used as a proxy for air temperature (Chabot-Couture et al, 2014). Land Surface Temperature (LST) is the temperature estimated using satellite imagery, accounting for the temperature at the top of the canopy. The highest spatial resolution for freely available satellite imagery with a thermal band is provided by the Landsat 8 Operational Land Imager (OLI) instrument, which offers a resolution of 30 meters.

LST was estimated using Landsat-8 OLI on the GEE using the procedure from (Jimenez-Munoz et al. 2009). This process involves calculating LST from the brightness temperature derived from the thermal band and a fractional green cover estimated from NDVI. Median LST for the period between 2018-2022 was generated and resampled to 10 m spatial resolution for further analysis.

3.2.3. Land Cover

To prepare the classification, training datasets containing six classes present in the study area was created. The classes that represent the land cover features in the study area were, grasslands, bare land, water, forests, herbaceous and built-up and they were extracted from Worldview-2 images.

We used the Random Forest (RF) algorithm for classification. RF is made up of many decision trees. Each tree is built using a random subset of training features. When classifying a data point, each tree votes for the most likely class. We found that using around 400 trees gives stable results, but we opted for 500 trees for caution. Additionally, we did not decide the number of features used for predictions, ensuring a minimum branching depth of 2 in the trees.

3.2.4. Topographic

The digital elevation model used for the study area was the ALOS PALSAR digital elevation model DEM with 12.5 m resolution images. The DEM was used to calculate the slope on ArcGIS Pro. Additionally, the elevation data from the DEM was used in modelling. Both datasets were resampled to a 10m pixel size for the modelling process.

3.2.5. Clay Content

We obtained a dataset for soil properties from ISDA soil data, specific to Africa, accessible through the GEE catalogue. The product contains information on soil properties and one of them is clay content (Miller et al. 2021). The dataset has a resolution of 30 meters and includes four layers: 0-20 cm and 20-50 cm depths, each with predicted mean and standard deviation values. We extracted the predicted mean clay content for the 0-20 cm layer and resampled it to 10m.

Additional data for annual precipitation weather data was obtained from the Worldclim database (https://www.worldclim.org/).
The variables were then all projected to the same coordinate system (UTM zone 36N), clipped to the extent of the regions, resampled and stored as .asc files for modelling.

3.3. Multicollinearity

Before fitting species distribution models, multicollinearity between the environmental variables is assessed. Multicollinearity arises when the explanatory variables are highly correlated, leading to misleading results and hindering analysis. With environmental variables, multicollinearity can lead to underestimation or overestimation of the effects of variables, causing misleading information (Kim 2019). The environmental variables were assessed using the variance inflation factor (VIF). High VIF values indicate multicollinearity, and following the rule of 10, variables with a score above 10 are dropped (O'Brien 2007).

We also conducted a Pearson correlation analysis to compare the remaining variables. This statistical method allowed us to measure the strength and direction of the linear relationship between two variables and detect multicollinearity among the variables. Variables with high correlation coefficients (0.8) were eliminated and not used for modelling.

3.4. Training data

As mentioned in the section 2.3.2, only five location occurrences were available for the three locations in the study area. Since these locations were the centroids of villages, there was a need to reformulate assigning the xy coordinates.

The distribution of disease is limited to the dispersal capacity or the "niche" of the parasite and the transmitting vector species (Escobar and Craft 2016). Niche can be further described using the Biotic Abiotic Movement (BAM) diagram in Figure 11. We first define the geographic and environmental space where the environmental space is the ENM, and geographic space is the real and georeferenced space (Sillero et al. 2021).

In the geographic space represented by G in Figure 11, A represents the abiotic factors that influence the growth of the species, B is the region where the species can coexist with its competitors and M represents sections in the environmental space the accessible areas for the parasite and their intersection J_o is the potential occupied area The species can disperse beyond J_o and go anywhere within the region of M so the niche can be defined as J_{ss} and $M \approx J_{SS}$. (Soberon et al. 2005).

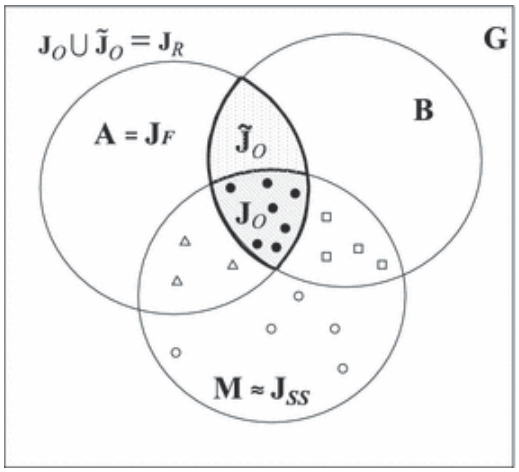

Figure 11: BAM model, G represents the total study area, A the region growth rate of species would be positive, B where the species can coexist with competitors and M where the species may be found.

For the sandfly *P. martini*, we defined J_{ss} in terms of its dispersal range. Since the sandfly vector can only move a few hundred meters from its preferred habitat, we defined J_{ss} using the ecological variables derived: the animal sheds, and the seasonal rivers together with the field data termite mounds. We created a buffer zone around these features for 200m and we defined this region as M.

The transmission of VL requires an infected human host. To represent human presence, we utilized the buildings dataset. We simulated potential XY coordinates for the modelling input by using the centroids of these buildings. Subsequently, we filtered out centroids located in enumeration areas with zero reported cases. Centroids that were in M formed the initial dataset. The study sites will be referred to as Region 1 and Region 2 as labelled in Figure 5. No locations were available for Region 3.

Defining M and selecting these locations leads to an uneven distribution of sampling points and might cause overfitting, biased significance, and inflated accuracy. Several bias correction methods exist such as adjusting the presence of data through geographic and environmental filtering or adjusting the background data through restrictions (Xu et al. 2024). We selected the geographic filtering using the spThin package in R. Figure 12 shows the distribution of the data for both the full and subset in both regions.

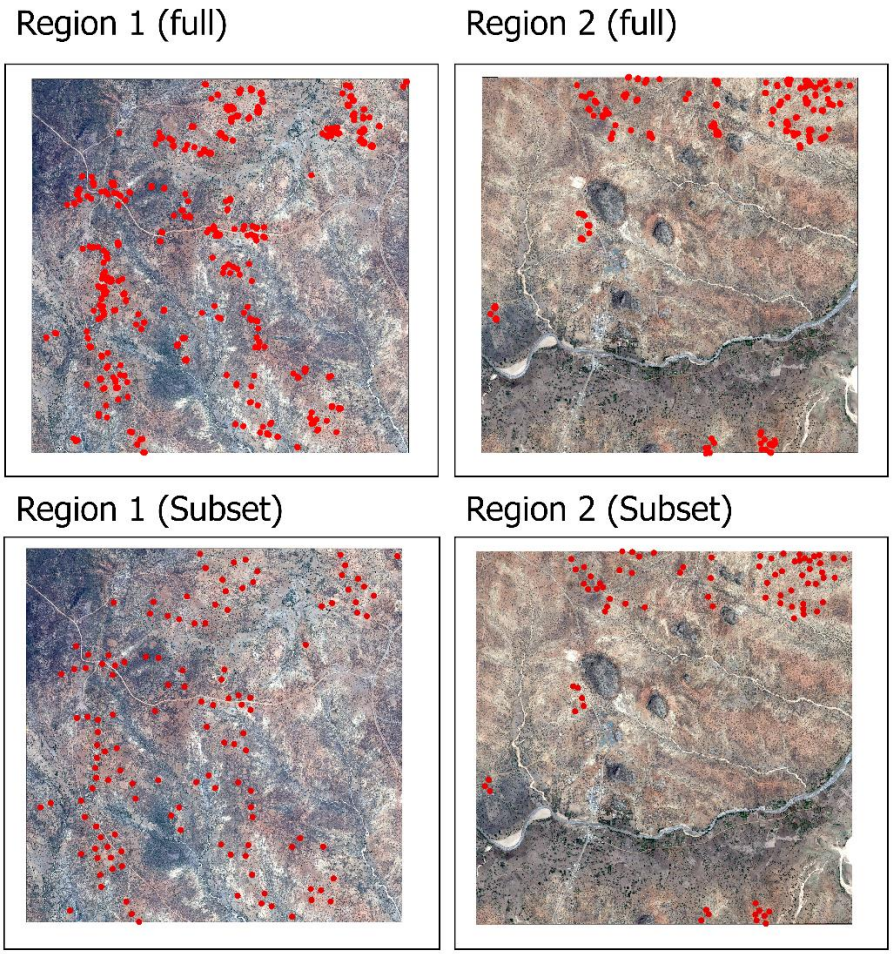

Figure 12: Simulated Occurrence data for VL in the two regions.

3.5. Model building.

3.5.1. Maximum entropy

Maximum entropy (maxent) is a machine-learning technique that originated from statistical methods. It can be used to estimate the probable distribution of a target by finding its maximum entropy which is the most spread out and closest to uniform distribution, subject to a set of conditions that represent the incomplete information of the distribution. This model requires presence-only data and environmental variables, can use continuous and categorical data, and incorporates the interaction between different variables (Phillips et al. 2006).

Model parameters.

The main parameters to be determined for model training are the feature classes (FC) and regularisation multiplier (RM). The feature class is the statistical transformation of the environmental variables on which the model constraints will be built. There are five feature types namely linear (L), quadratic (Q), product (P), threshold (T), hinge (H), and categorical features. The random multiplier (RM) is used to penalise for the inclusion of

additional parameters and prevent overfitting and underfitting. Low RM values will result in a model with many predictors and high RM values will lead to smoother more general models (Jiménez-Valverde 2020; Phillips et al. 2006).

To determine the feature class for the Maxent model, we used the ENMval package in R. It provides six ways to split the input training data which are variations of the K-fold cross-validation(Muscarella et al. 2014). The selected method for this study was the random K-fold. The evaluation method metrics are summarised in Figure 13.

Metric	Description	References
AUC_{TEST}	The threshold-independent metric AUC based on predicted values for the test localities (i.e. localities withheld during model training), averaged over *k* iterations. Higher values reflect a better ability for a model to discriminate between conditions at withheld (testing) occurrence localities and those of background localities (by ranking the former higher than the latter based on their predicted suitability values). The rank-based AUC does not indicate model fit	Hanley & McNeil (1982), Peterson *et al.* (2011)
AUC_{DIFF}	The difference between the AUC value based on training localities (i.e. AUC_{TRAIN}) and AUC_{TEST} ($AUC_{TRAIN} - AUC_{TEST}$). If $AUC_{TRAIN} < AUC_{TEST}$, the returned value is zero. Value of AUC_{DIFF} is expected to be positively associated with the degree of model overfitting	Warren & Seifert (2011)
OR_{MTP} ('Minimum Training Presence' omission rate)	A threshold-dependent metric that indicates the proportion of test localities with suitability values (MAXENT relative occurrence rates) lower than that associated with the lowest-ranking training locality. Omission rates greater than the expectation of zero typically indicate model overfitting	Fielding & Bell (1997), Peterson *et al.* (2011), Radosavljevic & Anderson (2014)
OR_{10} (10% training omission rate)	A threshold-dependent metric that indicates the proportion of test localities with suitability values (MAXENT relative occurrence rates) lower than that excluding the 10% of training localities with the lowest predicted suitability. Omission rates greater than the expectation of 10% typically indicate model overfitting	Fielding & Bell (1997), Peterson *et al.* (2011)
AICc	The Akaike Information Criterion corrected for small samples sizes reflects both model goodness-of-fit and complexity. The model with the lowest AICc value (i.e. $\Delta AICc = 0$) is considered the best model out of the current suite of models; all models with $\Delta AICc < 2$ are generally considered to have substantial support	Burnham & Anderson (2004), Warren & Seifert (2011)

Figure 13: Adapted from (Muscarella et al. 2014), evaluation metrics that will be useful to select the model parameter FC and RM

The FC and RM combination with the best AUC difference and AIC are selected to run the Maxent model. The variables used are shown in Table 2.

Table 2: Maxent model parameters

		RM	FC
Region 1	Full	1	H
	Subset	2	LQHP
Region 2	Full	1	H
	Subset	1	H

The Maxent Java application was used to run the model. The xy datasets and raster layers were loaded to the interface. Model parameters were adjusted. Other settings changed were the random seed to ensure consistent data splits, resampling techniques set to a ten-fold cross-validation and doing a jack-knife to estimate variable importance. Due to the size of the area, background points were sampled from a bias file created with ENMeval and reduced to 5000 points.

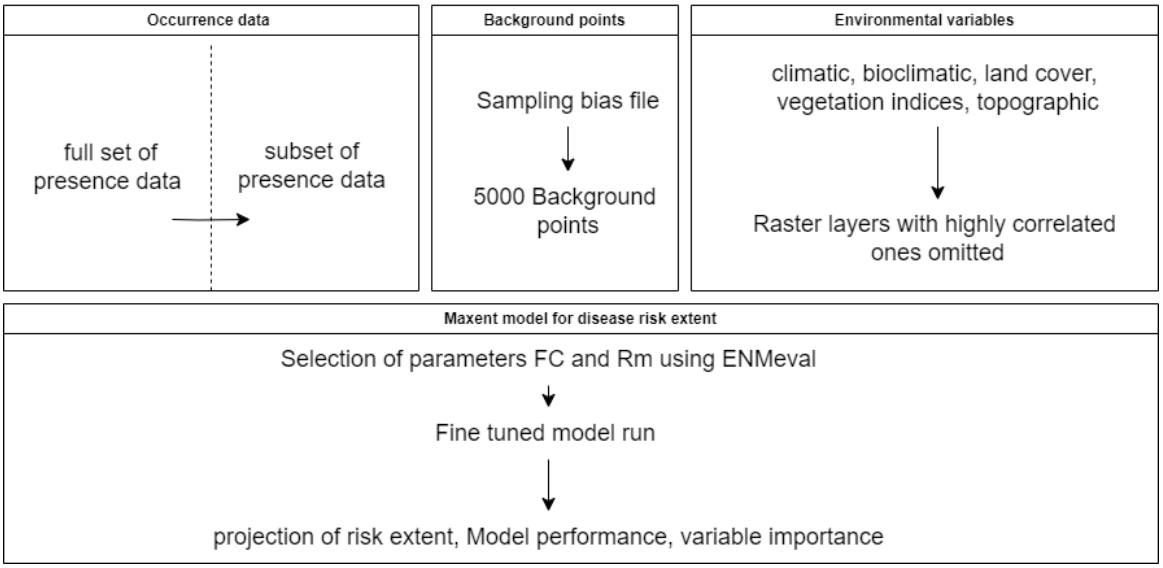

Figure 14: workflow for modelling with maxent.

4. Results

This chapter provides findings of the study based on the proposed methodology.

4.1. Vector habitat extraction.

We trained a deep learning model with the Detect object using a deep learning tool in Arcpro and used the generated model to detect animal enclosures. The average precision score was 0.705. The training and validation loss are shown in Figure 15.

From a visual inspection, the model successfully delineated animal enclosures in the study area and generated a polygon shapefile. There were no false positives, but larger enclosures were not captured by the model. A limitation of this process was the lack of ground truth validation data for animal enclosures.

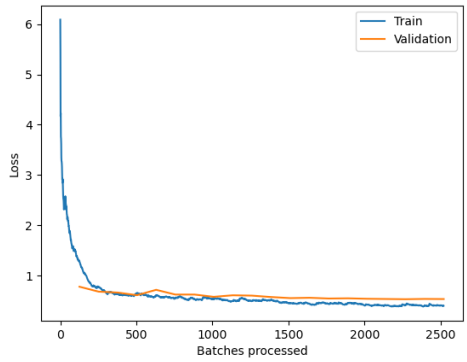

Figure 15: Training and validation loss. The lower the loss, the more reliable the model

4.2. Selection of Predictor Variables

We conducted a (VIF) test and retained the relevant environmental variables that had VIF values below 10. Only the median NDVI in Region 1 (highlighted red in Table 3) had a score of above 10 and it was not used in the modelling process.

Table 3: VIF scores

Environmental Data	**VIF**	
	Region 1	**Region 2**
Median NDVI	10.612029	9.279748
LST	3.516447	2.972034
Land cover	1.371782	1.680830

DEM	2.261939	2.451430
Soil	4.057682	2.724596
Slope	1.173661	1.057714
Min NDVI	2.670555	2.255488
Precipitation	2.044766	1.948943
Max NDVI	6.130484	5.337984

We also did a Pearson correlation, and the results are shown below in Figure 16 and Figure 17. There were no values above 0.8 for Region 1 and Region 2, Maximum and median NDVI were highly correlated and maximum NDVI was dropped from the modelling process.

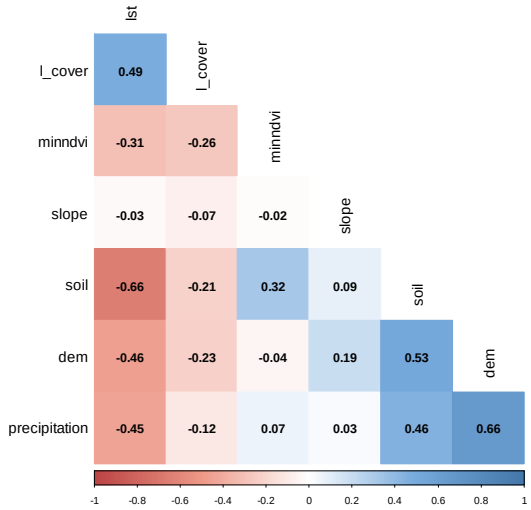

Figure 16: Pearson correlation matrix for region 1. All values are below 0.8.

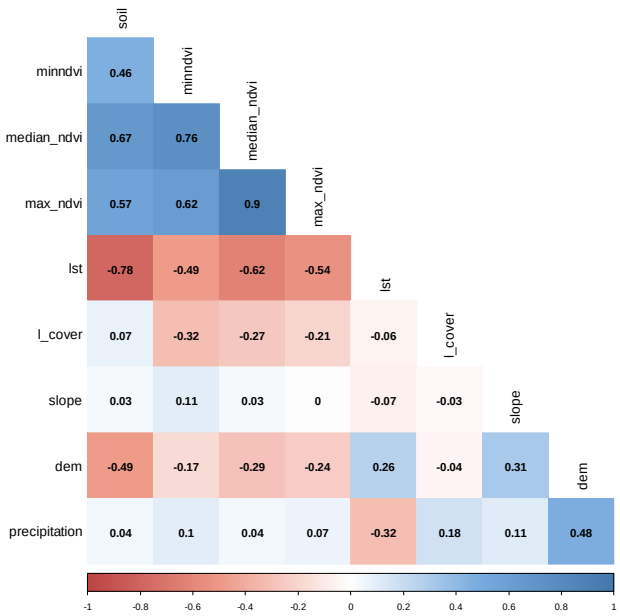

Figure 17: Pearson correlation matrix for region 2. Maximum and minimum NDVI are highly correlated and Maximum NDVI is dropped.

4.3. Model Performance

The predictive performance of the models was assessed using the area under the receiver operating characteristic curve (AUC) metric. The AUC is a valuable measure of a model's ability to distinguish between the presence and absence of a condition. Models with AUC scores between 0.75 and 0.9 are deemed highly effective. Our model demonstrated robust discriminatory power across both regions, except the subset in Region 1. A decrease in sample size corresponded to a reduction in AUC scores. The complex model performed the worst in this test.

Table 4: Model performance

		AUC
Region 1	Full	0.769
	Subset	0.662
Region 2	Full	0.805
	Subset	0.788

4.4. Predicted risk areas.

The model was reclassified into equal intervals with classes to see the general trend in the entire study area. The probability of risk ranges from ranges from 0-1 and the higher the value, the higher the risk.

As seen in Figure 18, with a reduction in sample size, larger areas were predicted as having more risk. In Region 1, our predictions indicated that risk was widespread in the full dataset but predominantly concentrated in the eastern part of the subset prediction. The high-risk areas were primarily located in the central region, characterised by rural communities living in 'manyattas' and numerous animal sheds.

In Region 2, the distribution of high-risk areas is similar for both datasets. The predicted risk areas for Region 2 were mainly in the Northern Region. This region is low-lying, with low NDVI, high temperatures and low rainfall. The south of this region is mainly agricultural, and it was predicted to have little to no risk. While there were no disease occurrences in the centre right of the region, this location was predicted to be high risk, meaning it may have conditions favouring vector survival.

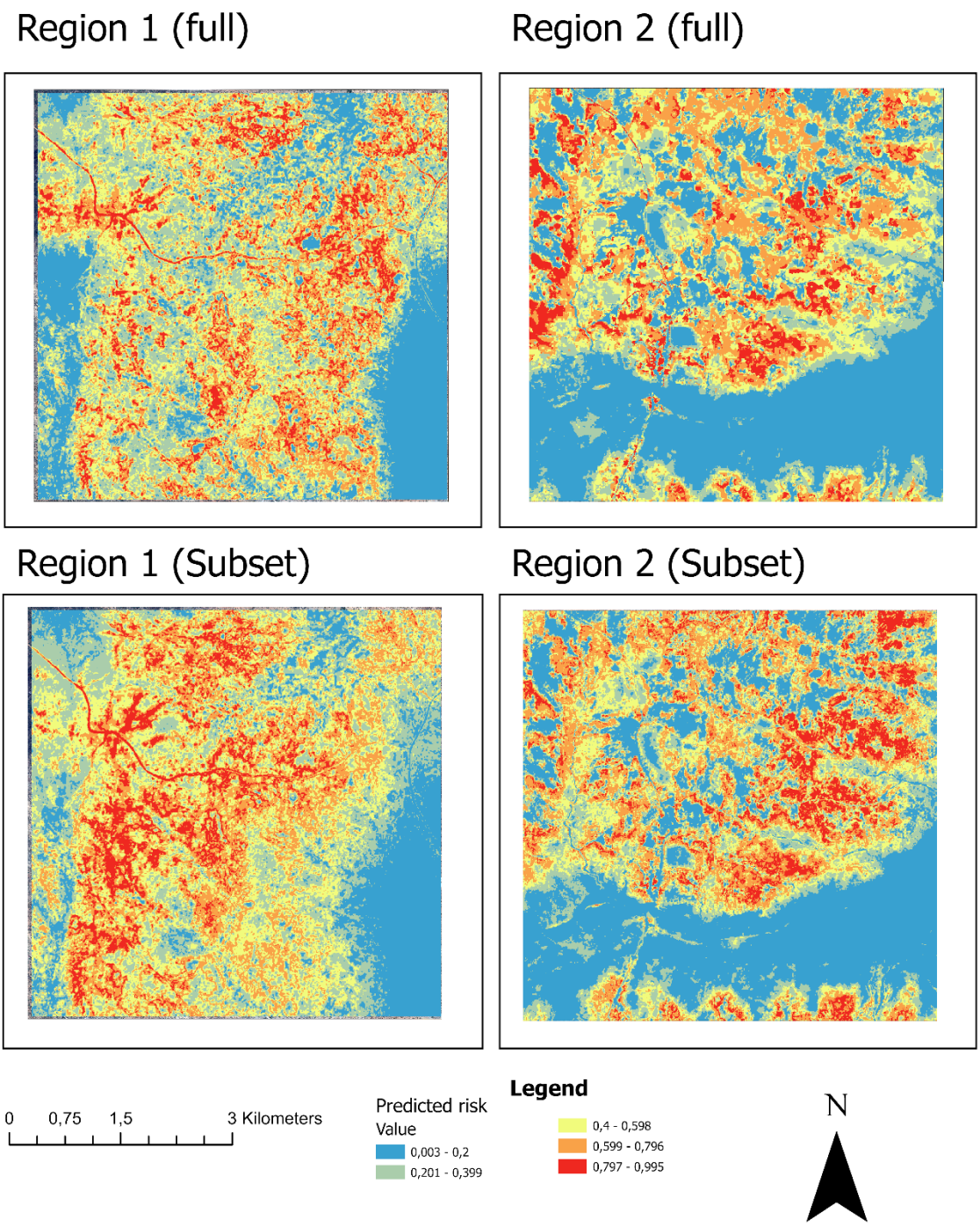

Figure 18: Predicted risk maps.

4.5. Variable importance

The Jackknife test gave information about predictor variables' contribution and relative importance to generate a MaxEnt model. The summaries are in Table 5. All the variables were different for each of the model runs. For all models, the percentage contribution on NDVI was quite high and it tested well for each subset. This is a proxy for vegetation, and it may indicate the relevance of the presence of acacia vector occurence.

Table 5: Variable contribution

		Variable	Percentage contribution
Region 1	Full	Maximum NDVI	53.6
	Subset	Precipitation	50.5
Region 2	Full	Median NDVI	37
	Subset	Median NDVI	58.2

Risk Map

The goal was to be able to predict risk at the house level. The pixels of the output map are 10m which covers a building. In Figure 19, in a section of Region 2, the pixel values are extracted for locations with buildings and in the top left corner, it is possible to see how risk values change with proximity to an animal shed.

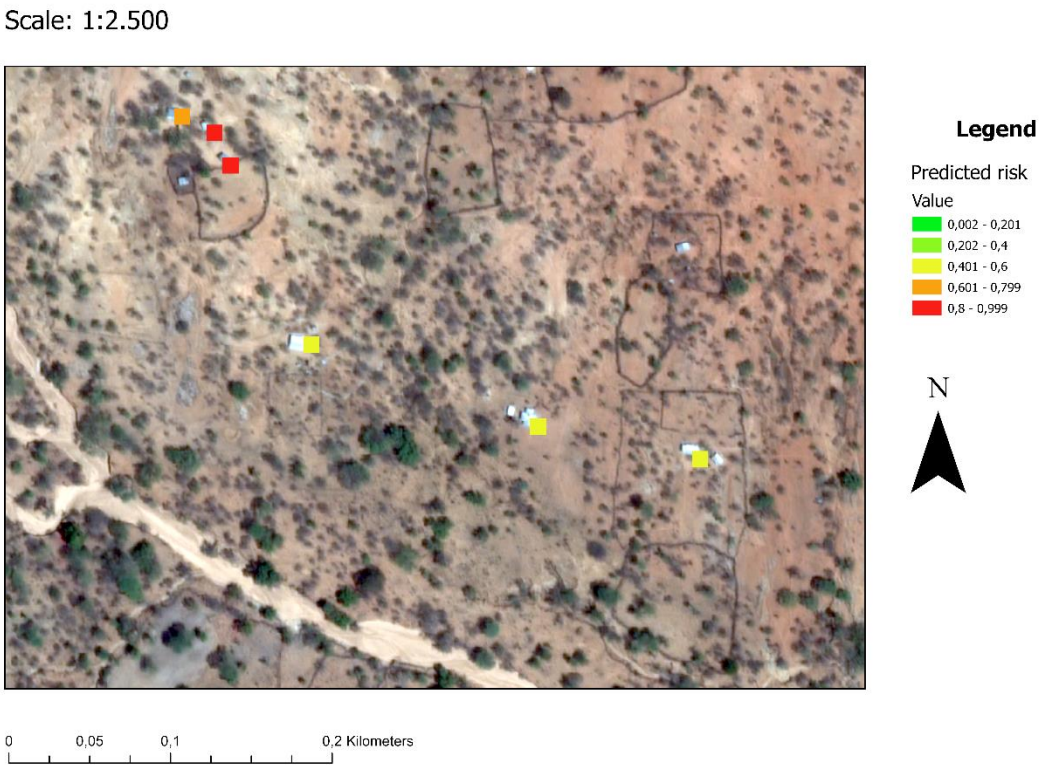

Figure 19: Risk map.

5. Discussion

This section aims to evaluate the results obtained from the modelling and the resulting predictions.

5.1. Uncertainty

This study had the challenge of having few input datasets, which informed the model choice and necessary steps to redistribute the model input. To map disease risk, occurrence data is usually in the form of disease cases, and in this case, only five locations were mapped, and this was insufficient.

When modelling cutaneous leishmaniasis, (Chavy et al. 2019) used the BAM framework for random redistribution to assign infection locations to regions where they are more likely to occur. The cutaneous leishmaniasis cases in his study region were georeferenced to cities and the disease is sylvatic, meaning it occurs near wild areas. He found that his results were robust and reliable in comparison to other studies. The modelling in this study had reliable results too, however, assessing their reliability was challenging due to the absence of existing maps or validation data for comparison.

The validity of the model predictions will be affected by the positional accuracy of the georeferenced locations. There may be variations in the environmental conditions of the simulated locations and the true location of the disease occurrence. The effect of this difference depends on the standard deviation of the explanatory variables in the spatial unit, and they can be minimised by using central tendency (mean, mode, median) values (Cheng et al. 2021). True presence is unknown, and it is impossible to quantify the distance away from true presence. This brings about uncertainty in the risk prediction we made.

The method chosen for this study for point redistribution was defining the full range of the species dispersal. In defining the BAM boundaries, only two variables were used. The inclusion of other ecological variables such as termite mounds, or acacia trees could have greatly affected the dispersal range generated.

With other simulated species, researchers often rely on bioclimatic variables or the species' real-world distribution (Xu et al. 2024). However, there is limited research on the suitable climatic conditions of sandflies in East Africa and we were only able to obtain 3 records from sandfly collection conducted by (Koskei et al. 2024) in 2016 which was insufficient for analysis. Consequently, the BAM approach, despite its uncertainty, was the only viable alternative for our study.

This approach additionally introduced movement into the risk model. The novelty of this study was the incorporation of movement at the local scale specific to this area. West Pokot is 95% rural and the choices for movement variables had to match the dynamics of the area. The predictions were more fine-tuned to the region.

5.2. Sample size

It is recommended that the input dataset of an SDM should contain at least 20 records per species, preferably more than 50, to ensure a robust model or accurately represent species with low occurrences, including those that are endemic and have naturally limited ranges. (Benavides Rios et al. 2024). In his study (Lamboley and Fourcade 2024) mentioned that model performance is worse in large, biased data than in smaller biased data. The simulated data for Region 1 had 600 locations and for Region 2, 229. We opted for filtering to compare the results of the full range to subset data.

Filtering the locations reduced the AUC score of the model. This was also the case with the study by (Fourcade et al. 2014). Alternatives for filtering are either in the environmental space or the geographic space but (Xu et al. 2024) found environmental filtering to be more effective. With validation data, other filtering methods could be explored to see which works best for the variables in this study.

5.3. Environmental data

This research attempted to make predictions at very fine scale and extracted variables that affect the distribution of the species. Variable selection is a source of model inaccuracies (Hanberry 2024)and we were careful not to include many correlated variables. We conducted a dual multicollinearity process to ensure the right predictor variables were chosen for the modelling process.

In socioeconomic studies for VL in West Pokot, sandflies have been observed to mostly bite at night(Abdullahi et al. 2022). Further, termite mounds have high humidity. We were not successful in retrieving datasets for humidity. Nighttime Land Surface Temperature (LST) data from MODIS, available at a 1 km resolution, was considered, however, including this dataset could have influenced our results or possible inaccuracies.

Notably, the poorest-performing model primarily relied on precipitation data with a resolution of 1km as the main predictor. Vector abundance is affected by rainfall (Koskei et al. 2024) and we chose to include it because of its importance in literature.

5.4. Extent

The size and boundaries of the study area greatly affect how well models can predict environmental suitability. Models perform well no matter how big the study area is, but larger areas show more differences in suitability across regions (Amaro et al. 2023). Modelling at larger geographic extents such as the county level could have reduced the amount of predicted risk areas in the current study area or caused a shift in where these suitable places were.

Where most models identify risk at a lower resolution (municipality or region), this does not allow for variation in risk within these regions. In this study, we attempted to develop an approach to model risk at a high resolution. The level of prediction is very high resolution and risk can be identified at the house level. We did not explore model transferability; however, we believe the model can be used to predict VL applied to predict visceral leishmaniasis (VL) in East Africa due to similar climatic, environmental and ecological conditions.

5.5. Wickedness of the Study.

The transmission of VL presents a wicked problem and with limited information on the true location of occurrence, few sample points to account for seasonal variation, and limited information on vector abundance and dispersal, the complexity of predicting VL in West Pokot was challenging. We also did not account for socio-economic dynamics such as pastoralism.

The study reduced some of the wickedness by exploring point redistribution and obtaining proxies for vector data. Defining M reduced uncertainty on vector presence, but it is still necessary to validate the output. There was no stakeholder input for the study apart from government reports and their measures to eradicate the disease through vector control (Mewara et al. 2022; Ministry of Health Kenya 2017). M

5.6. Limitations of the study

5.6.1. Occurrence data

Disease cases served as input data for the study, but its resolution was limited. Enhanced precision would have improved the spatial granularity of the research; however, ethical and privacy concerns prevented the collection of this dataset. Further, a lot of VL cases go unreported in endemic areas due to poor risk perception of the community, and the data collection might have not been the full range. Incorporating additional data, such as the geographical distribution of sandflies in the area, could have further refined the results.

5.6.2. Model validation.

An essential step in validating the model results would be to compare it to possibly other predictions in literature or systematically collected field data. The was no way of confirming the extent of the risk of VL predicted. Further, the predictions made were from simulated data which introduces uncertainty in the prediction.

5.6.3. Data unavailability

The study was also affected by the unavailability of data to quantify some risk factors, such as acacia trees, humidity, and higher resolution for rainfall data. These factors affect the life cycle of the transmitting vector and would have been important predictor variables.

6. Conclusions

6.1. Conclusion

This research developed an ENM and predicted risk for VL in West Pokot and in doing so we answered the following research questions.

1. Where are the vector habitats, and how can they be identified from high-resolution satellite imagery?

The identified vector habitats for sandfly *P. martini* were termite mounds, animal sheds and the banks of seasonal rivers. Using very high-resolution worldview imagery we trained a deep learning model that was able to clearly distinguish animal sheds. However, the method was not successful with termite mounds.

We used two proxies for vector habitats (seasonal riverbanks and animal shelters), a third vector habitat proxy (termite mounds) could not be derived from images. These proxies replaced actual habitat data.

2. What are the relevant environmental variables, what is their contribution to the occurrence of VL infections and can they be used to develop an accurate and reliable model for forecasting the risk?

We extracted environmental variables at very high resolution. After modelling, NDVI had the highest contribution. We were unable to incorporate humidity, high-resolution rainfall data, and acacia trees which are crucial for vector survival. We used NDVI as a proxy for vegetation.

3. What is the predictive accuracy of a machine learning model in forecasting the risk of VL based on the combination of vector ecology and environmental factors in West Pokot, and how can this model be applied to identify risk zones?

We simulated potential dispersal points for infection cases in Kacheliba using the BAM framework to generate input data. We ran a maxent model and the best score was an AUC of 0.805.

6.2. Recommendations

There are quite a few research topics that are logical follow-ups or relevant to this research. They are, but are not limited to the following topics:

- Systematically collect presence and absence points for sandflies in the case study areas and compare their extent to the ones derived from the BAM.
- Stakeholder engagement to assess their perceptions of environmental risk and control studies to obtain locations of susceptible individuals.
- Correction of sampling bias and its applicability to a BAM-derived dataset. We only applied one method of bias sampling correction; other methods may perform better than geographic filtering.
- Incorporating the movement of the nomadic people into the risk prediction.

References

Abdullahi, Bulle, Joshua Mutiso, Fredrick Maloba, John Macharia, Mark Riongoita, and Michael Gicheru. 2022. 'Climate Change and Environmental Influence on Prevalence of Visceral Leishmaniasis in West Pokot County, Kenya'. *Journal of Tropical Medicine* 2022(1):1441576. doi: 10.1155/2022/1441576.

Aklilu, Esayas, Solomon Yared, Araya Gebresilassie, Behailu Legesse, and Asrat Hailu. 2023. 'Phlebotomine Sandflies (Diptera: Psychodidae) of Ethiopia'. *Heliyon* 9(3):e14344. doi: 10.1016/J.HELIYON.2023.E14344.

Alford, John, and Brian W. Head. 2017. 'Wicked and Less Wicked Problems: A Typology and a Contingency Framework'. *Policy and Society* 36(3):397–413. doi: 10.1080/14494035.2017.1361634.

Alvar, Jorge, Margriet den Boer, and Daniel Argaw Dagne. 2021. 'Towards the Elimination of Visceral Leishmaniasis as a Public Health Problem in East Africa: Reflections on an Enhanced Control Strategy and a Call for Action'. *The Lancet Global Health* 9(12):e1763–69. doi: 10.1016/S2214-109X(21)00392-2.

Alves, Fabiana, Graeme Bilbe, Séverine Blesson, Vishal Goyal, Séverine Monnerat, Charles Mowbray, Gina Muthoni Ouattara, Bernard Pécoul, Suman Rijal, Joelle Rode, Alexandra Solomos, Nathalie Strub-Wourgaft, Monique Wasunna, Susan Wells, Eduard E. Zijlstra, Byron Arana, and Jorge Alvar. 2018. 'Recent Development of Visceral Leishmaniasis Treatments: Successes, Pitfalls, and Perspectives'. *Clinical Microbiology Reviews* 31(4):1–30. doi: 10.1128/CMR.00048-18/ASSET/D74A5965-E6AB-4624-B5FE-B1A9CBFF90FB/ASSETS/GRAPHIC/ZCM0041826440002.JPEG.

Amaro, George, Elisangela Gomes Fidelis, Ricardo Siqueira da Silva, and Cesar Augusto Marchioro. 2023. 'Effect of Study Area Extent on the Potential Distribution of Species: A Case Study with Models for Raoiella Indica Hirst (Acari: Tenuipalpidae)'. *Ecological Modelling* 483:110454. doi: 10.1016/J.ECOLMODEL.2023.110454.

Balint, Peter J., Ronald E. Stewart, Anand Desai, and Lawrence C. Walters. 2011. 'Wicked Environmental Problems'. *Wicked Environmental Problems*. doi: 10.5822/978-1-61091-047-7.

Benavides Rios, Eva, Jonathan Sadler, Laura Graham, and Thomas J. Matthews. 2024. 'Species Distribution Models and Island Biogeography: Challenges and Prospects'. *Global Ecology and Conservation* 51:e02943. doi: 10.1016/J.GECCO.2024.E02943.

Cao, Yangyang, Zuoxi Zhao, Yuan Huang, Xu Lin, Shuyuan Luo, Borui Xiang, and Houcheng Yang. 2023. 'Case Instance Segmentation of Small Farmland Based on Mask R-CNN of Feature Pyramid Network with Double Attention Mechanism in High Resolution Satellite Images'. *Computers and Electronics in Agriculture* 212:108073. doi: 10.1016/J.COMPAG.2023.108073.

Capucci, Débora Cristina, Aldenise Martins Campos, João Vítor Reis Soares, Vladimir Diniz Vieira Ramos, Camila Binder, Mariana Alves Lima, Carina Margonari, and José Dilermando Andrade Filho. 2023. 'Ecology and Natural Infection of Phlebotomine Sand Flies in Different Ecotopes and Environments in the Municipality of Pains, Minas Gerais, Brazil'. *Acta Tropica* 238:106789. doi: 10.1016/J.ACTATROPICA.2022.106789.

Cecílio, Pedro, Anabela Cordeiro-da-Silva, and Fabiano Oliveira. 2022. 'Sand Flies: Basic Information on the Vectors of Leishmaniasis and Their Interactions with Leishmania Parasites'. *Communications Biology 2022 5:1* 5(1):1–12. doi: 10.1038/s42003-022-03240-z.

Chabot-Couture, Guillaume, Karima Nigmatulina, and Philip Eckhoff. 2014. 'An Environmental Data Set for Vector-Borne Disease Modeling and Epidemiology'. *PLOS ONE* 9(4):e94741. doi: 10.1371/JOURNAL.PONE.0094741.

Chavy, Agathe, Alessandra Ferreira Dales Nava, Sergio Luiz Bessa Luz, Juan David Ramírez, Giovanny Herrera, Thiago Vasconcelos Dos Santos, Marine Ginouves, Magalie Demar, Ghislaine Prévot, Jean François Guégan, and Benoit De Thoisy. 2019. 'Ecological Niche Modelling for Predicting the Risk of Cutaneous Leishmaniasis in the Neotropical Moist Forest Biome'. *PLOS Neglected Tropical Diseases* 13(8):e0007629. doi: 10.1371/JOURNAL.PNTD.0007629.

Cheng, Tao, Xusheng Ji, Gaoxiang Yang, Hengbiao Zheng, Jifeng Ma, Xia Yao, Yan Zhu, and Weixing Cao. 2020. 'DESTIN: A New Method for Delineating the Boundaries of Crop Fields by Fusing Spatial and Temporal Information from WorldView and Planet Satellite Imagery'. *Computers and Electronics in Agriculture* 178:105787. doi: 10.1016/J.COMPAG.2020.105787.

Cheng, Yanchao, Nils Benjamin Tjaden, Anja Jaeschke, Stephanie Margarete Thomas, and Carl Beierkuhnlein. 2021. 'Using Centroids of Spatial Units in Ecological Niche Modelling: Effects on Model Performance in the Context of Environmental Data Grain Size'. *Global Ecology and Biogeography* 30(3):611–21. doi: 10.1111/GEB.13240.

Chollet Ramampiandra, Emma, Andreas Scheidegger, Jonas Wydler, and Nele Schuwirth. 2023. 'A Comparison of Machine Learning and Statistical Species Distribution Models: Quantifying Overfitting Supports Model Interpretation'. *Ecological Modelling* 481:110353. doi: 10.1016/J.ECOLMODEL.2023.110353.

Climatedata.org. n.d. 'Kacheliba Climate: Weather Kacheliba & Temperature by Month'. Retrieved 9 June 2024 (https://en.climate-data.org/africa/kenya/west-pokot/kacheliba-991749/#climate-table).

Collart, Flavien, and Antoine Guisan. 2023. 'Small to Train, Small to Test: Dealing with Low Sample Size in Model Evaluation'. *Ecological Informatics* 75:102106. doi: 10.1016/J.ECOINF.2023.102106.

Van Dijk, Norbert, Jane Carter, Wyckliff Omondi, Petra Mens, and Henk Schallig. 2023. 'Clinical Features, Immunological Interactions and Household Determinants of Visceral Leishmaniasis and Malaria Coinfections in West Pokot, Kenya: Protocol for

an Observational Study'. *BMJ Open* 13(4):e068679. doi: 10.1136/BMJOPEN-2022-068679.

Duncan, Erik C., Sergii Skakun, Ankit Kariryaa, and Alexander V. Prishchepov. 2023. 'Detection and Mapping of Artillery Craters with Very High Spatial Resolution Satellite Imagery and Deep Learning'. *Science of Remote Sensing* 7:100092. doi: 10.1016/J.SRS.2023.100092.

Elith, Jane, Steven J. Phillips, Trevor Hastie, Miroslav Dudík, Yung En Chee, and Colin J. Yates. 2011. 'A Statistical Explanation of MaxEnt for Ecologists'. *Diversity and Distributions* 17(1):43–57. doi: 10.1111/J.1472-4642.2010.00725.X.

Elnaiem, DE. 2011. 'Ecology and Control of the Sand Fly Vectors of Leishmania Donovani in East Africa, with Special Emphasis on Phlebotomus Orientalis'. *Journal of Vector Ecology : Journal of the Society for Vector Ecology* 36 Suppl 1:S23-31. doi: 10.1111/j.1948-7134.2011.00109.x.

Escobar, Luis E., and Meggan E. Craft. 2016. 'Advances and Limitations of Disease Biogeography Using Ecological Niche Modeling'. *Frontiers in Microbiology* 7(AUG):188208. doi: 10.3389/FMICB.2016.01174/BIBTEX.

ESRI. 2013. 'Arc Hydro: GIS for Water Resources'. Retrieved 8 April 2024 (https://books.google.nl/books?hl=en&lr=&id=07vH7Sf0v6MC&oi=fnd&pg=PP7&dq=related:CYuhCtfox40J:scholar.google.com/&ots=alKwBBfbix&sig=Z1oVBswoLiGFJWyCWy0iVdy2BuY&redir_esc=y#v=onepage&q&f=false).

ESRI. n.d. 'Train Deep Learning Model (Image Analyst)—ArcGIS Pro | Documentation'. Retrieved 21 June 2024 (https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/train-deep-learning-model.htm).

Feng, Yingchao, Wenhui DIao, Yi Zhang, Hao Li, Zhonghan Chang, Menglong Yan, Xian Sun, and Xin Gao. 2019. 'Ship Instance Segmentation from Remote Sensing Images Using Sequence Local Context Module'. *International Geoscience and Remote Sensing Symposium (IGARSS)* 1025–28. doi: 10.1109/IGARSS.2019.8897948.

Fourcade, Yoan, Jan O. Engler, Dennis Rödder, and Jean Secondi. 2014. 'Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias'. *PLOS ONE* 9(5):e97122. doi: 10.1371/JOURNAL.PONE.0097122.

Grimmett, Liam, Rachel Whitsed, and Ana Horta. 2020. 'Presence-Only Species Distribution Models Are Sensitive to Sample Prevalence: Evaluating Models Using Spatial Prediction Stability and Accuracy Metrics'. *Ecological Modelling* 431:109194. doi: 10.1016/J.ECOLMODEL.2020.109194.

Hanberry, Brice B. 2024. 'Practical Guide for Retaining Correlated Climate Variables and Unthinned Samples in Species Distribution Modeling, Using Random Forests'. *Ecological Informatics* 79:102406. doi: 10.1016/J.ECOINF.2023.102406.

Hassaballa, Iman B., Catherine L. Sole, Xavier Cheseto, Baldwyn Torto, and David P. Tchouassi. 2021. 'Afrotropical Sand Fly-Host Plant Relationships in a Leishmaniasis Endemic Area, Kenya'. *PLOS Neglected Tropical Diseases* 15(2):e0009041. doi: 10.1371/JOURNAL.PNTD.0009041.

Hassaballa, Iman B., Baldwyn Torto, Catherine L. Sole, and David P. Tchouassi. 2021. 'Exploring the Influence of Different Habitats and Their Volatile Chemistry in Modulating Sand Fly Population Structure in a Leishmaniasis Endemic Foci, Kenya'. *PLOS Neglected Tropical Diseases* 15(2):e0009062. doi: 10.1371/JOURNAL.PNTD.0009062.

Jiang, Dong, Tian Ma, Mengmeng Hao, Yushu Qian, Shuai Chen, Ze Meng, Liping Wang, Canjun Zheng, Xiao Qi, Qian Wang, and Fangyu Ding. 2021. 'Spatiotemporal Patterns and Spatial Risk Factors for Visceral Leishmaniasis from 2007 to 2017 in Western and Central China: A Modelling Analysis'. *Science of The Total Environment* 764:144275. doi: 10.1016/J.SCITOTENV.2020.144275.

Jimenez-Munoz, Juan C., Jordi Cristobal, José A. Sobrino, Guillem Sòria, Miquel Ninyerola, and Xavier Pons. 2009. 'Revision of the Single-Channel Algorithm for Land Surface Temperature Retrieval from Landsat Thermal-Infrared Data'. *IEEE Transactions on Geoscience and Remote Sensing* 47(1):339–49. doi: 10.1109/TGRS.2008.2007125.

Jiménez-Valverde, Alberto. 2020. 'Sample Size for the Evaluation of Presence-Absence Models'. *Ecological Indicators* 114:106289. doi: 10.1016/J.ECOLIND.2020.106289.

Kim, Jong Hae. 2019. 'Multicollinearity and Misleading Statistical Results'. *Korean Journal of Anesthesiology* 72(6):558–69. doi: 10.4097/KJA.19087.

Kirstein, Oscar David, Laura Skrip, Ibrahim Abassi, Tamara Iungman, Ben Zion Horwitz, Araya Gebresilassie, Tatiana Spitzova, Yoni Waitz, Teshome Gebre-Michael, Petr Volf, Asrat Hailu, and Alon Warburg. 2018. 'A Fine Scale Eco-Epidemiological Study on Endemic Visceral Leishmaniasis in North Ethiopian Villages'. *Acta Tropica* 183:64–77. doi: 10.1016/J.ACTATROPICA.2018.04.005.

Koskei, Edith, Solomon Langat, James Mutisya, Francis Mulwa, Joel Lutomiah, Hellen Koka, Samuel O. Oyola, Rebecca Waihenya, Sepha N. Mabeya, and Rosemary Sang. 2024. 'Isolation and Phylogenetic Characterization of Arboviruses Circulating among Phlebotomine Sandflies in Parts of North Rift, Kenya'. *Frontiers in Virology* 4:1289258. doi: 10.3389/FVIRO.2024.1289258/BIBTEX.

Lamboley, Quentin, and Yoan Fourcade. 2024. 'No Optimal Spatial Filtering Distance for Mitigating Sampling Bias in Ecological Niche Models'. *Journal of Biogeography* 00:1–12. doi: 10.1111/JBI.14854.

Mewara, Abhishek, Rajendra Gudisa, Bijaya Kumar Padhi, Pawan Kumar, Ranjit Sah, and Alfonso J. Rodriguez-Morales. 2022. 'Visceral Leishmaniasis Outbreak in Kenya—a Setback to the Elimination Efforts'. *New Microbes and New Infections* 49–50:101060. doi: 10.1016/J.NMNI.2022.101060.

Miller, Matthew A. E., Keith D. Shepherd, Bruce Kisitu, and Jamie Collinson. 2021. 'ISDAsoil: The First Continent-Scale Soil Property Map at 30 m Resolution Provides a Soil Information Revolution for Africa'. *PLoS Biology* 19(11). doi: 10.1371/JOURNAL.PBIO.3001441.

Ministry of Health Kenya. 2017. *PREVENTION, DIAGNOSIS AND TREATMENT OF VISCERAL LEISHMANIASIS (KALA-AZAR) IN KENYA* .

Mueller, Yolanda K., Jan H. Kolaczinski, Timothy Koech, Peter Lokwang, Mark Riongoita, Elena Velilla, Simon J. Brooker, and François Chappuis. 2014. 'Clinical Epidemiology, Diagnosis and Treatment of Visceral Leishmaniasis in the Pokot Endemic Area of Uganda and Kenya'. *The American Journal of Tropical Medicine and Hygiene* 90(1):33–39. doi: 10.4269/AJTMH.13-0150.

Muscarella, Robert, Peter J. Galante, Mariano Soley-Guardia, Robert A. Boria, Jamie M. Kass, María Uriarte, and Robert P. Anderson. 2014. 'ENMeval: An R Package for Conducting Spatially Independent Evaluations and Estimating Optimal Model Complexity for Maxent Ecological Niche Models'. *Methods in Ecology and Evolution* 5(11):1198–1205. doi: 10.1111/2041-210X.12261.

O'Brien, Robert M. 2007. 'A Caution Regarding Rules of Thumb for Variance Inflation Factors'. *Quality and Quantity* 41(5):673–90. doi: 10.1007/S11135-006-9018-6/METRICS.

Obwocha, Everlyne B., Joshua J. Ramisch, Lalisa Duguma, and Levi Orero. 2022. 'The Relationship between Climate Change, Variability, and Food Security: Understanding the Impacts and Building Resilient Food Systems in West Pokot County, Kenya'. *Sustainability 2022, Vol. 14, Page 765* 14(2):765. doi: 10.3390/SU14020765.

Phillips, Sharon B., Viney P. Aneja, Daiwen Kang, and S. Pal Arya. 2006. 'Maximum Entropy Modeling of Species Geographic Distributions'. *Ecological Modelling* 190(3–4):231–59. doi: 10.1016/J.ECOLMODEL.2005.03.026.

Sardar, Ashif Ali, Moytrey Chatterjee, Kingsuk Jana, Pabitra Saha, Ardhendu Kumar Maji, Subhasish Kamal Guha, and Pratip Kumar Kundu. 2020. 'Seasonal Variation of Sand Fly Populations in Kala-Azar Endemic Areas of the Malda District, West Bengal, India'. *Acta Tropica* 204:105358. doi: 10.1016/J.ACTATROPICA.2020.105358.

Shanzu, Irissheel. 2023. 'West Pokot Records the Highest Cases of Disease Spread by Sandflies - The Standard Health'.

Sillero, Neftalí, Salvador Arenas-Castro, Urtzi Enriquez-Urzelai, Cândida Gomes Vale, Diana Sousa-Guedes, Fernando Martínez-Freiría, Raimundo Real, and A. Márcia Barbosa. 2021. 'Want to Model a Species Niche? A Step-by-Step Guideline on Correlative Ecological Niche Modelling'. *Ecological Modelling* 456:109671. doi: 10.1016/J.ECOLMODEL.2021.109671.

Sirko, Wojciech, Sergii Kashubin, Marvin Ritter, Abigail Annkah, Yasser Salah Eddine Bouchareb, Yann Dauphin, Daniel Keysers, Maxim Neumann, Moustapha Cisse, and John Quinn. 2021. 'Continental-Scale Building Detection from High Resolution Satellite Imagery'.

Soberon, Jorge, and A. Townsend Peterson. 2005. 'Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas'. *Biodiversity Informatics* 2(0):1–10. doi: 10.17161/BI.V2I0.4.

Tyrrell, Peter, Irene Amoke, Koen Betjes, Femke Broekhuis, Robert Buitenwerf, Sarah Carroll, Nathan Hahn, Daniel Haywood, Britt Klaassen, Mette Løvschal, David Macdonald, Karen Maiyo, Hellen Mbithi, Nelson Mwangi, Churchil Ochola, Erick

Odire, Victoria Ondrusek, Junior Ratemo, Frank Pope, Samantha Russell, Wilson Sairowua, Kiptoo Sigilai, Jared A. Stabach, Jens Christian Svenning, Elizabeth Stone, Johan T. du Toit, Guy Western, George Wittemyer, and Jake Wall. 2022. 'Landscape Dynamics (LandDX) an Open-Access Spatial-Temporal Database for the Kenya-Tanzania Borderlands'. *Scientific Data* 9(1). doi: 10.1038/S41597-021-01100-9.

Tyrrell, Peter, Robin Naidoo, David W. Macdonald, and Johan T. du Toit. 2021. 'New Forces Influencing Savanna Conservation: Increasing Land Prices Driven by Gentrification and Speculation at the Landscape Scale'. *Frontiers in Ecology and the Environment* 19(9):494–500. doi: 10.1002/FEE.2391.

Vrieling, Anton, Francesco Fava, Sonja Leitner, Lutz Merbold, Yan Cheng, Teopista Nakalema, Thomas Groen, and Klaus Butterbach-Bahl. 2022. 'Identification of Temporary Livestock Enclosures in Kenya from Multi-Temporal PlanetScope Imagery'. *Remote Sensing of Environment* 279:113110. doi: 10.1016/J.RSE.2022.113110.

West Pokot. n.d. 'County Overview'. Retrieved 5 June 2024 (https://westpokot.go.ke/about-us).

Xu, Quanli, Xiao Wang, Junhua Yi, and Yu Wang. 2024. 'Bias Correction in Species Distribution Models Based on Geographic and Environmental Characteristics'. *Ecological Informatics* 81:102604. doi: 10.1016/J.ECOINF.2024.102604.

Zhang, Weixing, Chandi Witharana, Anna K. Liljedahl, and Mikhail Kanevskiy. 2018. 'Deep Convolutional Neural Networks for Automated Characterization of Arctic Ice-Wedge Polygons in Very High Spatial Resolution Aerial Imagery'. *Remote Sensing 2018, Vol. 10, Page 1487* 10(9):1487. doi: 10.3390/RS10091487.

Appendix 1

AI guidelines

In line with the AI guidelines from the University of Twente.

During the preparation of this work, the author used ChatGPT to debug the codes used in the thesis. After using this tool/service, the author reviewed and edited the content as needed and take(s) full responsibility for the content of the work."

Reproducibility

All relevant data and scripts will be added to the UT SurfDrive folder.

- The deep learning model can be used for transfer learning and can detect shed boundaries on base maps with <50cm resolution such as ESRI imagery.
- All environmental datasets used are freely available and can be obtained from Google Earth Engine
- The Maxent model is reproducible and can predict other regions with the correct environmental dataset.

Appendix 2

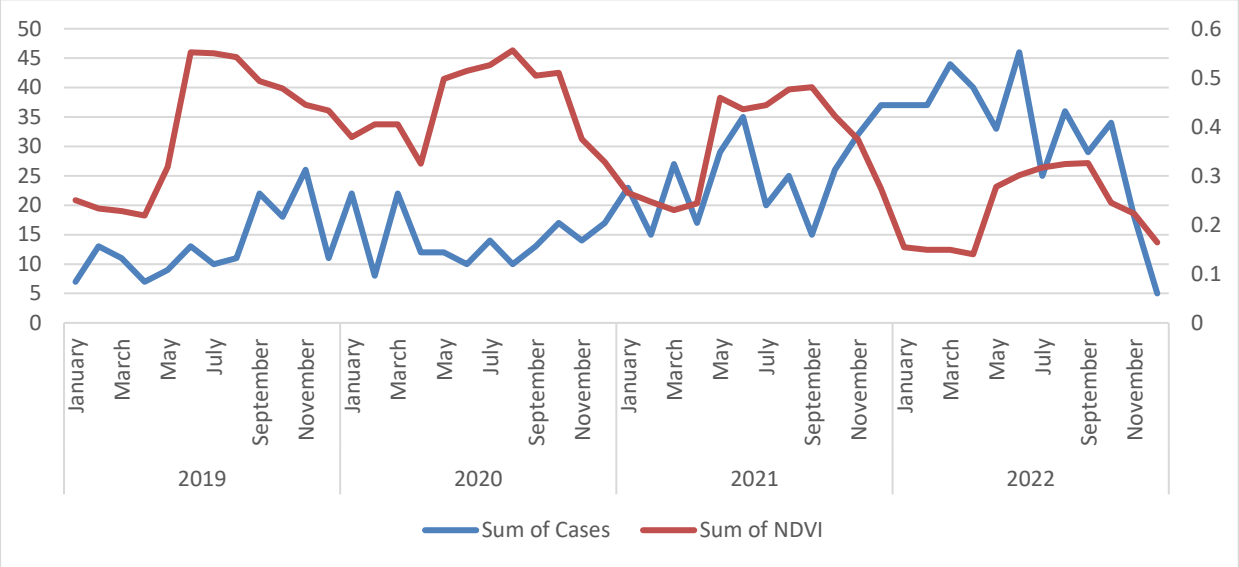

Figure 20: Temporal relationship of NDVI and VL occurrence, which may explain its dominance as a predictor variable.

ENM Evaluation

	fc	rm	tune.args	auc.train	cbi.train	auc.diff.av	auc.diff.sd	auc.val.av	auc.val.sd	cbi.val.avg	cbi.val.sd	or.10p.avg	or.10p.sd	or.mtp.avg	or.mtp.sd	AICc	delta.AICc	w.AIC	ncoef
1	L	1	fc.L_rm.1	0.658891	0.799	0.039479	0.026481	0.620681	0.019536	0.5052	0.278799	0.100615	0.084065	0.015385	0.034401	3231.661	23.69325	4.80E-06	9
2	LQ	1	fc.LQ_rm.1	0.673406	0.739	0.041676	0.030426	0.635447	0.022982	0.5818	0.183635	0.124923	0.062968	0.023077	0.051602	3228.579	20.612	2.24E-05	14
3	H	1	fc.H_rm.1	0.7355	0.982	0.080386	0.050698	0.659452	0.04465	0.661	0.11384	0.210462	0.074376	0.039077	0.027209	3235.582	27.61511	6.76E-07	32
4	LQH	1	fc.LQH_rm	0.734672	0.965	0.082644	0.043951	0.657002	0.037917	0.6868	0.252376	0.202769	0.062912	0.031385	0.032374	3232.414	24.44627	3.30E-06	32
5	LQHP	1	fc.LQHP_ri	0.733547	0.965	0.083294	0.043985	0.657643	0.037909	0.6914	0.232686	0.202769	0.062912	0.046769	0.042	3225.15	17.18281	0.000125	30
6	LQHPT	1	fc.LQHPT_	0.745188	0.978	0.110015	0.035103	0.64714	0.034166	0.5682	0.249387	0.257846	0.052821	0.054154	0.06427	3234.077	26.10991	1.43E-06	34
7	L	2	fc.L_rm.2	0.653094	0.701	0.033704	0.022627	0.622684	0.017503	0.5426	0.281456	0.100923	0.074434	0.015385	0.034401	3231.656	23.68866	4.81E-06	9
8	LQ	2	fc.LQ_rm.2	0.656891	0.894	0.033336	0.026323	0.629207	0.018836	0.5692	0.289787	0.092923	0.084136	0.023077	0.051602	3238.019	30.05158	2.00E-07	14
9	H	2	fc.H_rm.2	0.695883	0.944	0.056255	0.032671	0.647446	0.024461	0.5452	0.158464	0.156	0.046251	0.038462	0.054393	3213.749	5.781234	0.03725	16
10	LQH	2	fc.LQH_rm	0.700703	0.943	0.053062	0.031411	0.650436	0.024542	0.4808	0.349752	0.164	0.070279	0.015385	0.034401	3214.164	6.196962	0.030259	18
11	LQHP	2	fc.LQHP_ri	0.701047	0.941	0.051984	0.031847	0.652406	0.025682	0.5582	0.312517	0.140308	0.064049	0.015385	0.034401	3207.967	0	0.670659	16
12	LQHPT	2	fc.LQHPT_	0.702328	0.947	0.059193	0.033035	0.648527	0.026615	0.5102	0.322818	0.148308	0.056711	0.023077	0.051602	3212.921	4.953607	0.056343	18
13	L	3	fc.L_rm.3	0.647719	0.674	0.025202	0.020183	0.624119	0.016485	0.5746	0.235373	0.100923	0.074434	0.015385	0.034401	3227.75	19.78306	3.39E-05	7
14	LQ	3	fc.LQ_rm.3	0.647203	0.909	0.025848	0.020402	0.6285	0.016968	0.6064	0.231436	0.092923	0.079239	0.023077	0.051602	3227.191	19.22353	4.49E-05	9
15	H	3	fc.H_rm.3	0.675492	0.903	0.040531	0.021197	0.631619	0.016346	0.4898	0.363389	0.116308	0.071141	0.038462	0.054393	3211.832	3.864378	0.097132	8
16	LQH	3	fc.LQH_rm	0.677766	0.943	0.040585	0.027651	0.635999	0.022398	0.5144	0.334843	0.108308	0.078149	0.015385	0.034401	3215.325	7.357721	0.016935	12
17	LQHP	3	fc.LQHP_ri	0.679234	0.957	0.038729	0.028836	0.639063	0.024509	0.5338	0.2993	0.108923	0.073769	0.015385	0.034401	3214.33	6.362829	0.02785	12
18	LQHPT	3	fc.LQHPT_	0.679234	0.957	0.041624	0.025973	0.637527	0.022474	0.5254	0.290019	0.108923	0.073769	0.015385	0.034401	3214.33	6.362829	0.02785	12
19	L	4	fc.L_rm.4	0.642078	0.695	0.022704	0.019391	0.620892	0.017343	0.5498	0.218583	0.085231	0.06264	0.007692	0.017201	3226.523	18.55557	6.27E-05	6
20	LQ	4	fc.LQ_rm.4	0.640516	0.905	0.020935	0.017754	0.627502	0.018237	0.6446	0.200207	0.085231	0.06264	0.007692	0.017201	3223.499	15.53185	0.000284	7
21	H	4	fc.H_rm.4	0.644148	0.898	0.027053	0.017484	0.618599	0.011941	0.4876	0.478864	0.116615	0.046517	0.015385	0.021066	3225.129	17.16143	0.000126	7
22	LQH	4	fc.LQH_rm	0.656594	0.949	0.025756	0.026258	0.629699	0.022909	0.5716	0.320793	0.092923	0.063717	0.015385	0.034401	3217.229	9.2621	0.006535	8
23	LQHP	4	fc.LQHP_ri	0.658813	0.936	0.026227	0.025524	0.632821	0.024596	0.569	0.27588	0.100923	0.074434	0.015385	0.034401	3216.428	8.460144	0.009759	8
24	LQHPT	4	fc.LQHPT_	0.658813	0.936	0.025845	0.025917	0.632085	0.023614	0.5586	0.260204	0.100923	0.074434	0.015385	0.034401	3216.428	8.460144	0.009759	8
25	L	5	fc.L_rm.5	0.634516	0.675	0.022268	0.01599	0.619359	0.018207	0.5248	0.253762	0.085231	0.06264	0.007692	0.017201	3227.688	19.72015	3.50E-05	6
26	LQ	5	fc.LQ_rm.5	0.635813	0.91	0.021574	0.015726	0.62399	0.020339	0.613	0.199059	0.085231	0.06264	0.007692	0.017201	3220.272	12.30441	0.001428	5
27	H	5	fc.H_rm.5	0.627617	0.741	0.018213	0.015493	0.616136	0.01879	0.4046	0.447831	0.108923	0.056771	0.023077	0.021066	3222.547	14.57988	0.000458	4
28	LQH	5	fc.LQH_rm	0.639016	0.932	0.02378	0.020058	0.623475	0.022371	0.5862	0.234253	0.092923	0.063717	0.007692	0.017201	3221.793	13.82551	0.000667	7
29	LQHP	5	fc.LQHP_ri	0.641906	0.93	0.023794	0.020918	0.627083	0.024339	0.5866	0.231301	0.109231	0.063953	0.007692	0.017201	3218.668	10.70061	0.003183	6
30	LQHPT	5	fc.LQHPT_	0.641906	0.93	0.022863	0.020942	0.627115	0.024392	0.5858	0.229832	0.109231	0.063953	0.007692	0.017201	3218.668	10.70061	0.003183	6

Figure 21: Samples ENM Eval results for region 2 subset data. Model with delta.AIC = 0 is selected. In this case RM=2, features LQHP